Numerical methods of resonant dynamics for the Galaxy

Ralf Donner
Rudolf Peierls Centre for Theoretical Physics
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford

Hilary 2005

Numerical methods of resonant dynamics for the Galaxy

Ralf Donner
St Cross College

Rudolf Peierls Centre for Theoretical Physics
University of Oxford

Abstract

Numerical methods of resonant dynamics with applications to the Galaxy are considered in this thesis. We derive generating functions for first-order perturbation theory and the associated orbital frequencies by matrix calculus. For two action-angle spaces (J, θ) and (I, ϕ) related by a canonical map $I \cdot \theta + s$, we show that J can be averaged over ergodic orbits ϕ to provide an estimator of I to within $O(|s|^2)$. We provide examples in one and two dimensions and compare the technique to calculations of actions by numerical line integration in Poincaré sections.

We then use spectral dynamics and the Laskar frequency map (Laskar, 1993) to identify the dynamically important resonances of the 'flattened' axisymmetric isochrone potential. We simulate resonant capture in a low-order resonance by populating representative tori of a spherical isochrone Hamiltonian and integrating the orbits while adiabatically introducing axisymmetry. We use the averaging technique described above to observe the fraction of orbits captured, and we compare the result to a theoretical prediction.

We return to first-order perturbation theory to analyse its strengths and weaknesses, in particular near orbital pericentre, and when one action is significantly smaller than another. We also reproduce the expected pendulum dynamics in the resonant action-angle plane for orbits in our capture simulation.

We develop the concept of *adaptive dynamics*: we vary the initial orbital energy of the particles in the capture simulations and show that resonant and non-resonant orbits can be identified as clusters in the perturbed action plane. For a given Hamiltonian, we use the perturbed frequencies and a linear regression fit in the action plane as diagnostics of a set of model Hamiltonians on a grid in a suitable parameter space. We find we are able to constrain the parameters of a model Hamiltonian by this method. Finally, we reject the null hypothesis that resonant structures in phase space can be found by traditional methods of density estimation.

A thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford

Hilary 2005

Acknowledgements

Acknowledgements go to my supervisor, Professor James Binney F.R.S., for his collaboration. James is one of the most remarkable individuals I ever hope to meet, and both his astounding mental dexterity and lively sense of humour have contributed immeasurably to my enjoyment of this project. I also wish to thank him for providing me with the program code for first-order perturbation theory and spectral dynamics. These formed the basis for the numerical work of Chapters 3 and 4.

My office-mates should not go without honourable mention. Listening to Stuart Sim impeccably blasting out Mahler's Symphony No. 2 on the tuba in the Sheldonian, solving irrelevant physics problems on the blackboard to our utter satisfaction with Stuart and Simon Berman, and exploring the obscurities of Vasily Belokurov's worldly wisdom, were counterbalanced by the great comfort I took in knowing that we will all be saved from Armageddon by "Jonti" Horner. Further delights of office life included learning ä chli Bääääärndütsch from Sarah Sägesser, hearing an insider's account of both Sellafield AND Cambridge from Andy Eyre, and attempting under patient but futile instruction to intone Mimi Zhang's name. And where would I be without Rachel Koncewicz's delicious leaf tea and the best hair-cut in a theoretical physics sub-department anywhere?

I dedicate this thesis to my mother and father, who created the Universe.

R.M.D.
Oxford, March 2005

Contents

1 Introduction

In *The Republic*, Socrates argues[1] that '*the stars that decorate the sky...are far inferior...to the true realities, that is, to the true relative velocities, in pure numbers and perfect figures, of the orbits and what they carry in them, which are perceptible to reason and thought, but not visible to the eye.*' We should perhaps not infer too much from the muddled geocentric Platonic picture of the heavens, but there are two aspects of this quotation that are quite harmonious with modern astrophysics. First, there is the concept that the dynamics of their orbits exists independently of the individual stars, and second, that 2417 years after this Socratic dialogue, an unexpected wealth of structure was discovered in the velocity space of Solar neighbourhood stars (Dehnen, 1998). Famaey et al. (2005) revealed that stars with completely different histories are clustered in groups in velocity space, while de Simone, Wu, and Tremaine (2004) explained this phenomenon in the context of heating by spiral structure. The debate on the origin of these groupings is in its infancy, but it seems apparent that we should be looking for the culprit in a field that has already troubled celestial mechanics for more than a century: resonant dynamics.

Four hundred years ago, the only astrophysical system in which mathematically rigorous theories could be tested was the Solar System. In *Harmonices mundi* (1619), Kepler, convinced of the perfection of the Solar System, went to lengths to find harmonies in the eccentricities of orbits, and he found them. To Kepler, these discoveries were completely in keeping with the expectations of a rational Universe, and it was only with Newton that a systematic investigation of dynamics began to take preference. One of Newton's

---

[1] Plato. *The Republic*, transl. H.D.P. Lee, Penguin Books, London, 1987.

many famous triumphs was to derive Kepler's laws from dynamical principles, but, in the process, he was struck by a serious, potentially catastrophic flaw in the perfection of the Solar System. Newton realized that Jupiter introduced an element of dynamical instability, and there seemed to be no obvious way of limiting the damage. Indeed, it was fortunate that the contemporary view had the Creation of the Solar System as late as October 22nd, 4004 BC. Newton nevertheless postulated that there would need to be Divine intervention periodically to redress the destabilizing effects of a many-body problem.

By the 18th century, astrophysicists were more optimistic. Lagrange had invented a new and powerful mechanics in generalized coordinates (*Mécanique analytique*, 1788), and he used it to provide evidence that the Solar System was stable. His contemporary Laplace is perhaps more famous in this context: his five-volume *Mécanique céleste* (1799-1825) stood unopposed until the end of the 19th century as a 'proof' of the fundamental stability of the Solar System. Specifically, Laplace showed that if one considered only secular terms, with first-order perturbations in planetary mass and second-order perturbations in eccentricity, then the Solar System was stable. Lagrange extended the result to all orders in eccentricity, and Poisson showed that a second-order perturbation in planetary masses led to periodic variations in the semi-major axes of the disturbed orbits.[2] Because of this work, it was widely held that the Solar System was stable.

In the 19th century, the volume of work on the many-body problem grew enormously, and Bessel, Hamilton, Jacobi and Liouville all made significant contributions. In 1889, the journal *Acta mathematica* organised a competition to demonstrate the stability of the Solar

---

[2] For more details, see Murray, N. Solar System Dynamics.
http://www.cita.utoronto.ca/~murray/Papers/CAP/ms.ps

System in honour of the sixtieth birthday of King Oscar II of Norway and Sweden, and Poincaré won it with his diagrammatical solutions referred to as *Poincaré sections*. It was after he had submitted that a colleague pointed out a slight problem. The proofs that had already been circulated to Weierstraß, but fortunately not to his arch-enemy Kronecker, were collected, and the paper was extensively revised.[3] Poincaré's subsequent discovery of the ***homoclinic tangle*** (see Chapter 3) proved the opposite of his original intention. Far from bounding chaos, his results seemed to show that even minor departures from integrability produced instability. In the equations, this instability manifested itself in the ***problem of small divisors***: at a dense set of resonant locations, regular orbits did not exist.

This was still the lamentable state of celestial mechanics in 1950. In quantum electrodynamics the problem of divergent integrals for Feynman diagrams with closed loops of virtual particles had been eliminated, but there was no equivalent renormalization theory for the problem of small divisors in classical perturbation theory. The breakthrough came with the work of Kolmogorov in the 1950s (Kolmogorov, 1954), with subsequent improvements by Arnold and Moser in the 1960s. KAM theory provided a constructive proof of the presence of regular orbits in perturbed integrable Hamiltonians. Finally there existed a successful perturbation theory on a Cantor set, and in fact, since the distance from resonance required for the proof is a sufficient rather than a necessary condition, the regular domains could be quite large in practice.

It became apparent that the presence of resonances in a dynamical system does not necessarily result in global chaos: resonant trapping merely permits new dynamical structures

---

[3] A detailed account of this is to be found in (Peter Galison. 2003, *Einstein's Clocks, Poincaré's Maps*, W. W. Norton and Company, New York).

to arise. The focus therefore shifted towards estimates for the threshold of the departure from integrability needed for the destruction of all KAM tori and the onset of global chaos. The *Chirikov criterion* (Chirikov, 1979) was an attempt to quantify this transition in systems with two degree of freedom, on the basis of the overlap of the separatrices of two resonances. According to this argument, while the separatrices of two resonances remain non-intersecting, KAM tori must exist between them. Since an invariant KAM torus divides the isoenergetic 3-manifold $H = E$ into an inside and an outside, orbits integrated from initial conditions in either resonance are confined to that resonance. Numerical tests showed, however, that the *golden KAM tori*, the 'most irrational' and hence longest surviving family of tori with frequency ratio $\frac{\sqrt{5}-1}{2}$, are destroyed long before Chirikov's rather approximate overlap criterion (see references in Morbidelli, 2002). The *Nekhoroshev theorem* (Nekhoroshev, 1977), valid in any number of dimensions, remains more promising. It quantifies the transition to chaos by considering the overlap of resonances of different order.[4] Stable systems are said to have the *Nekhoroshev structure.* In three-dimensional systems, the presence of the Nekhoroshev structure is easily verifiable in a Laskar frequency map of a Hamiltonian's leading resonances (see Chapters 3 and 6).

While Chirikov and Nekhoroshev treated the transition to chaos, others, spurred by the birth of the electronic computer, were investigating to what extent local quasi-integrals provided stability. These also were the first small steps towards the application of near-integrable dynamics to Hamiltonians of relevance to the Galaxy. In the 1960s, Hénon and Heiles studied a variety of Hamiltonians (Hénon and Heiles, 1964 and subsequent papers), seeking demonstrations of their integrability. The results were often surprising: a

---

[4] These terms will be explained later in this chapter.

Hamiltonian that had long been conjectured to be non-integrable was found to have a third integral after all, while an ostensibly integrable system was shown to be non-integrable.[5] Inspired by the observations that the velocity dispersions in the Solar neighbourhood are not the same towards the Galactic centre and out of the plane, they investigated under what conditions a second isolating integral beyond the orbital energy E could exist in a 2-dimensional Hamiltonian with application to the Galaxy. Their tool was the Poincaré section, although they seem to have re-invented it, and did not refer to it by name. The method, re-stated in the language of action-angle variables, is simple. If at fixed energy, the consequents of a Poincaré section lie on a closed curve without self-intersection, then the area enclosed yields a well-defined action and demonstrates the existence of an additional isolating quasi-integral. Hénon and Heiles discovered that the existence of a second isolating integral was strongly dependent on E. They also observed and coined the term *chain of islands* to describe the projections of librational near-resonant tori that they saw.

Important work on the hunt for third isolating integrals in Galactic Hamiltonians was also carried out by Contopoulos, starting with (Contopoulos, 1960). At first, these were series solutions without a formal convergence proof; later, the question of convergence was addressed more closely, and the concept of an isolating integral was properly defined (Contopoulos, 1963). This work culminated in an analysis of ergodic and periodic orbits in barred Galactic potentials (Contopoulos, 1983), which revealed the complexity of their morphology.

The observation that elliptical galaxies exhibited triaxiality and hence admitted a third

---

[5] The approach to these two problems is obviously different. To demonstrate non-integrability, it is enough to find numerically one orbital energy at which regular orbits do not exist. Demonstrations of integrability require a formal existence proof.

isolating (quasi)-integral (Binney, 1978) inspired another important development. Schwarzschild (1979) integrated individual orbits in a triaxial potential consistent with a chosen real-space density over times long compared to the age of the Universe. Most orbits were found to have three 'effective' isolating integrals, i.e. quantities that were either isolating integrals in a formal sense, or approximated integrals over the integration time. It was then shown that the density distributions produced by the chosen orbits were consistent with the global density function.

The class of *Stäckel potentials* attracted considerable interest (e.g. de Zeeuw et al., 1986), because a) mathematically, they were separable and yielded an additional integral, and b) potentials of relevance to the Galaxy fit this form. On the other hand, Stäckel potentials were not convenient in the action-angle representation: obtaining both actions and angles for these systems was highly numerical work. Additionally, studies of resonant boxlets in a triaxial potential (Binney and Spergel, 1984) revealed that the dynamical picture could be more involved than the Stäckel systems had led one to believe.

The last twenty-five years have seen significant advances in resonant dynamics in the context of non-linear theory. Much of this work concerns the ever more complicated mechanisms of transition to chaos as the dimensionality of a system is increased, and it has limited application to physics. For further reading on near-integrable dynamics in applications of physical relevance, Lichtenberg and Lieberman (1991) and Arnold (1989) are recommended. For a contemporary text on developments in celestial mechanics that goes considerably beyond Arnold, I highly recommend Morbidelli (2002).

Resonant dynamics never achieved the success in Galactic applications that it had in celestial mechanics. The Solar System is filled with prominent examples of resonances, from the Kirkwood gaps in the asteroid belt, whose origins are known to be ongoing dynamical effects, to the Cassini and Encke divisions in Saturn's rings. The surprising harmonies that Kepler had reported in *Harmonices mundi* were in fact resonance conditions, and even planet-moon systems are dominated by resonant dynamics.[6] One might say that the analogy in Galactic dynamics would be to select a star and determine unequivocally from its phase space location whether it is resonantly trapped. We are a considerable distance from this goal.

The prospective launch of the GAIA satellite provides a new impetus. GAIA will furnish *velocity phase space*[7] data for 1% of the Galactic stellar population, so that the dynamical fine structure of the Galaxy should be visible. In mapping the techniques of celestial mechanics to the Galaxy, however, we must be aware of a number of issues. The Solar System is much older than the Galaxy in dynamical times, its resonances are narrow and well-defined, and the Hamiltonian is well-known. The double degeneracy of the Kepler potential means that the Solar System is a non-generic laboratory in which mean-motion resonances (e.g. three-body resonances) are significant. The Galactic potential, by contrast, is said to be largely generated by dark matter and is poorly defined. A successful approach must therefore involve *adaptive dynamics*. This concept (Binney, 2005) is elaborated in Chapter 4.

---

[6] A very interesting paper on resonances in the Earth-Moon system is Touma, J. 2000, *The Phase Space Adventure of the Earth and Moon*, in *The Origin of the Earth and Moon* (Canup, R.M., Righter, K., ed.), University of Arizona Press, Tucson.

[7] For configurations Q and momenta P, we distinguish *velocity phase space* $(Q, \dot{Q})$ from (symplectic) *phase space* (P, Q). Of course, velocity phase space can also be symplectic, e.g. when $Q \in \mathbb{R}^n$.

1.1 Outline of the thesis

In this thesis, we take steps towards detecting resonant structures in phase space and putting adaptive dynamics into practice. In the work of Kaasalainen (1994), it was shown that perturbation theory can be very useful when actions are generated by the so-called *torus machine*. We look at the uses and limitations of perturbation theory when only a simple set of spherical actions and angles is available.

In numerical work, perturbation theory provides a simple and well-studied approach (e.g. Morbidelli, 2002) to generating approximations to the invariant actions of a quasi-integrable system away from resonance. Generating functions provide one method of obtaining approximately invariant actions (Binney and Tremaine, forthcoming). We derive generating functions for a canonical map to approximately invariant actions by a simple technique involving only calculus, and we extend the method to finding the frequencies of the perturbed system.

Given numerical integrations of orbits, we find that the actions of a system can be approximated by averaging a set of canonically mapped actions over the angles conjugate to the original actions. We find explicit expressions for the errors in this approach for one- and two-dimensional systems for a general class of canonical maps, and show that in n-dimensional systems, the error is $O(|s|^2)$, for s the deviation of the generating function for the map from the generator of the identity map. We test the technique in a one-dimensional example, integrating the action of a Kepler potential over the angles of the parent family of isochrone orbits. We also consider a two-dimensional example for a Hamiltonian admitting a low-order resonance, and compare the effectiveness of the averaging technique with numerical integrations of orbits in the Poincaré section.

Given a potential, the first task is to identify the important resonances. A low-order resonance is often, but not necessarily dynamically important in the sense of capturing particles or creating an absence of particles. The importance of a resonance depends on the details of the time-evolution of the orbital frequencies in its vicinity. The behaviour of the frequencies near resonance is quite difficult to predict analytically because of the problem of small divisors, and one must resort to numerical techniques. For systems in two and three degrees of freedom, the relative importance of the resonances is easily revealed in a *Laskar frequency map.* A grid of initial conditions is chosen and the particles are integrated for some time in the potential. The final state frequencies are evaluated and the frequency ratios plotted. The frequencies can be determined by spectral analysis (Binney and Spergel, 1982), i.e. by Fourier transforming the orbits $Q(t)$ and investigating the most prominent lines of the frequency spectra. In the final chapter we return to Laskar frequency maps to investigate the resonances of a triaxial potential of some astrophysical relevance.

Having identified a key resonance of the chosen potential, we then consider applications of the averaging technique, focusing on the phenomenon of resonant trapping. We consider a simple spherical potential and populate a representative sample of its tori. We introduce axisymmetry adiabatically until some time at which the system permits a low-order resonance. We examine the orbits in two surfaces of section and in perturbed frequency. We apply the averaging technique introduced earlier to approximate invariant actions for untrapped orbits. We are able to observe the growth of the resonance over time in averaged action, and we can conclude the probability of capture into resonance. We propose a theoretical prediction for the capture probability from a straightforward macroscopic analysis, and we find theory and practice in agreement.

Now we turn to perturbation theory. We launch orbits in the zero-velocity torus and model their capture into resonance with first-order perturbation theory. We find that perturbation theory performs well in radial action away from pericentre and poorly when one action is significantly smaller than another. We also consider the effect of *aliasing* and quantify the errors resulting from the discretization of the generating function necessary for a computational implementation. We use generating functions to obtain a map to the resonant action-angle plane and are able to observe pendulum dynamics in our numerical simulation of resonant trapping.

Now we investigate adaptive dynamics. We vary the initial orbital energy and examine the orbits in the averaged action and the perturbed action planes. We determine the feasibility of identifying orbits as trapped or untrapped by looking for a cluster of orbits in the perturbed action-angle plane. Observing that these trapped orbits lie on what is very nearly a line in the averaged action plane, we use the scatter around a linear regression fit to perturbed actions as a diagnostic of the accuracy of a trial Hamiltonian. As a further diagnostic test, we use the scatter in perturbed frequency ratio around resonance for the trapped orbits. We find that the two diagnostic tests are sufficiently complementary to allow the trial Hamiltonian to be constrained in the chosen parameter plane.

The geometry of orbits in a symplectic space is simple; the map from velocity phase space to that space for a general quasi-integrable Hamiltonian, however, is not. We therefore test the null hypothesis that resonant substructures can be found through a traditional technique of velocity phase space density estimation. Here we apply the purportedly non-parametric *kernel method* (Silverman, 1986) to a velocity phase space distribution generated by a spherical potential and find that we are able to recover only the zero-order behaviour of the distribution function. We also demonstrate that the theoretically optimal *smoothing length* in the *fixed kernel method* (effectively the radius of the kernel) is far from the smoothing length found in the more evolved *adaptive kernel method*, in which the kernels are allowed to vary in size to compensate for a low- or high-density region. The optimal *sensitivity parameter*, a measure of the extent to which the kernels are allowed to vary in size with the density, is also found to be different from what might be expected. Adaptive kernels perform just as badly as fixed kernels in the sense that derivatives of the distribution function with respect to kinematic quantities are pure noise. We show that a recovery of the potential requires knowledge of these derivatives, and that we therefore are unable to reconstruct global features of the system, much less probe resonant structures. Some results from the numerical analysis community reveal that larger data sets are unlikely to significantly improve the results, and that there will never be a substantially better method of selecting the smoothing parameter. Most other data-based techniques involve linear maps and are therefore unsuited. In short, a successful approach to resonant phenomena must involve dynamical models. We therefore turn to Hamiltonian dynamics.

1.2 Overview of integrable dynamics

The phase space of an n-dimensional dynamical system with coordinates in a configuration n-manifold $Q \in \mathbb{Q}$ is the $2n$-dimensional space of momenta and configurations in its cotangent bundle $(P, Q) \in T^*\mathbb{Q}$.[8] Every initial condition $(P(t_0), Q(t_0))$ generates a unique

---

[8] The Lagrangian view is in many respects more intuitive: velocity phase space is just the tangent bundle to the configuration manifold $(Q, \dot{Q}) \in T\mathbb{Q}$. Unfortunately the Euler-Lagrange equation is clumsier than Hamilton's equations, and frequencies and hence resonant dynamics are more naturally expressed in phase

orbit by $(P(t_0), Q(t_0), H(P,Q)) \rightarrow (P(t), Q(t))$, the Hamiltonian $H(P,Q)$ providing the potential $\Phi(Q)$ to integrate the orbit. Specifically, $P(t), Q(t)$ are the solution to Hamilton's coupled differential equations $\dot{P}_i = [P_i, H]$, $\dot{Q}_i = [Q_i, H]$, $i = 1, \ldots, n$, where the Poisson bracket in a representation of phase space (P,Q) is[9] $[A,B] = \frac{\partial A}{\partial Q} \cdot \frac{\partial B}{\partial P} - \frac{\partial A}{\partial P} \cdot \frac{\partial B}{\partial Q}$. Then, $\dot{P} = -\frac{\partial H}{\partial Q}, \dot{Q} = \frac{\partial H}{\partial P}$. These are $2n$ first-order differential equations in $2n$ unknowns; given $2n$ initial conditions $P(t_0), Q(t_0)$, they can be solved.

An *integrable* system is one with n isolating *integrals of motion*[10] $J_j \in \mathbb{R}, j = 1, \ldots, n$, fulfilling the role of generalized momenta and satisfying $[J_j, H] = 0$ and $[J_i, J_j] = 0$; $i, j = 1, \ldots, n$. Functions whose Poisson bracket evaluate to zero are said to be *in involution.* Defining a conjugate set of coordinates θ, and choosing to evaluate the Poisson brackets in (J, θ), note that $[J_j, H] = -\frac{\partial H}{\partial \theta_j} = 0$. Then the Hamiltonian can be written $H = H(J)$. For such a system, Liouville showed that $\theta \in \mathbb{T}^n$, i.e. that θ are *angles* and thus define the n-torus.[11] Hamilton's equations tell us that $\dot{J}_i = [J_i, H] = 0$ and $\dot{\theta}_i = [\theta_i, H] = \frac{\partial H}{\partial J_i} = \omega_i(J)$. Hamilton's equations have been decoupled and their solution is $J(t) = J(t_0)$, $\theta(t) = \theta(t_0) + \omega(J)t$. Since θ are angles, $\omega(J)$ are correctly interpreted as frequencies. The J are called *actions* and the space of J is called *action space.* The n actions $J(t_0)$ are integrals and uniquely specify the orbit and its energy through $E = H(J)$. Action space is foliated in orbital n-tori. Because H maps a torus to itself, the tori are called *invariant.* The invariant n-tori are said to carry *conditionally periodic* orbital motion,

---

space. $T\mathbb{Q}$ maps to its dual space $T^*\mathbb{Q}$ by the Legendre transformation of the Lagrangian, $\mathcal{L}$: $P = \frac{\partial \mathcal{L}}{\partial Q}$.

[9] In the following, if $v = (v_1, \ldots, v_n)$, $\frac{\partial}{\partial v} \equiv \nabla_v = \left(\frac{\partial}{\partial v_1}, \ldots, \frac{\partial}{\partial v_n}\right)$.

[10] Systems with more than n and as many as $2n - 1$ isolating integrals are called *super-integrable.* These can be treated with integrable dynamics as described here, the super-integrability manifesting itself in degeneracies in the orbital frequencies. Spherical potentials $\Phi(r)$ generate super-integrable Hamiltonians. One representation of their integrals would be (L_x, L_y, L_z, E), for E orbital energy and L angular momentum.

[11] See Arnold (1989), Ch. 10 for the proof.

because the frequencies ω are constant in the tori. In those tori for which the orbital frequencies are incommensurable (rationally independent), the system is ergodic[12] and so the orbit must densely cover the n-torus: this is called *quasi-periodic* motion. For those J for which (some of) the $\omega(J)$ are commensurable, the motion is *(partially) periodic* and densely covers a lower-dimensional torus. The parent n-tori are said to be *resonant.* The set of points in action space corresponding to resonant tori is a Cantor set, but the resonant tori are dense in action space as long as the Hamiltonian satisfies a non-degeneracy condition.

Each resonant torus has frequencies satisfying m independent relations, $(1 \leq m \leq n-1)$ of the form $a^i \cdot \omega = 0$, for non-zero integer n-vectors $a^i \in \mathbb{Z}^n - \{0\}$, $i = 1, \ldots, m$, and where a^i is assumed to be irreducible: integer multiples of a^i are said to constitute the *resonant module.* The value m is the *multiplicity* of the resonance: resonances of multiplicity m have quasi-periodic orbits in $(n-m)$-tori, since it is possible to write m of their angles as functions of the $n - m$ other angles using the relationships $a^i \cdot \omega = 0$. These lower-dimensional tori form families belonging to the resonant n-torus. The *order* of the resonance is defined as the quantity $\min_{1\leq i\leq m} \sum_{j=1}^{n} |a_j^i|$, the subscript j denoting the components of a^i, the superscript i labelling the vector as before. As the order of the resonance increases, it fills a larger region of the n-torus: orbits in higher order resonant tori are increasingly indistinguishable from quasi-periodic motion in non-resonant tori.

In summary, when H is integrable, a map exists from momentum-configuration space to action-angle space $(P, Q) \overset{M}{\mapsto} (J, \theta)$, $T^*\mathbb{Q} \overset{M}{\mapsto} \mathbb{R}^n \times \mathbb{T}^n$, such that the Hamiltonian $H'(P, Q) = H(J)$. The orbits are diffeomorphic to n-tori away from a resonance condition and to $(n-m)$-tori at a resonance of multiplicity m, $a^i \cdot \omega(J) = 0$, $i = 1, \ldots, m$.

---

[12] For the proof, see Arnold (1989) or Binney & Tremaine (1987) on the time-averages theorem.

The obvious question is how to construct M. Arnold (1963) showed that when the actions are obtained by $J_i = \frac{1}{2\pi}\oint_{\gamma_i} P \cdot dQ$, for γ_i, $i = 1, \ldots, n$ a set of n topologically distinct closed paths on the orbital surface not deformable to a point, J_i takes just n non-zero values. Since $\theta \in \mathbb{T}^n$, Arnold's observation has a simple geometric interpretation. The n topologically distinct closed paths in an n-torus generate the n actions; closed paths η that can be deformed continuously to a point produce $\oint_\eta P \cdot dQ = 0$. Because of this property, the orbital tori are referred to as *null surfaces.* One set of labels of the non-trivial orbital families γ_i would be $\theta_i \in [0, 2\pi], \theta_j = \text{constant}, \forall j \neq i$. Any continuous deformation of γ_i leaves J_i unchanged.

Analytically, the angles θ are found by defining a generating function $S(J, Q)$ such that $(\theta, P) = \nabla_{J,Q} S$ and solving the resulting Hamilton-Jacobi equation $E = \frac{1}{2}(\nabla_Q S)^2 + \Phi(Q)$, with Φ the potential, for S. The solution is $S(J, Q) = \int P(J, Q; \Phi) \cdot dQ$. The use of the generating function to obtain θ guarantees that the map $(P, Q) \mapsto (J, \theta)$ is canonical. Unfortunately, in practice, obtaining this map analytically tends to be difficult except for the simplest potentials.

For $n = 2$, the orbits are best represented in a *Poincaré map* to a suitably chosen 2-submanifold of $T^*\mathbb{Q}$, called a *Poincaré section.* Points in the section are called *consequents* and record where the orbit punctures, for instance, the (P_i, Q_i) plane with $Q_j = 0$ and $\dot{Q}_j > 0$ in some interval Δt. In this view, the non-resonant orbital tori are closed curves for $\Delta t \to \infty$. Resonant tori generate a finite number of consequents in the surface of section: a resonant torus with m consequents in its (P_1, Q_1) section and n consequents in its (P_2, Q_2) section has a frequency ratio $\frac{\omega_1}{\omega_2} = \frac{m}{n}$. Except precisely at a resonance condition, all the orbits form a single family of n-tori.

2 Approximating invariant actions for quasi-integrable Hamiltonians

In this chapter we introduce the tools of resonant dynamics, by considering the fundamental problem in quasi-integrable systems: the determination of invariant actions. We analyse quasi-integrable systems and derive generating functions for perturbation theory by simple calculus. We also obtain the orbital frequencies of the perturbed system by this method. We propose a method for approximating invariant actions from numerical integrations of orbits by averaging a set of mapped actions over time. We derive expressions for the errors involved in this technique for one and two-dimensional systems and for the order of the error in n-dimensional systems, and we explore a simple one-dimensional example. We also compare the effectiveness of the averaging technique to performing numerical line integrals around curves inferred from consequents in Poincaré sections.

2.1 Quasi-integrable dynamics

The dynamics of quasi-integrable Hamiltonian systems

$$H(J,\theta) = H_0(J) + \epsilon H_1(J,\theta) \tag{1}$$

is considerably more complicated than that of integrable ones. KAM theory states (Morbidelli, 2002) that all sufficiently non-resonant orbital tori of H_0 persist in H for $\epsilon < \epsilon_{\mathrm{crit}}(\omega_0(J_{\mathrm{KAM}}))$, where a sufficient distance from resonance is expressible in a Diophantine condition of the form

$$|a \cdot \omega_0(J_{\mathrm{KAM}})| > \frac{\gamma}{|a|^\tau}, \; \forall a \in \mathbb{Z}^n - \{0\}, \; \tau > n-1. \tag{2}$$

The proof is constructive; Eq. 2 is a sufficient rather than a necessary condition for a *KAM torus*. The allowed actions J_{KAM} have $\dot{J}_{\mathrm{KAM}} \neq 0$ in a KAM orbit, but there exists a map to new actions and angles $(J, \theta) \mapsto (I, \phi)$, such that the new Hamiltonian $\hat{H}(I, \phi)$ admits *one* exact solution, $I = 0$, $\phi \in \mathbb{T}^n$ corresponding to the invariant KAM torus.

KAM theory provides invariant actions for a given torus. At the opposite extreme, (J, θ) are global action-angle coordinates, but the Hamiltonian $H(J, \theta)$ has angle-dependent terms to $O(\epsilon)$. A compromise is provided by perturbation theory, which can be used to decrease the order of the angle dependence, at the cost of being invalid at a dense set of resonant locations. Here, we derive the generating function for perturbation theory using only simple matrix calculus. Derivations by other methods include Binney and Tremaine (forthcoming).

2.1.1 A simple derivation of generating functions for perturbation theory

The map M introduced earlier generates the canonical transformation $(P, Q) \overset{M}{\mapsto} (J, \theta)$, such that $\dot{J} = 0$ for the Hamiltonian $H_0(J)$. Now we would like to find a canonical map $(J, \theta) \overset{P}{\mapsto} (J', \theta'), \mathbb{R}^n \times \mathbb{T}^n \overset{P}{\mapsto} \mathbb{R}^n \times \mathbb{T}^n$ such that for the quasi-integrable Hamiltonian

$$H(J, \theta) = H_0(J) + \epsilon^k H_1(J, \theta) \tag{3}$$

$$H(J, \theta) = G(J', \theta') = G_0(J') + \epsilon^{2k} G_2(J', \theta'). \tag{4}$$

As $k \to \infty$, we might expect to be left with a domain of action space foliated in invariant tori. For $A = (A_1, \ldots, A_n)$, $B = (B_1, \ldots, B_n)$, etc., define

$$\left(\frac{\partial A}{\partial B}\right)_C \equiv \begin{bmatrix} \left(\frac{\partial A_1}{\partial B_1}\right)_{B_2,\cdots,B_n,C_1,\cdots,C_n} & \cdots\cdots\cdots\cdots \\ \vdots & \vdots \\ \cdots\cdots\cdots\cdots & \left(\frac{\partial A_n}{\partial B_n}\right)_{B_1,\cdots,B_{n-1},C_1,\cdots,C_n} \end{bmatrix} . \tag{5}$$

Then

$$\begin{aligned} \left(\frac{\partial G}{\partial \theta'}\right)_{J'} &= \epsilon^{2k} \left(\frac{\partial G_2}{\partial \theta'}\right)_{J'} \\ &= \left(\frac{\partial J}{\partial \theta'}\right)_{J'} \frac{\partial H_0}{\partial J} + \epsilon^k \left(\frac{\partial J}{\partial \theta'}\right)_{J'} \left(\frac{\partial H_1}{\partial J}\right)_{\theta} + \epsilon^k \left(\frac{\partial \theta}{\partial \theta'}\right)_{J'} \left(\frac{\partial H_1}{\partial \theta}\right)_J . \end{aligned} \tag{6}$$

Every canonical map P is equivalently defined by generating functions $S(J, \theta')$ and $T(J', \theta)$. Choosing self-consistent generating functions of this form to be within $O(\epsilon^k)$ of the identity map $(J, \theta) \mapsto (J', \theta') : S(J, \theta') = J \cdot \theta'$,

$$S(J, \theta') = J \cdot \theta' + \epsilon^k s(J, \theta') \Rightarrow J'(J, \theta') = J + \epsilon^k \left(\frac{\partial s}{\partial \theta'}\right)_J \tag{7}$$

$$T(J', \theta) = J' \cdot \theta + \epsilon^k t(J', \theta) \Rightarrow \theta'(J', \theta) = \theta + \epsilon^k \left(\frac{\partial t}{\partial J'}\right)_{\theta} . \tag{8}$$

For δ the identity (Kronecker delta) matrix, we can write[13]

---

[13] Notice that the relationship between s and t is given by $\left(\frac{\partial t}{\partial J'}\right)_{\theta} = -\left(\frac{\partial s}{\partial J}\right)_{\theta'}$ and $\left(\frac{\partial t}{\partial \theta}\right)_{J'} = -\left(\frac{\partial s}{\partial \theta'}\right)_J$. Notice also that equations like Eq. 6 can be found by considering the expansion of Poisson brackets in the two phase-space representations (J, θ) and (J', θ'):

$$[J'_i, G] = [J'_i, H] = -\left(\frac{\partial G}{\partial \theta'_i}\right)_{J'} = \left(\frac{\partial J'_i}{\partial \theta}\right)_J \cdot \left(\frac{\partial H}{\partial J}\right)_{\theta} - \left(\frac{\partial J'_i}{\partial J}\right)_{\theta} \cdot \left(\frac{\partial H}{\partial \theta}\right)_J . \tag{9}$$

$$\left(\frac{\partial J}{\partial \theta'}\right)_{J'} = -\left(\frac{\partial J}{\partial J'}\right)_{\theta'}\left(\frac{\partial J'}{\partial \theta'}\right)_{J} = -\epsilon^k \left(\frac{\partial J}{\partial J'}\right)_{\theta'}\left(\frac{\partial^2 s}{\partial \theta'^2}\right)_{J,J}, \tag{10}$$

$$\begin{aligned}
\left(\frac{\partial J}{\partial J'}\right)_{\theta'} &= \left[\left(\frac{\partial J'}{\partial J}\right)_{\theta'}\right]^{-1} \\
&= \left[\delta + \epsilon^k \left(\frac{\partial^2 s}{\partial \theta' \partial J}\right)_{J,\theta'}\right]^{-1} \\
&= \delta - \epsilon^k \left(\frac{\partial^2 s}{\partial \theta' \partial J}\right)_{J,\theta'} + O(\epsilon^{2k}).
\end{aligned} \tag{11}$$

Similarly,

$$\begin{aligned}
\left(\frac{\partial \theta}{\partial \theta'}\right)_{J'} &= \left[\left(\frac{\partial \theta'}{\partial \theta}\right)_{J'}\right]^{-1} \\
&= \left[\delta + \epsilon^k \left(\frac{\partial^2 t}{\partial J' \partial \theta}\right)_{\theta,J'}\right]^{-1} \\
&= \delta - \epsilon^k \left(\frac{\partial^2 t}{\partial J' \partial \theta}\right)_{\theta,J'} + O(\epsilon^{2k})
\end{aligned} \tag{12}$$

$$\begin{aligned}
\therefore\ \epsilon^{2k}\left(\frac{\partial G_2}{\partial \theta'}\right)_{J'} &= -\epsilon^k \left[\delta - \epsilon^k \left(\frac{\partial^2 s}{\partial \theta' \partial J}\right)_{J,\theta'}\right]\left(\frac{\partial^2 s}{\partial \theta'^2}\right)_{J,J}\left[\omega_0(J) + \epsilon^k \left(\frac{\partial H_1}{\partial J}\right)_{\theta}\right] \\
&\quad + \epsilon^k \left(\frac{\partial H_1}{\partial \theta}\right)_{J} - \epsilon^{2k}\left(\frac{\partial^2 t}{\partial J' \partial \theta}\right)_{\theta,J'}\left(\frac{\partial H_1}{\partial \theta}\right)_{J} + O(\epsilon^{3k})
\end{aligned} \tag{13}$$

This is turned into Eq. 6 by evaluating the matrix M of Poisson brackets $M_{ij} = [J'_i, \theta_j]$ to reveal that $M = -\left(\frac{\partial J'}{\partial J}\right)_{\theta} = -\left(\frac{\partial \theta}{\partial \theta'}\right)^T_{J'}$ and similarly for $N_{ij} = [J'_i, J_j]$, which yields $N = \left(\frac{\partial J'}{\partial \theta}\right)_{J} = -\left(\frac{\partial J}{\partial \theta'}\right)^T_{J'}$.

for $\omega_0(J) = \frac{\partial H_0}{\partial J}$. This equality is satisfied when terms of $O(\epsilon^k)$ vanish. This provides the condition

$$\left(\frac{\partial H_1}{\partial \theta}\right)_J = \left(\frac{\partial^2 s}{\partial \theta'^2}\right)_{J,J} \omega_0(J). \tag{14}$$

To proceed to a useful expression for s, we expand $H_1(J,\theta)$ and $s(J,\theta')$ in Fourier series, omitting the $a = 0$ index, since this produces an integrable term:

$$s(J,\theta') = i \sum_{a \in \mathbb{Z}^n - \{0\}} s_a(J) e^{ia\cdot\theta'} \tag{15}$$

$$H_1(J,\theta) = \sum_{a \in \mathbb{Z}^n - \{0\}} h_a(J) e^{ia\cdot\theta}, \tag{16}$$

such that

$$\sum_{a \neq 0} \left[a h_a(J) e^{ia\cdot\theta} + a a^T s_a(J) \omega_0(J) e^{ia\cdot\theta'} \right] = 0. \tag{17}$$

Setting the summands to zero individually,

$$s_a(J) = -\frac{h_a(J)}{a \cdot \omega_0(J)} + O(\epsilon^k), \tag{18}$$

where the error arises from the replacement of $e^{ia\cdot\theta'}$ by $e^{ia\cdot\theta}$ in Eq. 17. Substituting in the expression for $s(J,\theta')$,

$$s(J,\theta') = -i \sum_{a \neq 0} \frac{h_a(J)}{a \cdot \omega_0(J)} e^{ia\cdot\theta'} + O(\epsilon^k). \tag{19}$$

Then

$$S(J,\theta') = J \cdot \theta' - i\epsilon^k \sum_{a \neq 0} \frac{h_a(J)}{a \cdot \omega_0(J)} e^{ia\cdot\theta'} + O(\epsilon^{2k}). \tag{20}$$

The corresponding Hamiltonian, with angle-dependent terms only to $O(\epsilon^{2k})$ is known as the *Birkhoff normal form.* We know how to generate (J, θ) for a Hamiltonian differing from integrability by $O(\epsilon)$, and we have just shown how to generate a map from a Hamiltonian with an angular dependence in $O(\epsilon^k)$ to one in $O(\epsilon^{2k})$. Why is this not simultaneously a proof by induction of the existence of KAM tori, i.e. of the map $(J, \theta) \overset{P^\infty}{\mapsto} (I, \phi)$? The problem is that Cantor set of resonant $\omega_0(J)$ through which we pass as we vary the frequency. At these locations $a \cdot \omega_0(J) = 0$ the generating function is singular, and as we iterate the above method to $O(\epsilon^{4k}), O(\epsilon^{8k}), \ldots$, the problem worsens. The consensus is (e.g. Saha, 1989) that iterations of perturbation theory are not in general terribly useful in applications. As $k \to \infty$, orbital KAM tori can still exist if $\epsilon < \epsilon_{\text{crit}}$. Closer to resonance than the region in which KAM tori are permitted by the Diophantine condition (Eq. 2) lie the *circulation tori* and the *libration tori.* We will return to these in the context of resonant capture (Chapter 3).

The substitution of $e^{ia\cdot\theta'}$ by $e^{ia\cdot\theta}$ in Eq. 19 is insignificant in physical applications, because the amplitude of the resulting error is $O(\epsilon^{2k})$ in S, but it remains unfortunate that generating functions do not in fact necessarily produce a formal solution[14] $J'(J, \theta)$, $\theta'(J, \theta)$.

---

[14] Finding *generating Hamiltonians* is a common technique for producing maps to perturbed action-angle space (e.g. Morbidelli 2002). (J', θ', H') are related to (J, θ, H) by *Lie series*, i.e. Taylor series with time derivatives $d^i f/dt^i$ replaced by nested Poisson brackets $[[\ldots[f, \chi]\ldots, \chi], \chi]$, for χ the generating Hamiltonian. The $O(\epsilon)$ terms in the Lie series for H' are made to equal zero. With the resulting constraint, the generating Hamiltonian is found to take the form $\chi(J', \theta') = -i\sum_a \frac{h_a(J')}{a\cdot\omega_0(J')} e^{ia\cdot\theta'}$. Because we have no knowledge of $\omega_0(J')$, the Lie series method is of no advantage over the method of generating functions.

2.1.2 Derivation of the perturbed frequencies

Now we obtain the perturbed frequencies. Defining $\omega'(J') \equiv \frac{\partial G_0}{\partial J'}$ and $\omega(J,\theta) \equiv \left(\frac{\partial H}{\partial J}\right)_\theta$, and with $\omega_0(J) \equiv \frac{\partial H_0}{\partial J}$ as usual,

$$\left(\frac{\partial G}{\partial J'}\right)_{\theta'} = \omega'(J') + \epsilon^{2k}\left(\frac{\partial G_2}{\partial J'}\right)_{\theta'}. \tag{21}$$

Proceeding as before, with δ the identity matrix,

$$\begin{aligned}\left(\frac{\partial G}{\partial J'}\right)_{\theta'} &= \left(\frac{\partial J}{\partial J'}\right)_{\theta'}\omega(J,\theta) + \epsilon^k\left(\frac{\partial \theta}{\partial J'}\right)_{\theta'}\left(\frac{\partial H_1}{\partial \theta}\right)_J \\ &= \left[\delta - \epsilon^k\left(\frac{\partial^2 s}{\partial\theta'\partial J}\right)_{J,\theta'} + O(\epsilon^{2k})\right]\omega(J,\theta) \\ &\quad -\epsilon^{2k}\left[\delta - \epsilon^k\left(\frac{\partial^2 t}{\partial J'\partial\theta}\right)_{\theta,J'} + O(\epsilon^{2k})\right]\left(\frac{\partial^2 t}{\partial J'^2}\right)_{\theta,\theta}\left(\frac{\partial H_1}{\partial\theta}\right)_J.\end{aligned} \tag{22}$$

Comparing orders,

$$\omega'(J') = \left[\delta - \epsilon^k\left(\frac{\partial^2 s}{\partial\theta'\partial J}\right)_{J,\theta'}\right]\omega(J,\theta), \tag{23}$$

and substituting in for the generating function, with the unavoidable approximation inherent in the generating function method that the Fourier expansion can be treated as an expansion in θ',

$$\omega'(J') \approx \left[\delta - \epsilon^k\sum_a \frac{1}{a\cdot\omega_0(J)}\left[\frac{\partial h_a}{\partial J} + \frac{h_a}{a\cdot\omega_0(J)}\frac{\partial\omega_0}{\partial J}\right]a^T e^{ia\cdot\theta}\right]\omega(J,\theta) \equiv \Lambda(J,\theta)\omega(J,\theta). \tag{24}$$

We can find $\omega(J,\theta)$ explicitly from the Hamiltonian:

$$H(P,Q) = \frac{1}{2}\left[P_r^2 + \left(\frac{P_\vartheta}{r}\right)^2 + \left(\frac{L_z}{r\sin\vartheta}\right)^2\right] + \Phi(Q), \tag{25}$$

such that

$$\begin{aligned}(\omega_r', \omega_\theta') \approx \Lambda\nabla_J H &= \Lambda\left[P_r\,\nabla_J P_r + P_\vartheta\,\nabla_J P_\vartheta\frac{1}{r^2} - \left(P_\vartheta^2 + \frac{L_z^2}{\sin^2\vartheta}\right)\frac{1}{r^3}\nabla_J r\right. \\ &- \left.\frac{L_z^2\cot\vartheta}{r^2\sin^2\vartheta}\nabla_J\vartheta - F_r\nabla_J r - rF_\vartheta\nabla_J\vartheta\right],\end{aligned} \tag{26}$$

where the forces $-\nabla\Phi = F_r\hat{e}_r + F_\vartheta\hat{e}_\vartheta$, and the necessary derivatives are found analytically from S. In practice, away from resonance, it will be sufficient to assume $\Lambda = \delta$.

2.2 Approximating invariant actions from numerical integrations of orbits

2.2.1 Averaging actions over mapped angles

In how far can one determine invariant actions for a non-resonant torus I by averaging a set of available actions J over angles ϕ conjugate to I? Consider a $2d$-dimensional phase space, and assume the existence of a canonical map $(J,\theta) \overset{S}{\mapsto} (I,\phi)$ provided by a generating function $S(I,\theta) = I\cdot\theta + \sum_{n\in\mathbb{Z}^d-\{0\}} s_n(I)e^{in\cdot\theta}$, with $s_n = \bar{s}_{-n}$, for $\bar{\ }$ the complex conjugate, such that the imaginary component of S, $\Im[S] = 0$. We want an expression for the deviation $\langle J\rangle_t - I$ for $\langle J\rangle_t$ the time-average of J over the orbit $\{I(t) = \text{const}, \phi(t)\}$. In the following, we denote $\sum_{n\in\mathbb{Z}^d-\{0\}}$ by $\sum_{n\neq 0}$. First consider this problem in one dimension. The orbit in a non-resonant torus is ergodic and we can apply the time-averages theorem:

$$\langle J\rangle_t = \frac{1}{2\pi}\int_0^{2\pi} J\,d\phi = \frac{1}{2\pi}\int_\alpha^\beta J\left(\frac{\partial\phi}{\partial\theta}\right)_I d\theta, \tag{27}$$

observing that $\phi(I = \text{const}, \theta) \rightarrow d\phi = \left(\frac{\partial\phi}{\partial\theta}\right)_I d\theta$. From S,

$$\left(\frac{\partial\phi}{\partial\theta}\right)_I = 1 + \sum_{n\neq0}\frac{ds_n}{dI}ine^{in\theta}, \tag{28}$$

$$J = I + \sum_{n\neq0} s_n(I)ine^{in\theta} \tag{29}$$

$$\therefore\ \langle J\rangle_t = \frac{1}{2\pi}\int_{\phi=0}^{\phi=2\pi} J\,d\theta + \frac{1}{2\pi}\int_\alpha^\beta\left(I + \sum_{n\neq0}s_n(I)ine^{in\theta}\right)\sum_{m\neq0}\frac{ds_m}{dI}ime^{im\theta}\,d\theta. \tag{30}$$

Since $\phi = \theta + \sum_{n\neq0}\frac{ds_n}{dI}e^{in\theta}$, the boundaries of integration are given by

$$2\pi = \beta - \alpha + \sum_{n\neq0}\frac{ds_n}{dI}(e^{in\beta} - e^{in\alpha}), \tag{31}$$

yielding the unique solution $\beta - \alpha = 2\pi$. Substituting Eq. 29 into the first integral of Eq. 30, we see this evaluates to I.[15] The second integral consists of two summands, the first of which will integrate to zero. The second summand produces a non-zero term for $n = -m$. Putting all this together,

$$\langle J\rangle_t = I + \sum_{n\neq0} n^2 s_n(I)\frac{ds_{-n}}{dI}. \tag{33}$$

---

[15] In fact, in one dimension, the first integral is equivalent to I for *all* canonical maps, by the conservation of Poincaré invariants:

$$2\pi I = \int_\zeta I d\phi = \int_S dI\wedge d\phi = \int_S dJ\wedge d\theta = \int_\zeta Jd\theta \tag{32}$$

for $\zeta : \phi \in [0, 2\pi]$ and S the surface bounded by ζ. The first integral of Eq. 30 can therefore be rewritten as $\frac{1}{2\pi}\int_{\phi=0}^{\phi=2\pi} J\,d\theta = I$.

For $s(\theta) \equiv \sum_{n\neq 0} s_n e^{in\theta}$,

$$\begin{aligned} s_n &= \frac{1}{2\pi}\int_0^{2\pi} s(\theta)e^{-in\theta}\,d\theta & (34)\\ \therefore\ \bar{s}_{-n} &= \frac{1}{2\pi}\int_0^{2\pi} \bar{s}(\theta)e^{-in\theta}\,d\theta. & (35)\end{aligned}$$

To guarantee $\Im[s(\theta)] = 0$, we require $s_n = \bar{s}_{-n}$. Rewriting our expression for $\langle J\rangle_t$,

$$\begin{aligned} \langle J\rangle_t &= I + \sum_{n>0} n^2\left[s_n(I)\frac{ds_{-n}}{dI} + s_{-n}(I)\frac{ds_n}{dI}\right] \\ &= I + \sum_{n>0} n^2\left[s_n(I)\frac{d\bar{s}_n}{dI} + \bar{s}_n(I)\frac{ds_n}{dI}\right] \\ &= I + \sum_{n>0} n^2\frac{d}{dI}\left|s_n\right|^2. & (36)\end{aligned}$$

The condition $\langle J\rangle_t = I$ is therefore only satisfied in the trivial case of the identity map, $s_n = 0, \forall n$. Notice, however, that as long as the series converges, if $J = I + O(\epsilon)$, then $\langle J\rangle_t = I + O(\epsilon^2)$. This confirms that we can use $\langle J\rangle_t$ as an estimator of I.

In two dimensions, the change of angle variables requires the Jacobian $\frac{\partial(\phi)}{\partial(\theta)}_I \equiv \left|\left(\frac{\partial\phi}{\partial\theta}\right)_I\right|$, and we have

$$\langle J_a\rangle_t = \frac{1}{4\pi^2}\int_{[0,2\pi]^2} J_a \frac{\partial(\phi)}{\partial(\theta)}_I\, d^2\theta \qquad (37)$$

for $a = 1, 2$. Noting that

$$\begin{aligned} \left(\frac{\partial\phi_i}{\partial\theta_j}\right)_I &= \delta_{ij} + \sum_{n\neq 0}\frac{\partial s_n}{\partial I_i} i n_j e^{in\cdot\theta}, & (38)\\ \frac{\partial(\phi)}{\partial(\theta)}_I &= \frac{\partial\phi_1}{\partial\theta_1}\frac{\partial\phi_2}{\partial\theta_2} - \frac{\partial\phi_1}{\partial\theta_2}\frac{\partial\phi_2}{\partial\theta_1}\end{aligned}$$

$$
\begin{aligned}
&= \left(1+\sum_{n\neq 0}\frac{\partial s_n}{\partial I_1} i n_1 e^{in\cdot\theta}\right)\left(1+\sum_{m\neq 0}\frac{\partial s_m}{\partial I_2} i m_2 e^{im\cdot\theta}\right) \\
&\quad -\sum_{p\neq 0}\frac{\partial s_p}{\partial I_1} i p_2 e^{ip\cdot\theta}\sum_{q\neq 0}\frac{\partial s_q}{\partial I_2} i q_1 e^{iq\cdot\theta}. \qquad (39)
\end{aligned}
$$

Proceeding analogously to the one-dimensional case, we find that

$$
\begin{aligned}
\langle J_a\rangle_t = I_a &+ \sum_{n\neq 0} n_1 n_a s_n(I)\frac{\partial s_{-n}}{\partial I_1} \\
&+ \sum_{n\neq 0} n_2 n_a s_n(I)\frac{\partial s_{-n}}{\partial I_2} \\
&+ \sum_{n\neq 0, m\neq 0} i n_a(m_1 n_2 - m_2 n_1) s_n(I)\frac{\partial s_m}{\partial I_1}\frac{\partial s_{-n-m}}{\partial I_2} \qquad (40) \\
\therefore \langle J_1\rangle_t = I_1 &+ \sum_{\mathcal{D}} n_1^2\left[s_n(I)\frac{\partial s_{-n}}{\partial I_1} + s_{-n}(I)\frac{\partial s_n}{\partial I_1}\right] \\
&+ \sum_{\mathcal{D}} n_1 n_2\left[s_n(I)\frac{\partial s_{-n}}{\partial I_2} + s_{-n}(I)\frac{\partial s_n}{\partial I_2}\right] + O(s^3), \qquad (41)
\end{aligned}
$$

with $\mathcal{D}$: $n_1 \in \mathbb{Z}_+$, $n_2 \in \mathbb{Z} - \{0\}$. Again, the condition $\Im[s(\theta)] = 0$ requires $s_n = \bar{s}_{-n}$. Ignoring third order terms,[16]

$$
\langle J_1\rangle_t \approx I_1 + \sum_{\mathcal{D}} n_1^2 \frac{\partial}{\partial I_1}|s_n|^2 + \sum_{\mathcal{D}} n_1 n_2 \frac{\partial}{\partial I_2}|s_n|^2. \qquad (42)
$$

Similarly,

$$
\langle J_2\rangle_t \approx I_2 + \sum_{\mathcal{D}} n_2^2 \frac{\partial}{\partial I_2}|s_n|^2 + \sum_{\mathcal{D}} n_1 n_2 \frac{\partial}{\partial I_1}|s_n|^2. \qquad (43)
$$

In d dimensions, obtaining an expression for the error in closed form is more challenging,

---

[16] Note that this result holds given our initial assumption of physically relevant generating functions that differ from the identity map by a separable function of I and θ. For a general canonical map, it is easy to construct a generating function for which averaged actions bear no relationship to true actions. An example is $S(I,\theta) = I_1\theta_2 + I_2\theta_1$, which generates the map $(J_1, J_2) \mapsto (J_2, J_1)$, so $\langle J_1\rangle_t = I_2$ and $\langle J_2\rangle_t = I_1$.

but it is straightforward to show that $\langle J\rangle_t - I \sim O(|s|^2)$ here also: in matrix notation,

$$\left(\frac{\partial\phi}{\partial\theta}\right)_I = \delta + \sum_{n\neq 0} \frac{\partial s_n}{\partial I} i n^T e^{in\cdot\theta}, \tag{44}$$

where the Kronecker delta matrix δ is obviously just the identity matrix. For a square matrix M and $\epsilon \ll 1$, $|\delta + \epsilon M| = 1 + \epsilon\mathrm{Tr}(M) + O(\epsilon^2)$, so[17]

$$\begin{aligned}
\langle J_a\rangle_t &= \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d} J_a \left[1 + \sum_{n\neq 0}\sum_{j=1}^{d} \frac{\partial s_n}{\partial I_j} i n_j e^{in\cdot\theta} + O(|s|^2)\right] \; d^d\theta \\
&= I_a + \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d} \left(I_a + \sum_{n\neq 0} s_n(I) i n_a e^{in\cdot\theta}\right) \left[\sum_{n\neq 0}\sum_{j=1}^{d} \frac{\partial s_n}{\partial I_j} i n_j e^{in\cdot\theta} + O(|s|^2)\right] \; d^d\theta \\
&= I_a + \sum_{n\neq 0}\sum_{j=1}^{d} n_j n_a s_n(I) \frac{\partial s_{-n}}{\partial I_j} + \frac{1}{(2\pi)^d}\int_{[0,2\pi]^d} \left(I_a + \sum_{n\neq 0} s_n(I) i n_a e^{in\cdot\theta}\right) O(|s|^2) \; d^d\theta \\
&= I_a + O(|s|^2).
\end{aligned} \tag{45}$$

2.2.2 A 1-dimensional example

Consider the following example in one dimension. For details of the derivations of actions and angles of the isochrone potential see the references listed in Appendix C, especially Kaasalainen (1994). A general isochrone potential

$$\Phi(r) = -\frac{1}{b + \sqrt{b^2 + r^2}} \tag{46}$$

---

[17] In the expansion of $|\delta + \epsilon M|$ one finds d terms of $O(\epsilon^2)$, so that a simple closed form for $\langle J_a\rangle_t - I_a$ seems unlikely.

can be treated as a perturbation of a Kepler potential. Let us investigate what we obtain when we integrate the Kepler action J_0 over a torus of an isochrone potential, i.e.

$$\langle J_0 \rangle_{\theta_b} = \frac{1}{2\pi} \int_0^{2\pi} J_0(\theta_b)\, d\theta_b, \tag{47}$$

for θ_b the angle of the isochrone potential. The Kepler case is recovered for $b = 0$. For E_b the orbital energy and L the angular momentum,

$$\begin{aligned} J_b &= (-2E_b)^{-\frac{1}{2}} - \frac{L}{2} - \sqrt{\frac{L^2}{4} + b} && (48) \\ J_0 &= (-2E_0)^{-\frac{1}{2}} - L, && (49) \end{aligned}$$

(see Appendix C and references therein), the first equation yielding

$$E_b = -\frac{1}{2}\left[J_b + \frac{L}{2} + \sqrt{\frac{L^2}{4} + b}\right]^{-2}. \tag{50}$$

The angle variable of a generic isochrone potential is slightly more challenging. For

$$\begin{aligned} a &= -\frac{1}{2E_b} - b && (51) \\ e &= \sqrt{1 - \frac{L^2}{a}\left(1 + \frac{b}{a}\right)}, && (52) \end{aligned}$$

the independent variable θ_b determines r somewhat circuitously through the quantity η, by the Kepler equation

$$\theta_b(\eta) = \eta - \frac{ae}{a+b}\sin\eta. \tag{53}$$

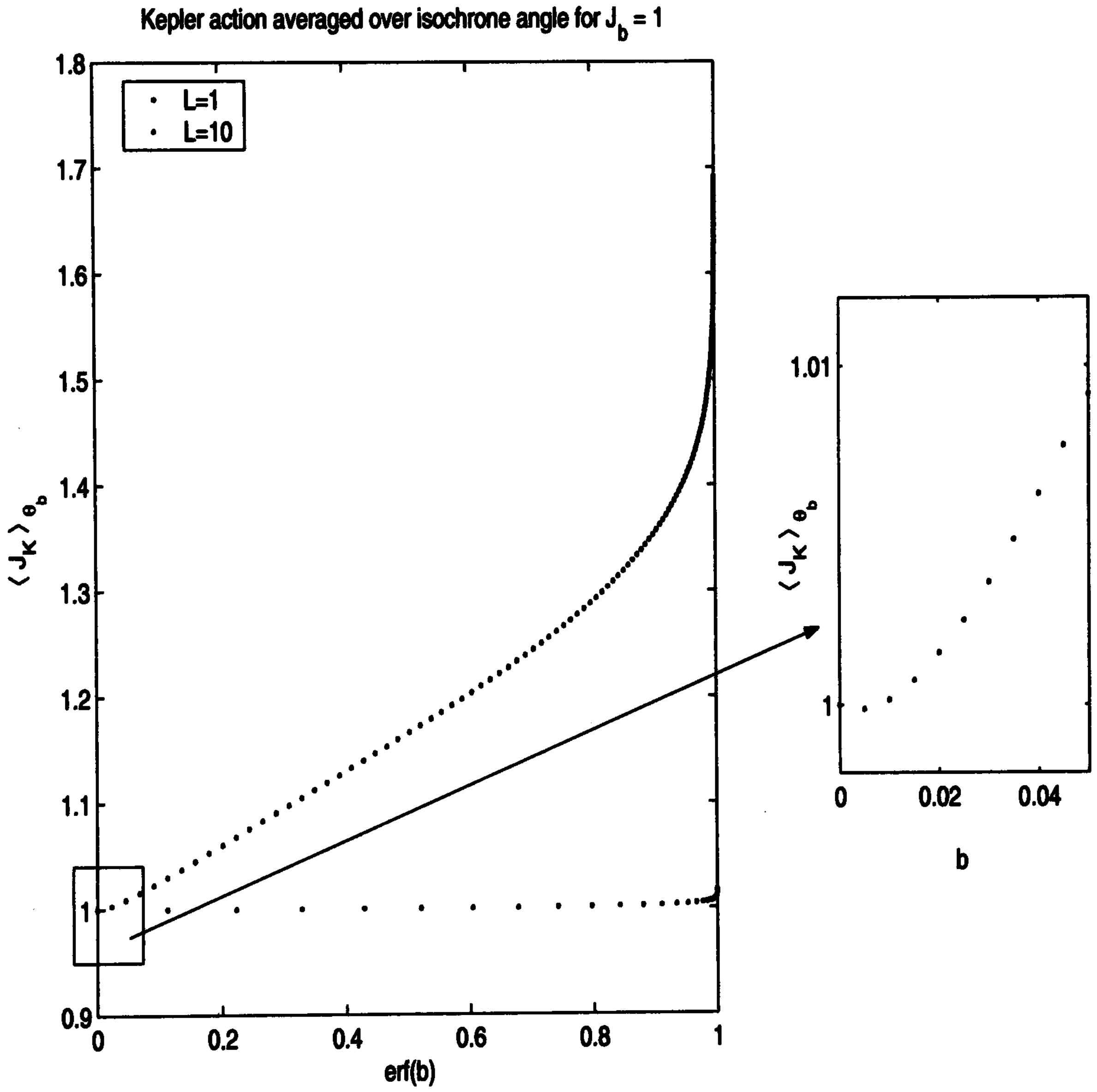

Figure 1: The Kepler action averaged over isochrone tori as a function of the isochrone parameter reveals a gradual deviation from the chosen J_b. As the eccentricity of the orbit is reduced by increasing L, the deviation becomes more gradual.

We solve for η, and with $u(\theta_b) = a(1 - e\cos\eta)$, r and P are given by[18]

$$\begin{aligned} r(\theta_b) &= \sqrt{u(u+2b)} && (54) \\ P(\theta_b) &= \left[(a+b)(u+2b)u\right]^{-1/2} ae\sin\eta. && (55) \end{aligned}$$

We have

$$\langle J_0 \rangle_{\theta_b} = \frac{1}{2\pi}\int_0^{2\pi} \left[\frac{2}{r(\theta_b)} - P(\theta_b)^2 - \left(\frac{L}{r(\theta_b)}\right)^2\right]^{-\frac{1}{2}} d\theta_b - L. \tag{56}$$

We can make no further analytical progress, but given a choice of J_b, L, and b, the numerical integration is straightforward, and a plot of averaged Kepler action over a range of isochrone tori is shown in Fig 1. The figure reveals a gradual deviation of averaged action from J_b, with $\langle J_0 \rangle_{\theta_{b\to\infty}} \to \infty$. Since $\langle J \rangle_\theta - I = O(|s|^2)$, we expect $\langle J \rangle_{\theta_b} - I = O\left(\left(\frac{d\Phi}{db}\delta b\right)^2\right)$. As $b \to 0$, $\langle J \rangle_{\theta_b} - I \to O(b^2)$. Then we should see $\frac{d}{db}\Big|_{b=0} \langle J \rangle_{\theta_b} = 0$. This is indeed seen to be the case in Fig. 1.

For the blue curve, $L = 1$ and $J_b = 1$, and the eccentricity of the Kepler orbit is $e = \sqrt{3}/2$, so the axis ratio is 2 and our chosen orbits are 'quite' non-circular in some physically relevant sense. Over a range in b of 0.5, the error in approximating J_b by $\langle J_0 \rangle_{\theta_b}$ is only 15%. For the red curve, $L = 10$ and the orbits are quite circular. The averaging technique is correspondingly successful. Averaging mapped actions over orbits is therefore looking increasingly promising as a technique of approximating the true actions on these orbits.

---

[18] Eq. 55 is derived as follows: $P = \dot{r} = (u+b)\left[u(u+2b)\right]^{-\frac{1}{2}} ae\sin\eta \frac{d\eta}{d\theta_b}\dot{\theta}_b$, where $\dot{\theta}_b = \omega_b = (-2E_b)^{3/2}$ and $\frac{d\eta}{d\theta_b} = \left[1 - \frac{ae}{a+b}\cos\eta\right]^{-1} = \frac{a+b}{u+b}$.

2.2.3 Areas in the Poincaré map compared with the averaging technique

The standard method of obtaining the true actions of a quasi-integrable Hamiltonian is calculating the numerical line integral $\oint P \cdot dQ$ around a curve inferred from the consequents of the Poincaré map. It can be shown by a simple geometric argument (e.g. Binney, Gerhard, and Hut 1985) that the map from the orbit to the surface of section is canonical. Then the area enclosed by a curve through the consequents (P_i, Q_i) with $Q_j = 0$ and $\dot{Q}_j > 0$ must equal the true action I_i.

We can study the relative effectiveness of the method of calculating areas in the surface of section and the method of averaging actions over mapped angles in the context of an example, to which we will return in Chapter 3. We populate ten representative orbital tori of a spherical system J^i, $i = 1, \ldots, 10$, with four initial conditions each: that is, we choose four conditions $\theta^{i,j} \in [0, 2\pi]^2$, $j = 1, \ldots, 4$. The orbits are integrated in time as axisymmetry is introduced into the potential. At some chosen time, the 'flattening' mechanism is switched off, and the orbits are integrated at fixed axis ratio. We use the action-angle notation (J, θ) to describe the non-invariant actions and their conjugate angles, and (I, ϕ) for the invariant actions and corresponding angles. Fig. 2 is a plot of true radial action I_r for orbits in the resulting two-dimensional quasi-integrable Hamiltonian as estimated by $\oint_\gamma P \cdot dQ$, for γ the curve through the consequents.[19] A plot of $\langle J_r \rangle_t$ is superimposed. The data are plotted in groups corresponding to the four initial angle conditions chosen for each torus.

This system will be studied in greater detail in the next chapter, but a few features can

---

[19] Since all the orbits are near the 1:1 resonance of this Hamiltonian, it is sufficient to define γ by the shortest line connecting the i-th consequent (P_r^i, r^i) to the $(i+1)$-th consequent (P_r^{i+1}, r^{i+1}), $\forall i$. In general, however, the temporal order of the consequents does not coincide with the spatial order in which we need them for the line integration. The points then have to be ordered computationally.

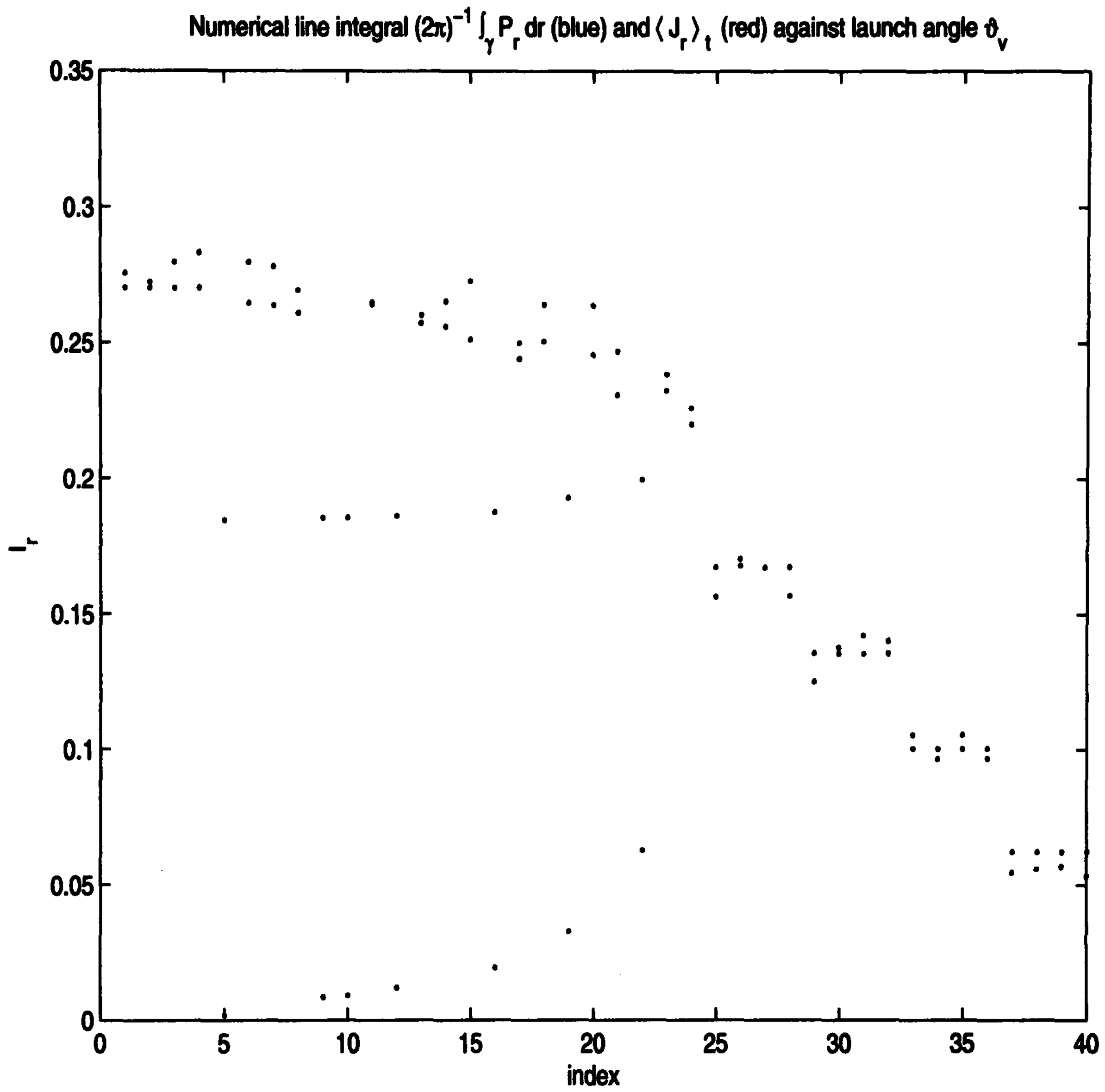

Figure 2: Estimators of true radial action I_r for orbits in a two-dimensional quasi-integrable Hamiltonian $H(J_r, J_\vartheta, \theta_r, \theta_\vartheta)$ admitting an important low-order resonance. Values of numerical line integrals around the curve generated by the consequents of the (P_r, r) Poincaré map are shown in blue. Averaged actions $\langle J_r \rangle_t$ are shown in red. Ten representative orbital tori are chosen, corresponding to the different velocity space spherical polar angles ϑ_v at which the particles are launched. We henceforth call ϑ_v the *launch angle*. Groups of four points correspond to randomly selected initial angles on each torus.

already be noted now. The fit is in general quite good, except for seven outlying points. The Hamiltonian admits an important resonance, and the outliers correspond to resonant orbits. The method of integrating areas in surfaces of section always produces the true actions I_r, so the seven blue points correspond to the true actions on these resonant orbits. The averaging technique cannot reproduce the true actions of resonant orbits, because the I_r form three distinct sets, depending on whether the orbital tori *circulate* (inside or outside the resonance) or *librate* around the resonance. This will also be explained in the next chapter. The fit is least accurate for non-resonant orbits in the vicinity of the resonance. This is because these orbits circulate very slowly, producing many consequents before completing a circuit.

It is striking how constant and reliable our method of averaging actions seems to be. We can see from the figure that the $O(|s|^2)$ error terms are not a concern. In fact, for non-resonant orbits, the averaging technique is seen to be as effective as numerical line integration.[20] This now provides us with a useful tool. Away from resonance, we can check the effectiveness of techniques such as perturbation theory by comparing perturbed actions to those found by averaging the spherical actions over orbits. This avoids the more laborious numerical line integrals over curves in Poincaré sections and is less prone to error. In fact averaged actions provide sufficient information to conclude resonant capture probabilities and make other observations on resonant dynamics, as we shall see.

---

[20] In systems of dimension three and higher, we no longer have invariant curves in Poincaré sections, and the averaging method provides the only simple route to the actions.

3 Capture into resonance and averaged actions

We use generating functions to reproduce the argument (Morbidelli, 2002) that the non-trivial dynamics of a resonance of multiplicity one in a region of action space containing only one principal resonance can be mapped into a phase plane. We generate the spectra and frequency maps for a simple toy Hamiltonian to illustrate the important resonances. Then we pick up the model studied in Section 2.2.3. We simulate resonant capture by launching particles in representative tori of a spherical Hamiltonian and slowly introducing axisymmetry. We plot Poincaré sections for these orbits and observe them in frequency ratio. Using the averaging technique of the previous chapter, we are able to observe the growth of the resonance. A simple macroscopic argument allows us to calculate the probability of capture into resonance theoretically and to confirm theory with practice.

3.1 The resonant action-angle plane

Consider a quasi-integrable Hamiltonian in a $2n$-dimensional angle-action space

$$H(J,\theta) = H_0(J) + \epsilon \sum_{a\in\mathbb{Z}^n-\{0\}} h_a(J)e^{ia\cdot\theta}. \tag{57}$$

In Chapter 2 we constructed the Birkhoff normal form

$$G(J',\theta') = G_0(J') + \epsilon^2 G_2(J',\theta') \tag{58}$$

by means of a canonical map $(J,\theta) \mapsto (J',\theta')$. Consider instead the condition

$$H(J,\theta) = G(J',\theta') = G_0(J') + \epsilon G_1(J',\theta') + \epsilon^2 G_2(J',\theta'), \tag{59}$$

for $G_1(J',\theta') = \sum_{a \in \mathcal{A}} g_a(J') e^{ia\cdot\theta'}$, $\mathcal{A} \subset \mathbb{Z}^n - \{0\}$. The analysis of Chapter 2 applies here, so that in place of Eq. 13, we have

$$\begin{aligned} \epsilon \left(\frac{\partial G_1}{\partial \theta'}\right)_{J'} + \epsilon^2 \left(\frac{\partial G_2}{\partial \theta'}\right)_{J'} &= -\epsilon \left[I - \epsilon \left(\frac{\partial^2 s}{\partial \theta' \partial J}\right)_{J,\theta'}\right] \left(\frac{\partial^2 s}{\partial \theta'^2}\right)_{J,J} \left[\omega_0(J) + \epsilon \left(\frac{\partial H_1}{\partial J}\right)_{\theta}\right] \\ &\quad + \epsilon \left(\frac{\partial H_1}{\partial \theta}\right)_{J} - \epsilon^2 \left(\frac{\partial^2 t}{\partial J' \partial \theta}\right)_{\theta,J'} \left(\frac{\partial H_1}{\partial \theta}\right)_{J} + O(\epsilon^3). \end{aligned} \tag{60}$$

Equating $O(\epsilon)$ terms,

$$\left(\frac{\partial G_1}{\partial \theta'}\right)_{J'} = \left(\frac{\partial H_1}{\partial \theta}\right)_{J} - \left(\frac{\partial^2 s}{\partial \theta'^2}\right)_{J,J} \omega_0(J), \tag{61}$$

and substituting $s(J,\theta') = i \sum_{a \in \mathbb{Z}^n - \{0\}} s_a(J) e^{ia\cdot\theta'}$ and $H_1(J,\theta) = \sum_{a \in \mathbb{Z}^n - \{0\}} h_a(J) e^{ia\cdot\theta}$ as before,

$$\sum_{a \in \mathcal{A}} i a g_a(J') e^{ia\cdot\theta'} = i \sum_{a \in \mathbb{Z}^n - \{0\}} \left[a h_a(J) e^{ia\cdot\theta} + a a^T s_a(J) \omega_0(J) e^{ia\cdot\theta'}\right]. \tag{62}$$

There is no simple choice of s to satisfy this equality. If we choose

$$\begin{aligned} s_b &= -\frac{h_b(J)}{b \cdot \omega_0(J)}, \quad b \in \mathbb{Z}^n - \mathcal{A} - \{0\} \\ s_b &= 0, \quad b \in \mathcal{A}, \end{aligned} \tag{63}$$

however, then the error in the satisfaction of Eq. 62 is

$$\sum_{a \in \mathcal{A}} \left[a g_a(J') e^{ia\cdot\theta'} - a h_a(J) e^{ia\cdot\theta} \right], \tag{64}$$

where $J = J' + O(\epsilon)$ and $\theta = \theta' + O(\epsilon)$. Now

$$G(J', \theta') = G_0(J') + \epsilon \sum_{a \in \mathcal{A}} g_a(J') e^{ia\cdot\theta'} + R(J', \theta'), \tag{65}$$

where the order of the remainder term R is complicated by the aforementioned approximation.

Consider J in a region $\mathcal{U}$ characterized by one principal resonance of multiplicity one, $a \cdot \omega_0(J^*) = 0$, $a \in \mathbb{Z}^n$. Returning to our generating functions, let the resonant module of a up to order K be $\mathcal{A}$, defined by

$$\mathcal{A} = \{a : a \in \mathbb{Z}^n - \{0\},\ |a| < K,\ a \cdot \omega_0(J^*) = 0,\ J^* \in \mathcal{U}\}. \tag{66}$$

For this choice of $\mathcal{A}$ the Hamiltonian G is said to be in *resonant normal form*. Let the canonical map $(J', \theta') \overset{S}{\mapsto} (K, \psi)$, $\mathbb{R}^n \times \mathbb{T}^n \mapsto \mathbb{R}^n \times \mathbb{T}^n$ be defined by the generating function $S(K, \theta') = K^T U \theta'$, for $U \in \mathbb{Z}^{n \times n}$. Then

$$\begin{aligned} \psi &= \frac{\partial S}{\partial K} = U\theta' && (67) \\ J'^T &= \frac{\partial S}{\partial \theta'} = K^T U && (68) \\ \therefore K &= (U^T)^{-1} J'. && (69) \end{aligned}$$

When $\det(U) = 1$ and the first row of U is given by the vector[21] a, then $\psi_1 = a \cdot \theta'$ is called the *resonant angle.* Substituting into (Eq. 65),

$$G(U^T K, U^{-1}\psi) = G_0(U^T K) + \epsilon \sum_{b \in \mathbb{Z}-\{0\}} g_{b\cdot a}(U^T K) e^{ib\psi_1} + R(U^T K, U^{-1}\psi). \tag{70}$$

Ignoring the remainder term and re-defining H, H_0, and H_1 such that $H(K,\psi) = G(U^T K, U^{-1}\psi)$, etc., we now have the integrable Hamiltonian

$$H(K,\psi) \approx H_0(K) + H_1(K,\psi_1), \tag{71}$$

for which, by Hamilton's equations, $\dot{K}_i = 0$, $i = 2,\ldots,n$. In this Hamiltonian, all non-trivial dynamics therefore take place in the (K_1, ψ_1) plane. In summary, resonant dynamics in a region of phase space dominated by one resonance of multiplicity one can be projected into a single action-angle plane.

Following Morbidelli (2002), we expand H around the exact resonance condition $K^* = (U^T)^{-1} J'^*$,

$$H \approx \alpha(K^*)(K_1 - K_1^*) + \frac{1}{2}\beta(K^*)(K_1 - K_1^*)^2 + \gamma(K^*)\cos\psi_1. \tag{72}$$

Since $\dot{\psi}_1(K_1) = \alpha + \beta(K_1 - K_1^*)$, at resonance, $\dot{\psi}_1(K_1^*) = \alpha$, but by definition of ψ, $\dot{\psi}_1(K_1^*) = 0$, so $\alpha = 0$. The form of H is seen to be equivalent to the Hamiltonian of the pendulum, and we expect pendulum dynamics in the resonant plane. Fig. 3 illustrates this idealized pendulum dynamics (also see Morbidelli, 2002) for a variety of tori $H = E_{1,2,\ldots}$. We

---

[21] See Morbidelli (2002) and references therein for an alternative derivation of the resonant normal form in terms of Lie series and generating Hamiltonians, and also for the proof that U subject to these conditions always exists. Morbidelli concludes that the size of R depends on secondary resonances in U.

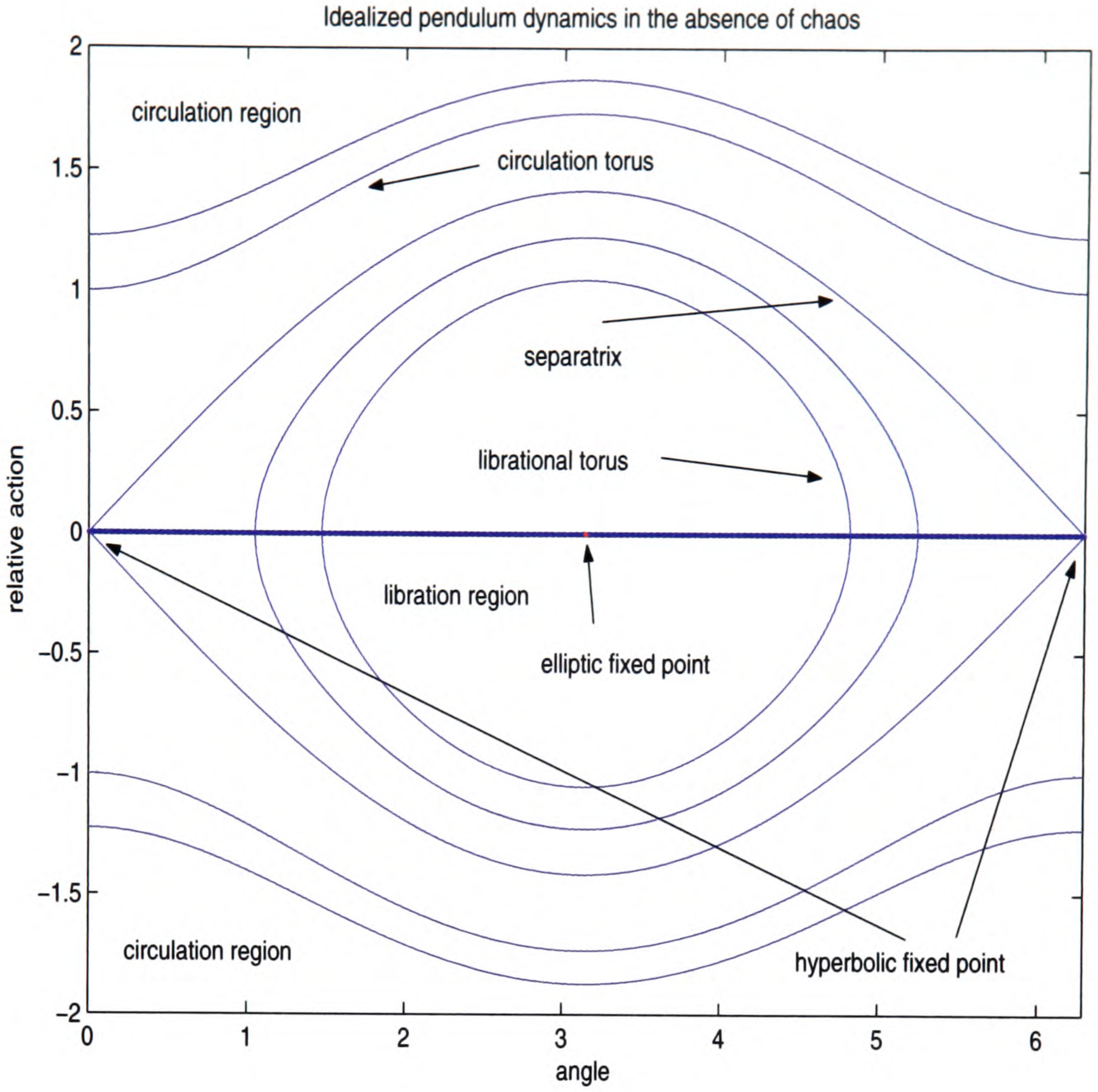

Figure 3: Idealized pendulum dynamics in the absence of separatrix splitting. The KAM picture is valid as one tends out into the circulation regions on either side of the resonance. The thick central line is the exact resonance condition. Relative action $K_1 - K_1^*$ is graphed against ψ_1.

notice immediately that the tori fall into three groups. Those that librate around the exact resonance condition (specifically, around the elliptic fixed point of the Hamiltonian, as is easily verified by taking second derivates of H) are called *libration tori.* Tori that circulate outside the immediate vicinity of the resonance are called *circulation tori.* These are KAM tori if their distance from the resonance satisfies the Diophantine condition (Eq. 2). The non-smooth curve separating the libration region from the circulation regions above and below the resonance is the *separatrix.* It is defined by an upper and a lower manifold that meet at the hyperbolic fixed point of the Hamiltonian. As long as the remainder term $R(K, \psi)$ is negligible, this is the entire dynamical picture of a multiplicity one resonance, and thus of all resonances in two-dimensional systems.[22]

True actions can of course be defined through the usual line integral for all tori, whether librational or circulating. It is readily seen from the diagram that the areas enclosed by isoenergetic curves, and so the actions, form three distinct sets. We can use our averaging method for all circulation tori; attempting to perform $\langle J_i \rangle_t$ for librational tori will give values for the actions close to those of a hypothetical circulation torus at the exact resonance condition.

So far we have considered static Hamiltonians, but what happens when time-variation is introduced in the potential? The law of *adiabatic invariance* tells us that if the time-variation is sufficiently slow compared with dynamical times, the actions remain constant and all the above results still apply. This result is proved in Binney & Tremaine (1987) and Arnold (1989).

---

[22] The phenomenon of *separatrix splitting*, in principle for all $\epsilon > 0$, introduces chaos to the region near the separatrix. It was observed by Poincaré that if the separatrix intersects itself in a point, it must intersect in infinitely many points, looping chaotically on itself. This has a very elegant and simple explanation. The separatrix can be considered to consist of a stable and an unstable manifold, M_+ and M_- which converge at the hyperbolic fixed point. Consider an orbit that contains an additional point of intersection of these two manifolds (called a *homoclinic point* because M_+ and M_- are part of the same separatrix.) Then all points on the orbit must lie in $M_+ \cap M_-$. Since the orbit must tend to the hyperbolic fixed point like all orbits on the separatrix, the images it leaves on the Poincaré section cannot coincide, and there must be an infinite number of homoclinic points. Additionally, 1) as the orbit slows, the separation between homoclinic points is reduced, and 2) incompressibility of Hamiltonian flow requires that the volume (i.e. area on the section) between the stable and unstable manifolds be conserved. Imagine two curves intersecting transversally. Since the direction of motion on M_+ is towards and the motion on M_- is away from the hyperbolic fixed point, the only way for the next homoclinic intersection to lie in the future of both M_+ and M_- is for one of the manifolds to loop back on itself. Because of 1), for subsequent homoclinic points, the distance between the loops is shrinking, and because of 2) the length of the loops is increasing. The result is that the loops start to produce higher-order intersections with $M_\pm$, creating infinitely long and infinitely convoluted stable and unstable manifolds, a topology known as the *homoclinic tangle*.

3.2 Simulations of resonant capture

The first step in simulating resonant capture is to determine which resonances are important in the chosen potential. In axisymmetric and triaxial potentials, one way to proceed is to construct a Laskar frequency map for the potential. Particles from a grid of initial conditions are integrated in the potential and their final state frequencies are plotted in the frequency ratio plane. These frequencies can be determined by spectral analysis of the orbits (Binney and Spergel, 1982). The method is summarized below.

3.2.1 Spectral dynamics and the Laskar frequency map

Consider a representation of the coordinates of a dynamical system $(Q_1, \ldots, Q_n)$. Whether angles or Euclidean quantities, the range of $Q(t)$ is bounded for a bound orbit, so it must be possible to write

$$Q_1(J, \theta) = \sum_{a \in \mathbb{Z}^n} A'_a(J) e^{ia\cdot\theta}. \tag{73}$$

For regular orbits, $\theta(t) = \theta(0) + \omega t$ and $\dot{J} = 0$, so

$$\begin{aligned} Q_1(J,t) &= \sum_a A_a(J) e^{ia\cdot\omega t} && (74) \\ Q_2(J,t) &= \sum_a B_a(J) e^{ia\cdot\omega t} && (75) \\ &\vdots \end{aligned}$$

Specifically, if we choose a representation of actions corresponding to these coordinates $J = (J_{Q_1}, \ldots, J_{Q_n})$ then we normally expect the principal terms in these series to be

$$\begin{aligned} Q_1(J,t) &\approx A_{100\ldots}(J)e^{i\omega_{Q_1}t} && (76) \\ Q_2(J,t) &\approx B_{010\ldots}(J)e^{i\omega_{Q_2}t} && (77) \\ &\vdots \end{aligned}$$

Consider a generic coordinate Q and record $Q(J,t_b)$ for $t_b = b\delta t$, $b = 0, 1, \ldots, N$, and some chosen time-step δt. The *spectrum* of Q is defined to be

$$\tilde{Q}(\omega) = \delta t \sum_{b=0}^{N} Q(t_b)e^{i\omega t_b} \equiv \mathcal{Q}(\omega)e^{i\phi_Q(\omega)}. \tag{78}$$

Sharp peaks or *lines* in the spectrum $|\tilde{Q}|$ indicate pure sinusoidal components in $Q(t)$. As $N \to \infty$ these of course become delta functions. For $T = \frac{1}{2}(N+1)\delta t$, the contribution of a harmonic term $Q(t) = A\cos(\omega_0 t + \phi)$ to its (positive-frequency) spectrum[23] is

$$\tilde{Q}(\omega) = A\frac{\sin(\omega - \omega_0)T}{\sin(\omega - \omega_0)}e^{i\left[(\omega-\omega_0)\frac{T-1}{2\delta t}-\phi\right]} \tag{79}$$

(Binney and Spergel, 1982), so one can fit $\mathcal{Q}(\omega) = |\tilde{Q}(\omega)|$ and $\phi_Q(\omega) = \arg[\tilde{Q}(\omega)]$ to find A, ω_0, and ϕ. This procedure is repeated for $\tilde{Q}(\omega^m)$, with $\omega^m = m\delta\omega = \frac{2\pi m}{N\delta t}$, for $m = 1, 2, \ldots, N$. This choice of ω ensures that $\mathcal{Q}(\omega^m)$ due to a sinusoidal component at ω_0

---

[23] Note that the continuous Fourier transform $\mathcal{F}[\cos(\omega_0 t + \phi)] = \pi\left[e^{i\phi}\delta(\omega - \omega_0) + e^{-i\phi}\delta(\omega + \omega_0)\right]$, so that each harmonic term strictly speaking contributes two lines to the spectrum, one with a positive and the other with a negative frequency. This has implications for the spectral analysis, because a line at $\omega_0 = \alpha \cdot (\omega_{Q_1}, \omega_{Q_2}, \omega_{Q_3})$ can be misinterpreted as $-\alpha \cdot (\omega_{Q_1}, \omega_{Q_2}, \omega_{Q_3})$. See, for instance, footnote 26.

falls monotonically around ω_0.

Consider a specific example. The isochrone potential is given by

$$\Phi'(r) = -\frac{1}{b + \sqrt{b^2 + r^2}} \tag{80}$$

and its density distribution can be found from Poisson's equation, $\rho(r) = \frac{1}{4\pi}\nabla^2\Phi'(r)$. The axisymmetric, or 'flattened' isochrone potential $\Phi(R, z; q)$ with axis ratio $q \in (0, 1]$ is generated from the density distribution $\frac{1}{q}\rho(R, \frac{z}{q})$ and can be obtained numerically by the *method of homoeoids*[24] (Binney & Tremaine, 1987). In Fig. 4 and Fig. 5 we see the spectra of $R(t)$, $x(t)$ and $z(t)$ for the flattened isochrone potential. There are three prominent peaks in $|\tilde{R}(\omega)|$. From Eq. 78 we expect the largest amplitude peak at some $m = p$ to give us $\omega^p = \omega_R$. The other two peaks occur at ω^{2p} and ω^{3p}. In the representation $\omega^m = \alpha \cdot (\omega_R, \omega_\vartheta, \omega_\varphi)$, therefore, these three peaks correspond to $\alpha = (1, 0, 0)$, $(2, 0, 0)$, $(3, 0, 0)$. This is in accordance with expectation: $|\tilde{R}(\omega)|$ should be completely dominated by multiples of the epicyclic frequency, $\omega^p \equiv \kappa$. In $|\tilde{x}(\omega)|$ the leading line is Ω, the rotational frequency of the epicentre. It is expected to be a good representation of ω_φ, while in $|\tilde{z}(\omega)|$ the leading line is normally ω_ϑ. The other lines are found to be linear combinations of the fundamental frequencies ω_R, ω_ϑ, ω_φ as shown.[25] Notice that $|\tilde{x}(\omega)|$ contains both $(1, 0, 1)$

---

[24] Homoeoids are shells in real space bounded by two similar concentric ellipsoids, in which isodensity surfaces are also concentric with and similar to the bounding ellipsoids. Then the potential in the interior of a homoeoid can be shown to be constant, and the potential of the system can be calculated by summing over homoeoids. Equipotential surfaces outside the homoeoid are ellipsoids that are confocal with the homoeoid. See Binney & Tremaine (1987) for details.

[25] Better estimates of the frequencies can be obtained by looking at the second difference of the spectrum, $\tilde{Q}''(\omega^m) = \tilde{Q}(\omega^{m+1}) + \tilde{Q}(\omega^{m-1}) - 2\tilde{Q}(\omega^m)$. Note from Eq. 79 that near ω_0, $\tilde{Q} \propto (\omega - \omega_0)^{-1}$. It turns out that $\tilde{Q}'' \propto (\omega - \omega_0)^{-3}$ (Binney and Tremaine, forthcoming). The lines of the second difference spectrum are thus considerably sharper than those of the spectrum. We use this second difference scheme to produce the Laskar frequency map.

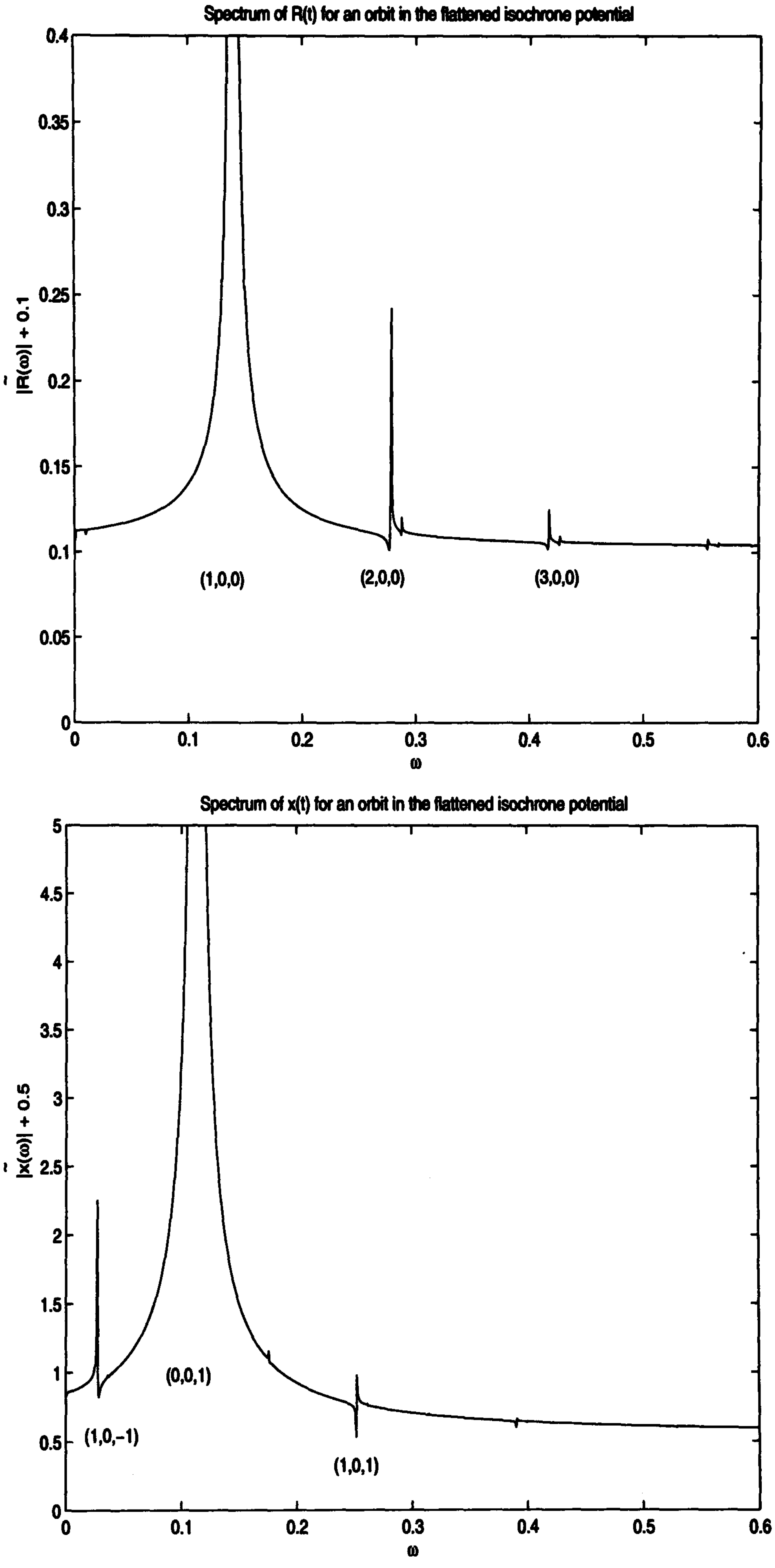

Figure 4: The spectra of cylindrical radius $R(t)$ and $x(t)$. The points of the discrete spectra $|\tilde{R}(\omega^m)|$, $|\tilde{x}(\omega^m)|$ are connected to form the curves depicted. The prominent peaks are labelled by the vector α for which $\omega^m = \alpha \cdot (\omega_R, \omega_\vartheta, \omega_\varphi)$. The spectrum of spherical radius $|\tilde{r}(\omega)|$ is essentially indistinguishable from $|\tilde{R}(\omega)|$ for this orbit.

and $(1,0,-1)$, i.e. $\kappa+\Omega$ and $\kappa-\Omega$. In the lower plot of Fig. 5 we superpose $|\tilde{R}(\omega)|$ and $|\tilde{z}(\omega)|$. Notice that the $\alpha=(1,0,0)$ and $(0,1,0)$ lines are very close together compared to their width: our orbit is very nearly in the $\omega_R=\omega_\vartheta$ resonance.

To produce a Laskar frequency map we thus need the locations of the principal line in each of the R, x, and z spectra for a grid of initial conditions. The result is shown in Fig. 6. We are able to identify the $a=(1,-1)$, $(3,-4)$, and $(2,-3)$ resonances in $a\cdot(\omega_R,\omega_\vartheta)=0$ as lines of collimated points on the diagram.[26] By far the most conspicuous resonance is the $(1,-1)$ line, which stretches the length of the diagram. Notice that for an axisymmetric potential, all resonances can be represented as lines through the origin in the frequency ratio plane. Equivalently, intersections between resonant lines are forbidden in a two-dimensional system, because if such an intersection existed, then at the point of intersection $a^1\cdot\omega=a^2\cdot\omega=0$, and the resonance would have multiplicity two (see Chapter 1).

The spherical isochrone potential has

$$\omega_\vartheta=\omega_\varphi=\frac{1}{2}\left(1+\frac{L}{\sqrt{L^2+4b}}\right)\omega_r, \tag{81}$$

where we can define ω_r and ω_R to be interchangeable for most orbits in this potential. As q, the axis ratio of the mass distribution, is reduced, the frequency of the vertical oscillations ω_ϑ becomes larger than the rotational frequency of the epicentre, ω_φ. Then the lower cut-

---

[26] Some care is needed in the spectral analysis. For some orbits near, but not in the $(1,-1)$ resonance, the $|\tilde{z}(\omega)|$ spectrum is dominated not by ω_ϑ, but by the barely resolved $|\omega_\vartheta-\omega_R|$ seen in Fig. 5. It turns out to be $\omega_R-\omega_\vartheta$, as can be confirmed, for instance, from the spectrum of $\vartheta(t)$. Because $|\tilde{\vartheta}(\omega)|$ can sometimes produce a very prominent peak at $2\omega_\vartheta$, in effect the construction of this Laskar map required four spectra: $\tilde{x}$, $\tilde{R}$, $\tilde{z}$, and $\tilde{\vartheta}$.

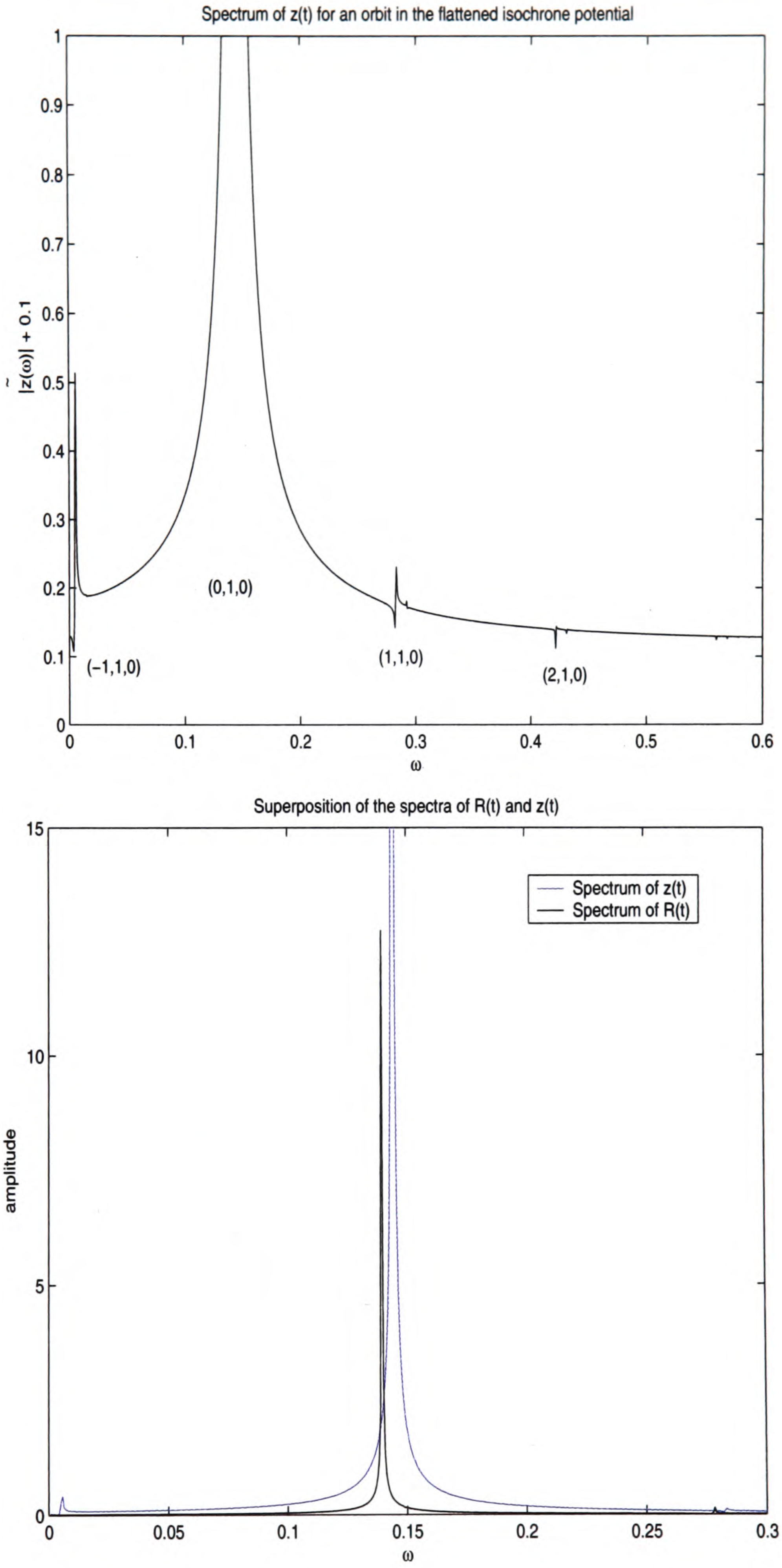

Figure 5: The spectrum of $z(t)$ (above) and a superposition of the spectra $|\tilde{R}(\omega)|$ and $|\tilde{z}(\omega)|$ (below). The leading lines of these two spectra are close compared to their widths, indicating the proximity of the orbit to a 1:1:0 resonance in $(\omega_R, \omega_\vartheta, \omega_\varphi)$. The scale of the plot renders the other lines invisible.

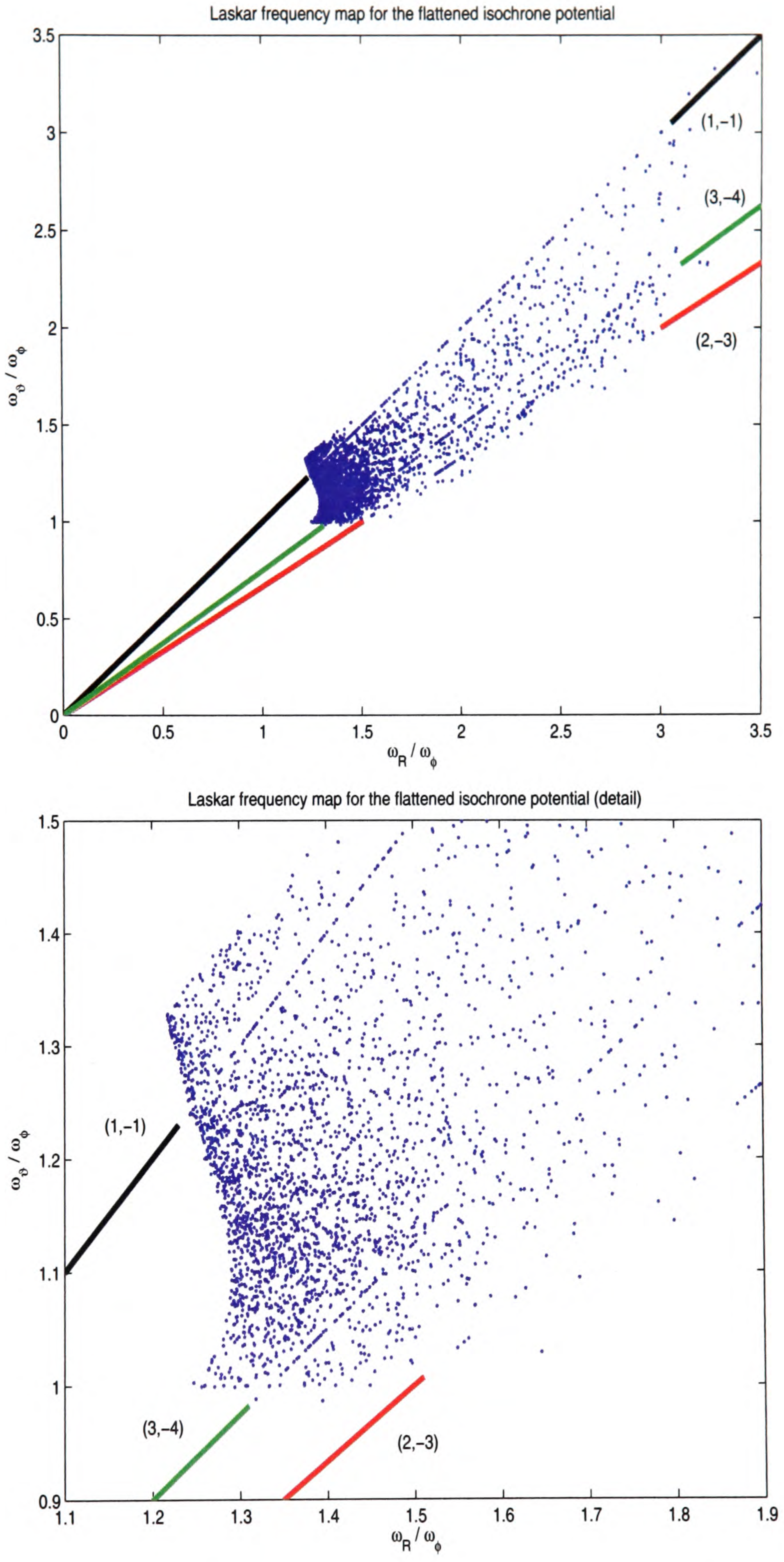

Figure 6: The Laskar frequency map for the flattened isochrone potential. The observed resonances appear as lines of points and are extended in the plots for clarity. The notation (a_1, a_2) denotes the resonance $a_1\omega_R + a_2\omega_\vartheta = 0$. At the chosen axis ratio, the $(1, -1)$ resonance has just appeared. The map is constructed for $E = -0.135$.

off at $\frac{\omega_\vartheta}{\omega_\varphi} = 1$ with the condition $\frac{\omega_R}{\omega_\varphi} \in (1,2)$ corresponds to the allowed frequencies of the spherical system $q = 1$. All resonances up to order seven are either a) on this diagram, b) lie below this cut-off line (most notably $(1,-2)$) and are not permitted, or c) lie above the $(1,-1)$ line and have not yet appeared at this axis ratio.

3.2.2 Simulating resonant trapping: populating the orbital tori

Having identified the $(1,-1)$ resonance of the flattened isochrone potential as dynamically important and likely to trap particles, we are now ready to simulate the trapping phenomenon. We start in the spherical isochrone potential ($q = 1$). In order to populate the invariant tori of the spherical system, we find the actions of the spherical isochrone potential by $\oint_{\gamma_i} P \cdot dQ$ as described in Chapter 1. The path γ_i is chosen to lie in the (P_i, Q_i) plane. It is found (see references in Appendix C) that

$$J_\varphi = |L_z| \tag{82}$$

$$J_\vartheta = L - |L_z| \tag{83}$$

$$J_r = \frac{1}{\sqrt{-2E}} - \frac{L}{2} - \sqrt{\frac{L^2}{4} + b}. \tag{84}$$

Substituting $L = J_\vartheta + J_\varphi$ into Eq. 168 and treating $E = H(J)$,

$$H(J_r, J_\vartheta + J_\varphi) = -2\left(2J_r + (J_\vartheta + J_\varphi) + \sqrt{(J_\vartheta + J_\varphi)^2 + 4b}\right)^{-2}. \tag{85}$$

The frequencies are given by $\omega_0 = \nabla_J H(J)$, so $\omega_\vartheta = \omega_\varphi$. Thus all spherical potentials have at least one degeneracy in frequencies (see footnote 10). In the case of the isochrone potential, Eq. 81 reveals that the spherical system does not experience the low order and thus potentially very important $\omega_r = \omega_\vartheta$ resonance. As shown in the previous subsection, however, at some numerically determinable axis ratio $q = q_{\mathrm{res}}$, the 1 : 1 resonance appears. This resonance has previously been studied (Sridhar and Touma, 1996) as a possible heating mechanism for the Galactic disk.

Consider starting with the spherical case $q = 1$ and gradually flattening the system. The tori of the initially spherical system are populated as follows. We choose a surface of section $z = 0$, $\dot{z} > 0$ and populate initial conditions in the xy-plane[27] with various velocity *launch angles* ϑ_v. By ϑ_v we mean the velocity space spherical polar angle. We choose an initial cylindrical radius R, orbital energy E and special component of the angular momentum L_z such that the velocity in the meridional plane v_m is given by

$$v_m = \sqrt{2\left[E - \Phi(R,0)\right] - \left(\frac{L_z}{R}\right)^2} \tag{86}$$

with the cylindrical polar radius $R = r$. Then the initial conditions for these orbits with initial azimuthal angle $\varphi_0 = 0$ are $(x, y, z, \dot{x}, \dot{y}, \dot{z}) = (R, 0, 0, v_m \sin\vartheta_v, \frac{L_z}{R}, v_m \cos\vartheta_v)$. Having determined the initial velocity phase space point, we apply M, the map of Chapter 1, to give us $(r, \vartheta, P_r, P_\vartheta) \overset{M}{\mapsto} (\theta_r, \theta_\vartheta, J_r, J_\vartheta)$. The J we recover label the torus corresponding to the orbit, so that we now randomly choose n particles in the torus with $(\theta_r, \theta_\vartheta) \in [0, 2\pi] \times [0, 2\pi]$. From M^{-1} we recover $(r, \vartheta, P_r, P_\vartheta)$. We also need φ, which is obtained by

---

[27] Axisymmetric systems are of course fundamentally two-dimensional, but the computational implementation is easiest in Cartesian space.

observing[28] that for $\vartheta < \pi/2$

$$\varphi = \theta_\vartheta - \theta_\varphi + \arcsin\left[\cot\vartheta \frac{J_\varphi}{\sqrt{(J_\varphi + J_\vartheta)^2 - J_\varphi^2}}\right], \tag{87}$$

where $J_\varphi = L_z$ and $J_\varphi + J_\vartheta = L$. Because a spherical potential is characterized by $\omega_\vartheta = \omega_\varphi$, $I = \theta_\vartheta - \theta_\varphi$ is an integral of motion. For initial spherical polar angle $\vartheta_0 = \pi/2$, and with $\vartheta_v < \pi/2$ such that $\dot{z} \neq 0$, $I = \varphi_0$ so that

$$\varphi = \varphi_0 + \arcsin\left[\cot\vartheta \frac{J_\varphi}{\sqrt{(J_\varphi + J_\vartheta)^2 - J_\varphi^2}}\right]. \tag{88}$$

We are thus able to recover the Cartesian velocity phase space point from

$$x = r\sin\vartheta\cos\varphi \tag{89}$$

$$y = r\sin\vartheta\sin\varphi \tag{90}$$

$$z = r\cos\vartheta \tag{91}$$

$$\dot{x} = P_r\sin\vartheta\cos\varphi + \frac{P_\vartheta}{r}\cos\vartheta\cos\varphi - \frac{L_z\sin\varphi}{r\sin\vartheta} \tag{92}$$

$$\dot{y} = P_r\sin\vartheta\sin\varphi + \frac{P_\vartheta}{r}\cos\vartheta\sin\varphi + \frac{L_z\cos\varphi}{r\sin\vartheta} \tag{93}$$

$$\dot{z} = P_r\cos\vartheta - \frac{P_\vartheta}{r}\sin\vartheta \tag{94}$$

and can integrate the orbit forwards by the Runge-Kutta technique while gradually flattening the isochrone potential.

---

[28] The analytic expressions for the angles θ are clumsy and quite unintuitive and are not included here. For the details see Binney & Tremaine, or Appendix C.

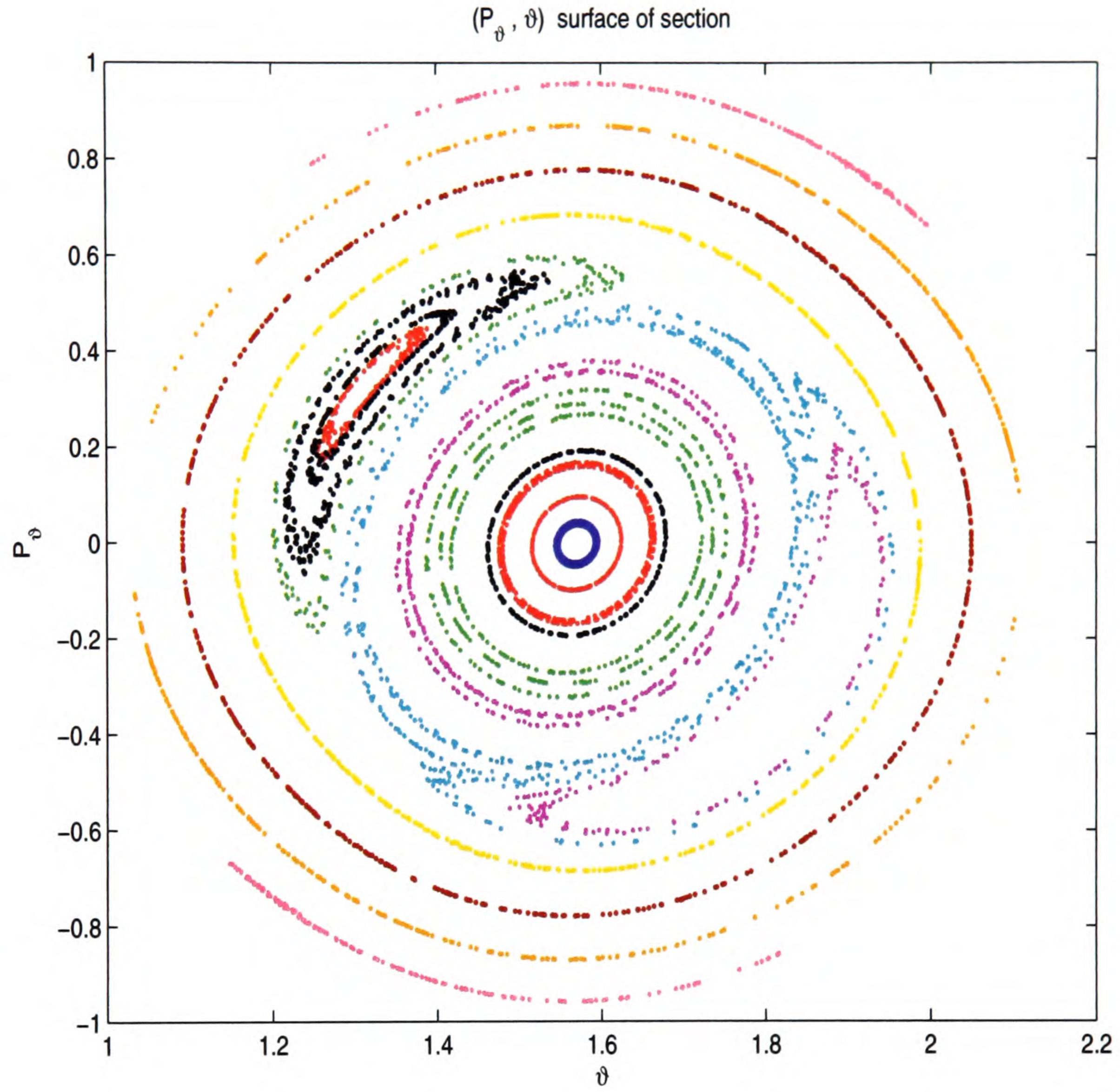

Figure 7: Poincaré section (P_ϑ, ϑ) for orbits at identical initial energy and angular momentum and a range of initial velocity polar angles ϑ_v. For the orbital torus defined by each choice of ϑ_v, four orbits uniform in phase $(\theta_r, \theta_\vartheta)$ over the torus were integrated. With decreasing polar velocity launch angle from the zero-velocity torus, the colours used were blue, red, black, green, magenta, cyan, yellow, brown, orange, pink. The section is generated for consequents satisfying $R = \frac{1}{2}(R_{min} + R_{max})$ and $\dot{R} > 0$, with R_{min} and R_{max} the minimum and maximum radii of the zero-velocity orbit (see Eq. 167). The initial conditions are $L_z = 1$, $E = -0.135$.

Consider a representative sample of tori of a spherical isochrone potential, with a given E and L_z, but differing in initial velocity polar launch angle ϑ_v. In each torus choose four initial conditions varying in angle $(\theta_r, \theta_\vartheta)$. Integrate the system, while slowly flattening the potential to guarantee the adiabatic invariance of non-resonant orbits. For some $q_{\text{final}} < q_{\text{res}}$ stop the flattening and integrate forwards[29] at $q = q_{\text{final}}$. Plots of the resulting consequents in the (P_ϑ, ϑ) Poincaré section are shown in Fig. 7. Consequents of orbits in the same initial torus are given the same colour.

KAM tori are seen as nested closed curves circulating around a common midpoint. Within the separatrix, the dynamics is altered completely. Concentric resonant islands are seen on the Poincaré map in place of the expected KAM torus. These are the libration tori discussed above, and they librate around the exact resonance condition. All initial conditions in KAM tori interior to the resonance at $q = q_{\text{final}}$ lie on the same final curves in the (P_ϑ, ϑ) section. This confirms the adiabatic invariance of the flattening procedure. Initial conditions in the same torus are seen to form quite distinct orbits. This is a reflection of the breakdown of adiabatic invariance arising from their interaction with the resonance. As in the spherical case, in the immediate vicinity of a resonance (the yellow curves), a circuit in the section takes significantly longer than further from resonance, evidence that in the finite time over which the orbit is computationally integrated, the orbit in a near-resonant torus fills a smaller volume of angle-space than in a torus far from resonance. From the appearance of two resonant islands in this and the subsequent (P_R, R) and (P_ϑ, ϑ) sections, we know that the resonant orbit punctures both surfaces of section every half-period so that, in fact, we are seeing the 2:2 resonance. The distinction is irrelevant here,

---

[29] The resonance is found numerically to appear for $q_{\text{res}} = 0.531$.

but is important in the Arnold webs of higher-dimensional systems (see Chapter 6).

We can read the evolution of the orbits as the system was flattened from the librating tori. In the spherical system, the resonance condition was not satisfied anywhere. For $q > q_{\rm res}$ the surface of section remained identical through adiabatic invariance. At $q = q_{\rm res}$ the resonance appeared at the zero-velocity torus and had infinitesimal width. The resonance was not wide enough to capture particles in the initial torus colour-coded blue. As the resonance grew and moved inwards, it caused a non-adiabatic $\delta I_\vartheta > 0$ in the red, green etc. orbits and succeeded in capturing some of these into one of its resonant islands. The red librational orbit was captured first: as the separatrix grew, it librated closer to the elliptic fixed point of the exact resonant condition. Black and green orbits were captured subsequently. The cyan orbit has the largest action of the librational tori. The hyperbolic fixed points and the associated chaotic regions of separatrix splitting lie in the empty regions between the two resonant islands. At $q = q_{\rm final}$, the resonance had not yet encompassed the yellow, brown, orange and pink orbits.

In Fig. 8 we see the orbits of Fig. 7 in perturbed frequency ratio. We are able to confirm from the diagram that the system contains the $1:1$ resonance condition. The captured red orbit, having been captured the longest, librates closest to the elliptic fixed point. We observe, however, that perturbed frequency is not a good diagnostic of resonant capture: among the magenta and cyan orbits that the surface of section clearly reveals to be captured, we can say little more from Fig. 8 at first sight than that they hover near the resonance in perturbed frequency.

Despite the crudeness of the calculation of the perturbed frequencies, we can extract a little more information, by considering the nearest low-order resonance $a_1 : a_2$ to the final

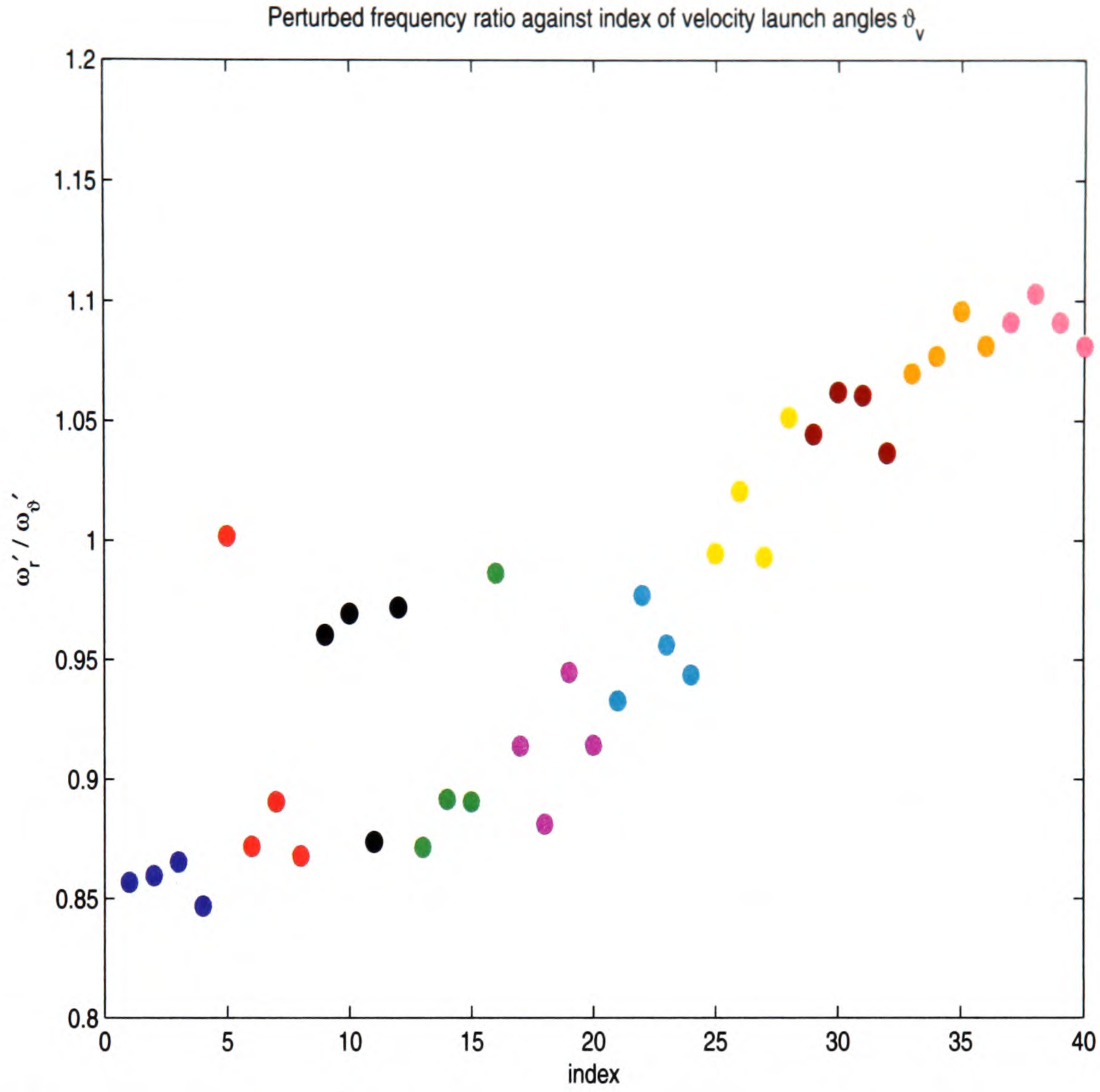

Figure 8: Perturbed frequency ratios for the orbits of Fig. 7. The same colour coding is used. For computational convenience, we chose to set $\Lambda = \delta$, the identity matrix, in (Eq. 26), so we do not expect very accurate perturbed frequencies; nevertheless, three black, one red, and a green orbit are obviously captured into resonance, and a degree of adiabatic invariance is manifest, especially in orbits away from resonance.

state of a particle after its integration in the gradually flattening potential. Consider all resonances in $\frac{\omega_r}{\omega_\vartheta} \in (0, 1]$ up to order n. The highest order resonance will be $n-1 : n+1$ for odd n and $n-2 : n$ for even n. The biggest gap between resonances up to order n will lie between this highest order resonance and $1 : 1$. Thus the largest error in estimating the resonance[30] up to order n is $1 - \frac{n-1}{n+1}$, so to get an error of at most δ in $\frac{a_1}{a_2}$ we need $n \geq \frac{2}{\delta} - 1$.

---

[30] Although a_1 and a_2 are uniquely defined by $(\frac{\omega_r}{\omega_\vartheta}, n)$, there is to my knowledge no simple way of producing them analytically.

For orbits through which the resonance has passed, i.e. interior to the yellow orbits in the (P_ϑ, ϑ) section, Table 1 shows the nearest resonance condition to within a chosen δ. We notice that for a given initial torus, if an orbit is trapped, then its value of $a_1 : a_2$ is closer to $1:1$ than that of any untrapped orbit.

3.2.3 The probability of capture into resonance

We can get a clearer picture by applying our method for approximating true actions I away from resonance. Fig. 9 and Fig. 10 are orbits similar to those of Fig. 7, seen in averaged action. In each plot, orbits were launched at a range of velocity polar angle $\vartheta_v \in (0, \pi/2)$. For each choice of launch angle and hence torus, four orbits random in phase $(\theta_r, \theta_\vartheta)$ were chosen and integrated forwards. Each is graphed separately on the plot. In Fig. 9 the resonance has not yet appeared at the zero-velocity torus. All the points correspond to KAM tori and exhibit the expected properties: their final true actions as approximated by our averaging procedure are independent of initial angles in accordance with the expected adiabatic invariance, and they tend towards the maximum zero-velocity action.

Fig. 10, a plot for $q < q_{\mathrm{res}}$, shows a completely different picture. Here it is immediately apparent that the actions are divided into three groups. Those towards the right of the plot correspond to the circulation region of KAM tori interior to the resonance on the Poincaré section. They are interior to and unaffected by the resonance and exhibit adiabatic invariance as before. Orbits launched near the zero-velocity torus fall into two groups, depending on whether they have been captured into the principal $1:1$ resonance. Even uncaptured orbits are affected by the presence of the resonance: their true actions do not lie on the pre-resonance path seen in Fig. 9, but seem to have suffered an increment in

colour	index	$a_1 : a_2$, $\lvert\frac{a_1}{a_2} - \frac{\omega'_r}{\omega'_\vartheta}\rvert < \delta$
blue	1	6:7
	2	6:7
	3	13:15
	4	11:13
red	**5**	**1:1**
	6	7:8
	7	8:9
	8	13:15
black	**9**	**19:20**
	10	**19:20**
	11	7:8
	12	**19:20**
green	13	7:8
	14	17:19
	15	8:9
	16	**1:1**
magenta	17	11:12
	18	15:17
	19	**17:18**
	20	11:12
cyan	21	14:15
	22	**1:1**
	23	19:20
	24	17:18

Table 1: For each orbit of Fig. 7 interior to the yellow orbits, the table shows the nearest resonance condition, $a_1 : a_2$, $|\frac{a_1}{a_2} - \frac{\omega'_r}{\omega'_\vartheta}| < \delta$, for $\delta = 0.05$. Resonantly trapped orbits are in bold-face. For each inital torus, i.e. for each colour, the resonantly trapped orbit(s) are found to be those with $a_1 : a_2$ closest to $1 : 1$.

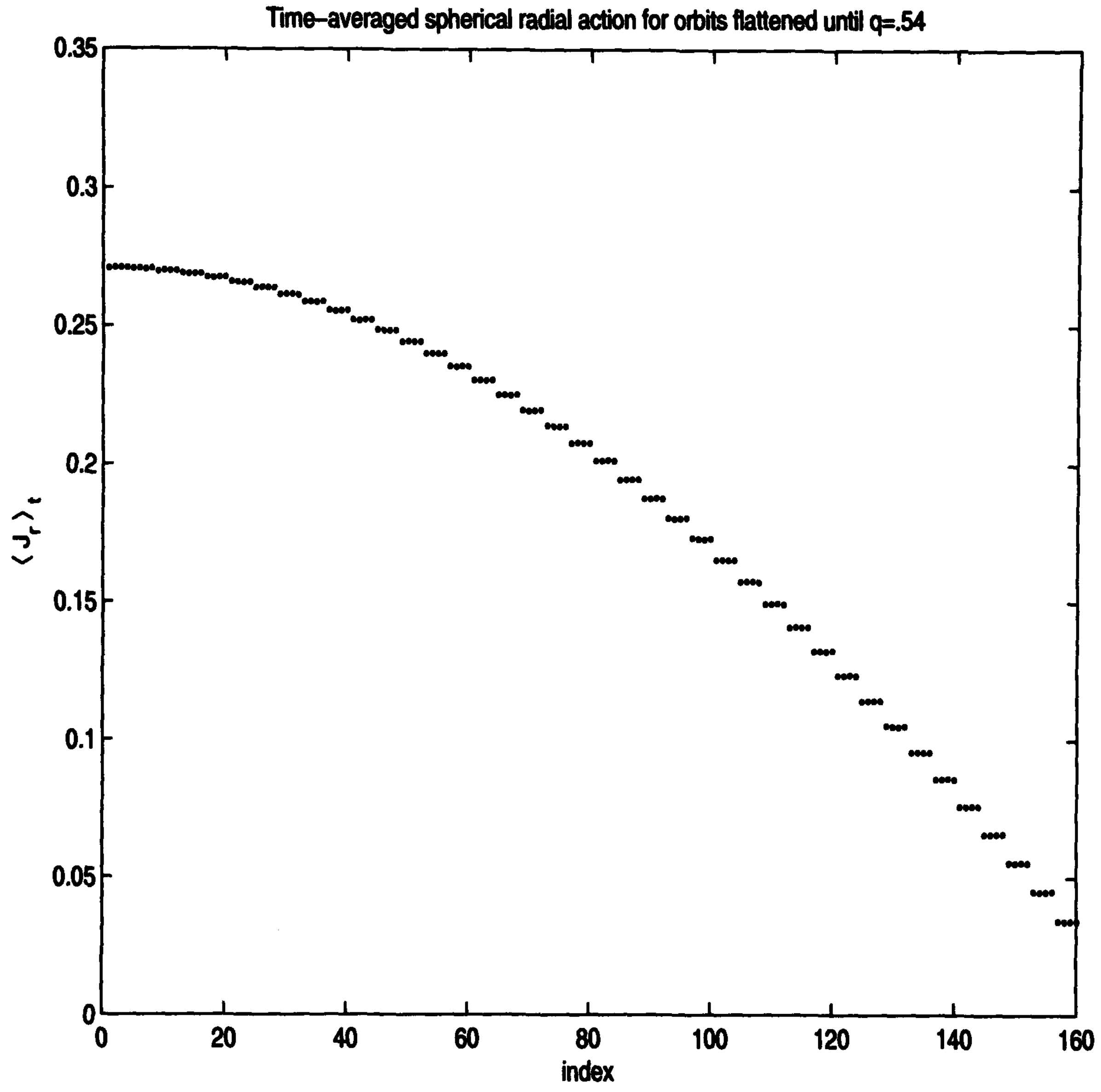

Figure 9: Spherical radial action averaged over axisymmetric orbits $\langle J_r \rangle_t$ at an axis ratio at which the 1:1 resonance has not yet arisen. For each initial polar launch angle ϑ_v four orbits random in phase are launched. These are graphed individually, with a spread introduced for clarity. The velocity launch angle decreases from left to right: the highest set of four indices corresponds to $\vartheta_v = 0$, the lowest four to the zero-velocity torus $\vartheta_v = \pi/2$. In this and subsequent figures, $L_z = 1$ and the initial orbital energy $E = -0.135$.

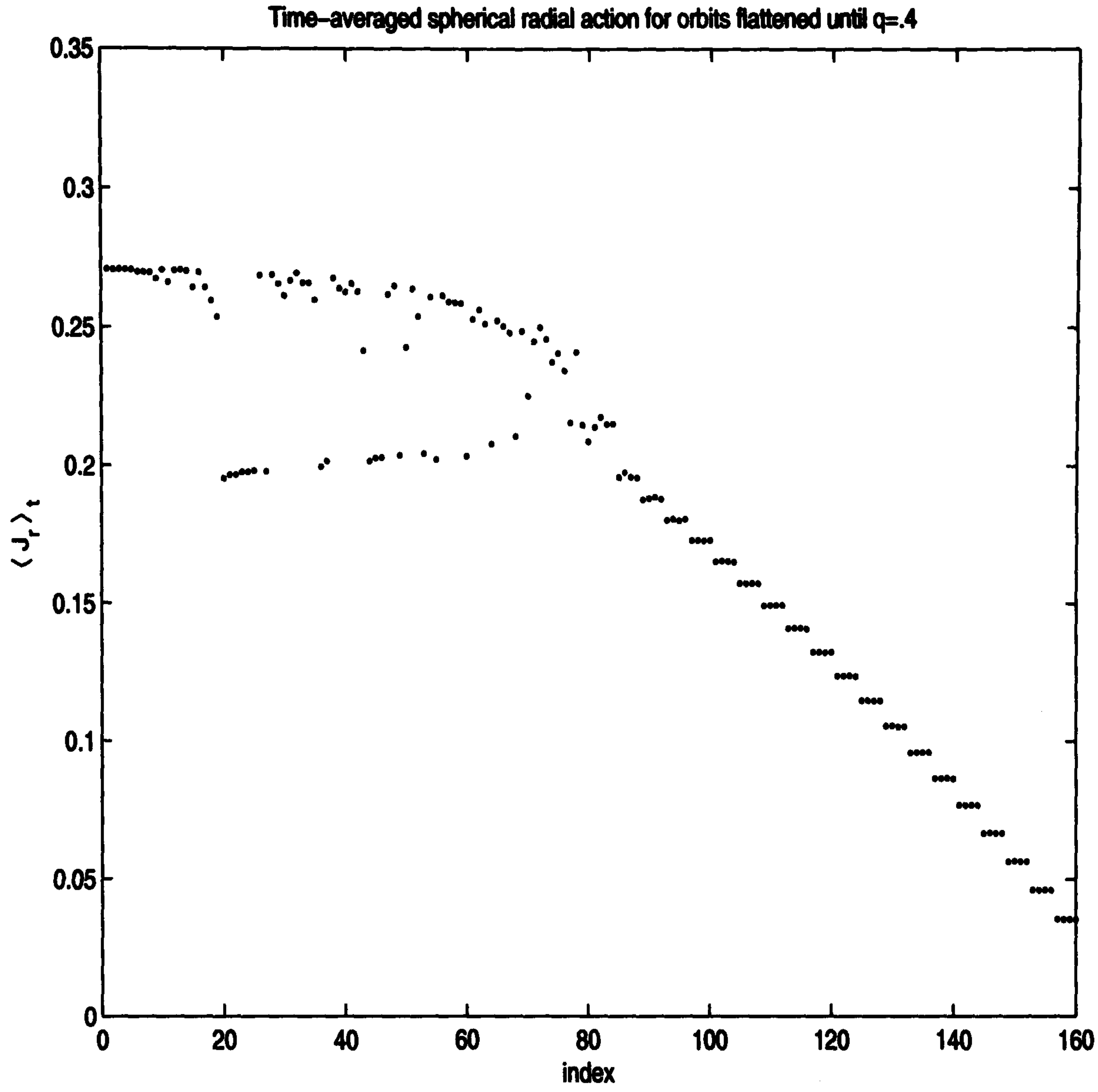

Figure 10: Spherical radial action averaged over axisymmetric orbits $\langle J_r \rangle_t$ for each of the integrated orbits of Fig 7. For each initial polar launch angle ϑ_v four orbits random in phase are launched. These are graphed individually, with a spread introduced for clarity. The velocity launch angle decreases from left to right: the highest set of four indices corresponds to $\vartheta_v = 0$, the lowest four to the zero-velocity torus $\vartheta_v = \pi/2$.

J_r, $\Delta J_r > 0$ asymptotically towards the maximum allowed action corresponding to the zero-velocity torus. As discussed earlier, the averaged actions of the third group, namely the trapped orbits, do not represent the true actions of these orbits, but nevertheless serve as a clear indicator of the libration region, lying approximately on a line corresponding to the exact resonance condition. The gap between the last adiabatically invariant orbit of the lower circulation region and the first untrapped orbit in the upper circulation region

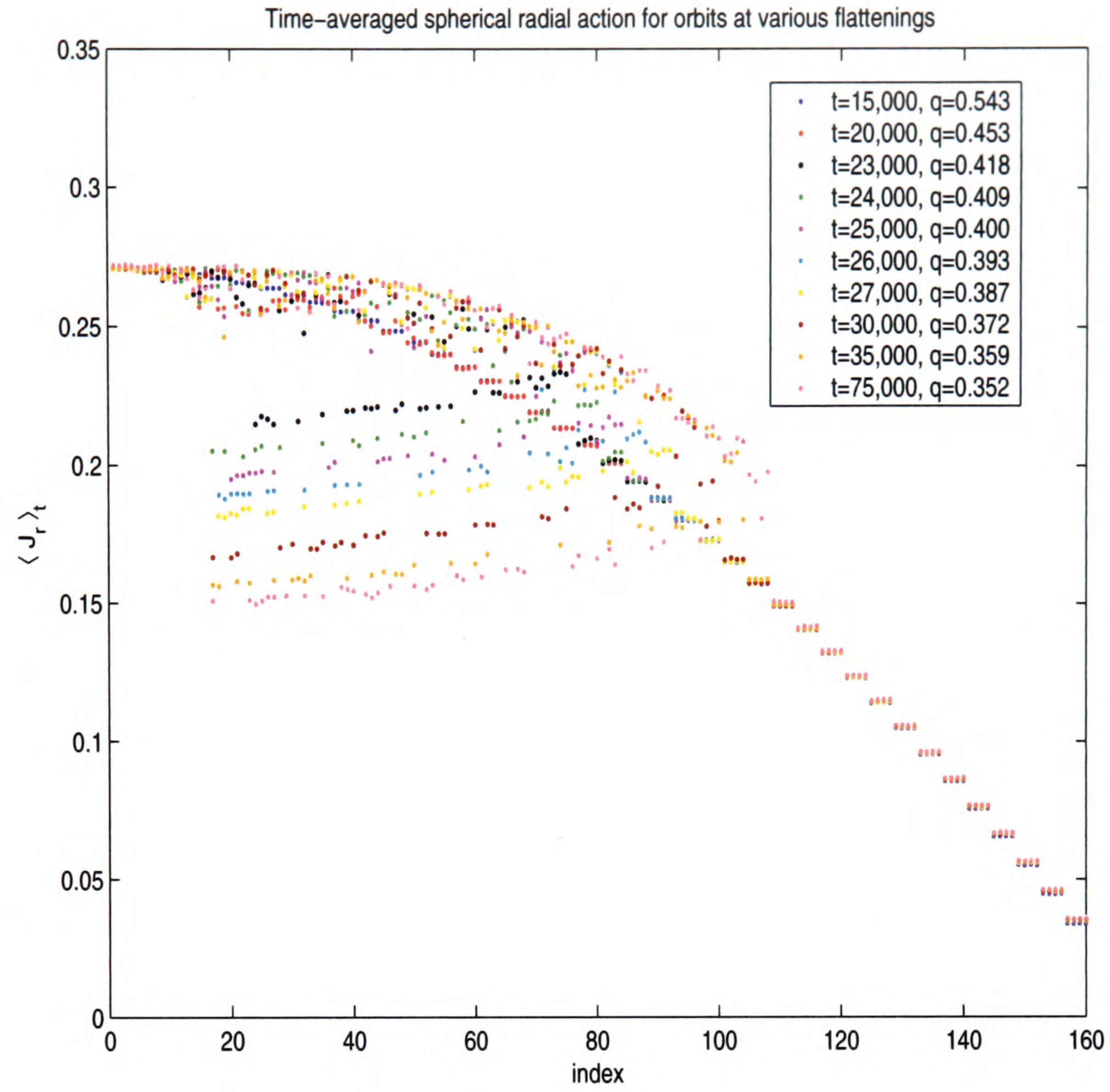

Figure 11: Spherical radial action averaged over axisymmetric orbits $\langle J_r \rangle_t$ at various axis ratios. Data from this plot are used to determine the probability of capture into resonance. The plot also demonstrates the growth with time of the resonance as it descends in radial action from the zero-velocity torus. The flattening function used here is $q(t) = [1 + \mathrm{erf}(1) + \mathrm{erf}(\frac{t}{15000} - 1)]^{-1}$.

corresponds to the width of the libration region, i.e. to the width of the resonance.

We can follow the time-evolution of the resonance in Fig. 11. The averaged actions of the trapped orbits decrease with time as they are dragged down by the descending resonance. Particles in the circulation region below the resonance are unaffected until the lower manifold of the separatrix descends to meet them. Particles not captured by the resonance are flung out behind the resonance with a sufficiently large $\dot{I}$ that they are observed on these plots only in their final state of release into the upper circulation region. Since the resonance not only descends but also grows, particles previously released into the upper circulation region are pushed upwards in action. The well-defined curve formed by these upper circulation region orbits indicates a smooth growth of the resonance.

At the largest polar velocity launch angle (smallest index) no orbits were trapped. When the resonance passed these initial conditions, it was too narrow to trap orbits effectively. The trapped orbits of lowest index have been resonantly trapped the longest and are particularly well-collimated in averaged action. Their librational tori are initially close to the the separatrix, but as the resonance grows, more particles are trapped behind them so that after the flattening, the orbits trapped first librate closest to the elliptic fixed point, while orbits trapped later librate near the separatrix. We observed this phenomenon previously in Fig. 7, where the black orbits were first to be trapped, and the cyan orbits were trapped later. In averaged action, the trapped orbits are seen to follow a clear line, with the longest-trapped orbits having smaller averaged actions than the more recently trapped ones. This indicates a discrepancy with idealized pendulum dynamics: the introduction of flattening causes the librational orbits to be not quite symmetric with respect to the exact resonance condition, but instead distorted somewhat in the direction of the growth of the resonance. The asymmetry is most conspicuous for those librational orbits that have been distorted the longest.

In Fig. 12 we see plots of averaged latitudinal action against initial polar velocity launch angle, again with four inital conditions in each torus. Resonantly trapped orbits are manifest on the horizontal branch of the diagram. Resonant trapping occurs first at an index corresponding to $\vartheta_v = 83°$. For $\vartheta_v \in (83°, 90°)$ all orbits are untrapped. Notice that there is

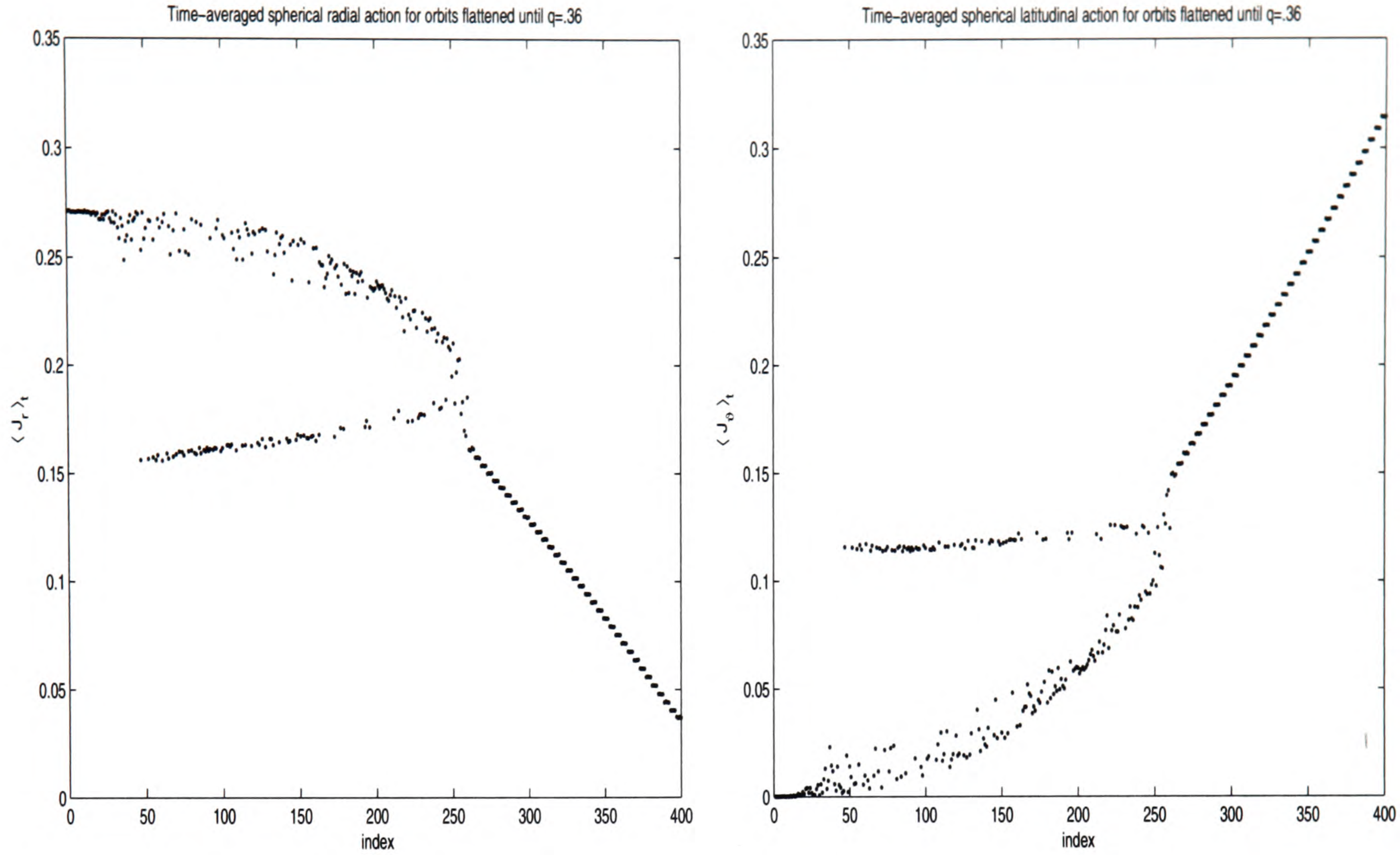

Figure 12: $\langle J_r \rangle_t$ and $\langle J_\vartheta \rangle_t$ averaged over trapped and untrapped orbits in an axisymmetric potential at a given axis ratio. Again polar velocity launch angle ϑ_v decreases stepwise from $\frac{\pi}{2}$ to 0 as the index increases.

very little spread in $\langle J_\vartheta \rangle_t$ and $\langle J_r \rangle_t$ in this range. For $\vartheta_v < 83°$, however, there is suddenly a wide range of actions among untrapped orbits. Particles not captured into resonance are being spread in final action by the presence of the resonance. We also see this in Fig. 7, in the spread of tori interior to the resonance. Notice from Fig. 12 that there is action overlap for these orbits, i.e. orbits in different initial tori can occupy the same final torus.

In Fig. 13 and Fig. 14 we see the evolution of the 2:2 resonance as the axis ratio is reduced. We populate identical initial conditions in each plot and integrate and flatten for the stipulated time. The surface of section is then generated at fixed potential to produce the plots shown. The plots affirm the results of our earlier averaged action plots, but

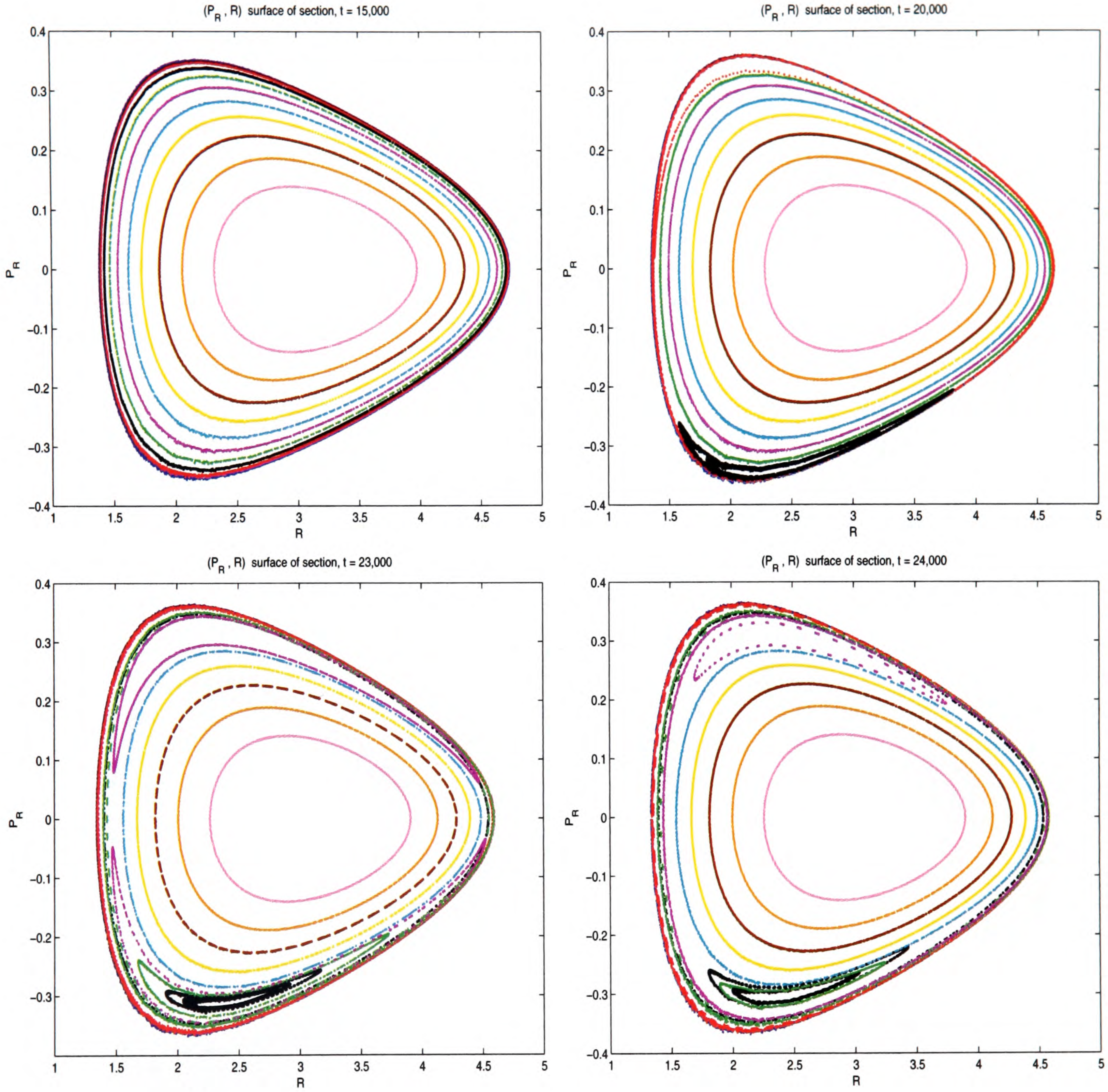

Figure 13: Poincaré maps to the (P_R, R) plane for a chosen set of initial conditions. As before, each of the initial orbital tori corresponds to a colour and is populated by four particles with different phases $(\theta_r, \theta_\vartheta)$. The particles are integrated for the stipulated times as the axis ratio of the potential is adiabatically reduced. The sections are then generated at fixed potential. All the orbits satisfy $L_z = 1$ and $E = -0.135$.

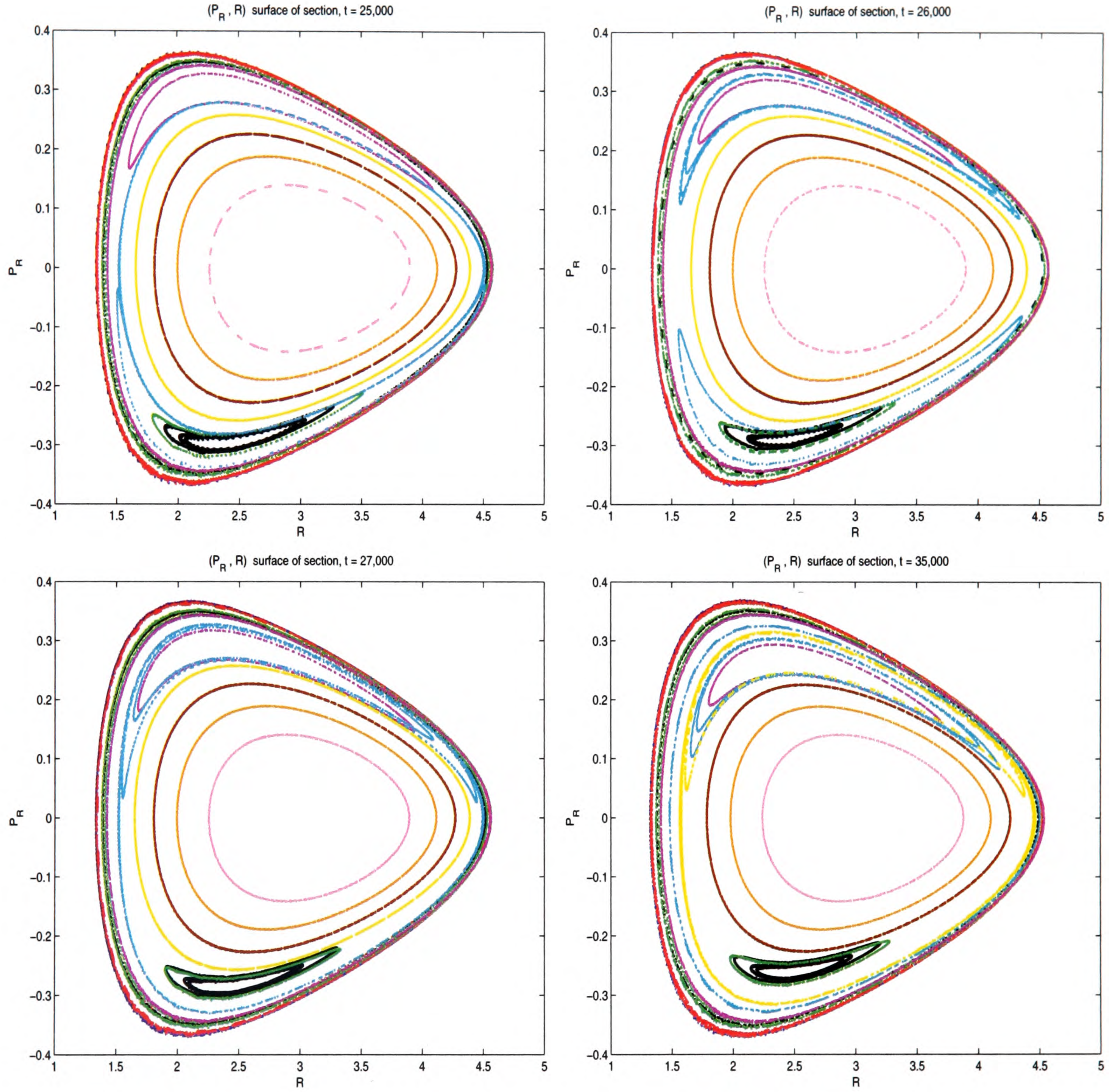

Figure 14: Poincaré maps to the (P_R, R) plane for the initial conditions of Fig. 13.

also yield some additional information about chaotic effects: we notice that the resonance leaks particles across its upper manifold despite its growth throughout the flattening. The red orbit trapped at $t = 2 \cdot 10^4$ (the corresponding q are listed in Fig. 11) is released by $t = 2.3 \cdot 10^4$. All four black orbits are trapped at $t = 2 \cdot 10^4$, but only three remain trapped by $t = 3.5 \cdot 10^4$. We observe similar effects in the cyan orbits.

We can see this result with greater clarity in the corresponding (P_ϑ, ϑ) sections displayed in Fig. 15. At $t = 1.5 \cdot 10^4$, the resonance has not appeared. All four black orbits and a red orbit are captured at $t = 2 \cdot 10^4$. The trapped red orbit receives a non-adiabatic change in action $\Delta I_\vartheta > 0$ and emerges as the outermost untrapped red orbit in the plot at $t = 2.3 \cdot 10^4$. We can assume that the spread in the blue, red, and green is due to a similar trapping phase. By $t = 2.7 \cdot 10^4$, the resonance has arrived at the yellow orbits and has discarded the magenta particles it trapped earlier. From the lack of spread in the magenta orbits, we can conclude that the trapping phase for these particles was brief compared with that of the black or green particles. This is confirmed in the (P_R, R) section at $t = 2.4 \cdot 10^4$ of Fig. 13.

In real space, the 2:2 resonant orbit oscillates in and out of the $z = 0$ plane, executing radial motion in tandem with its vertical motion. In Fig. 16 we see the innermost black trapped orbit of Fig. 13 in the Rz-plane.

The surfaces of sections permit an analysis of the growth of the resonance, but our averaged action plots can take us a step further. From Fig. 11 it is possible to read the actions corresponding to the upper and lower separatrices and the maximum action. By comparing plots at various axis ratios and thus building up a picture of the time-dependence of the actions, one might hope to recover the probability of capture into resonance as follows.

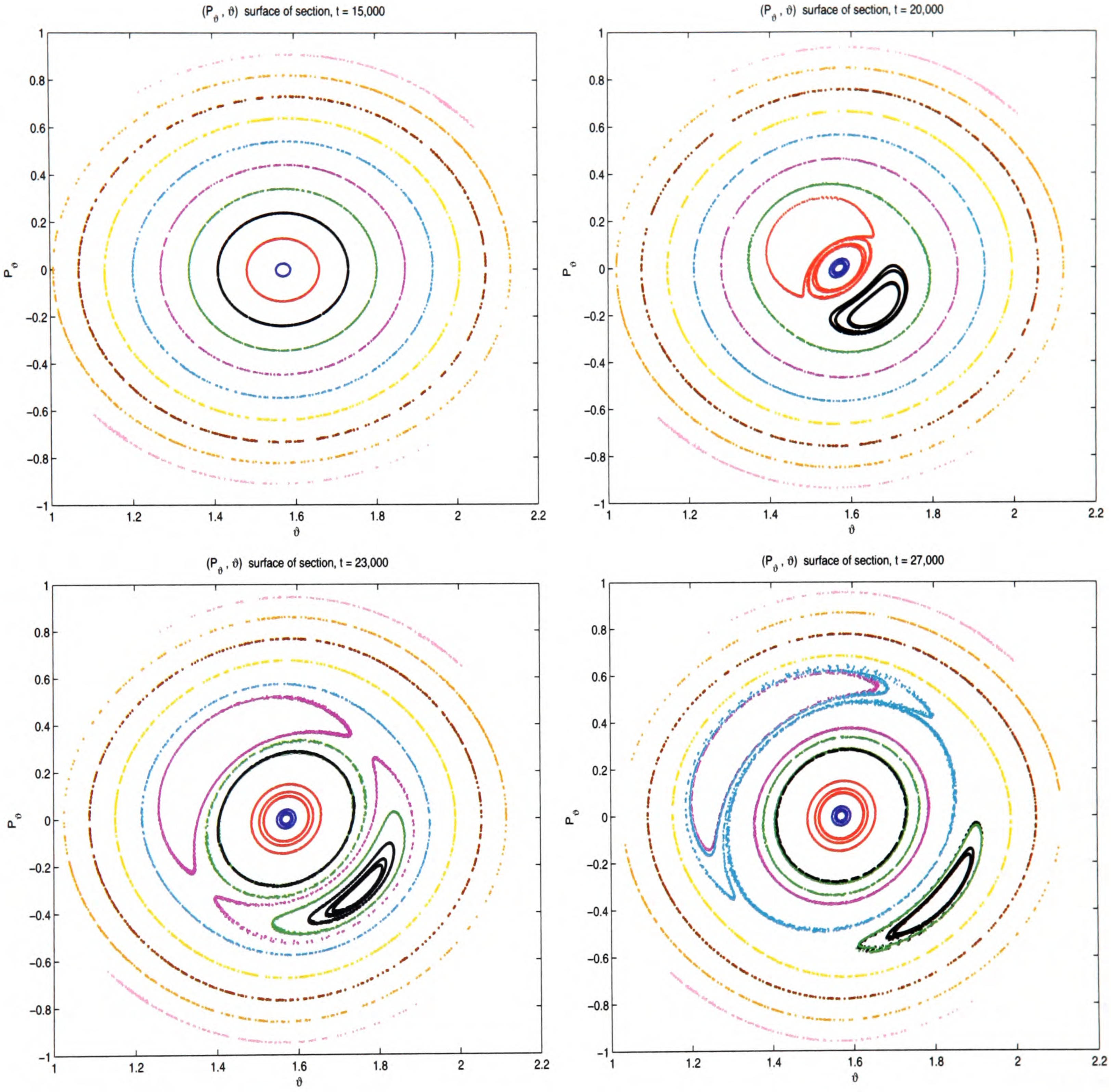

Figure 15: Poincaré maps to the (P_ϑ, ϑ) plane for the initial conditions of Fig. 13.

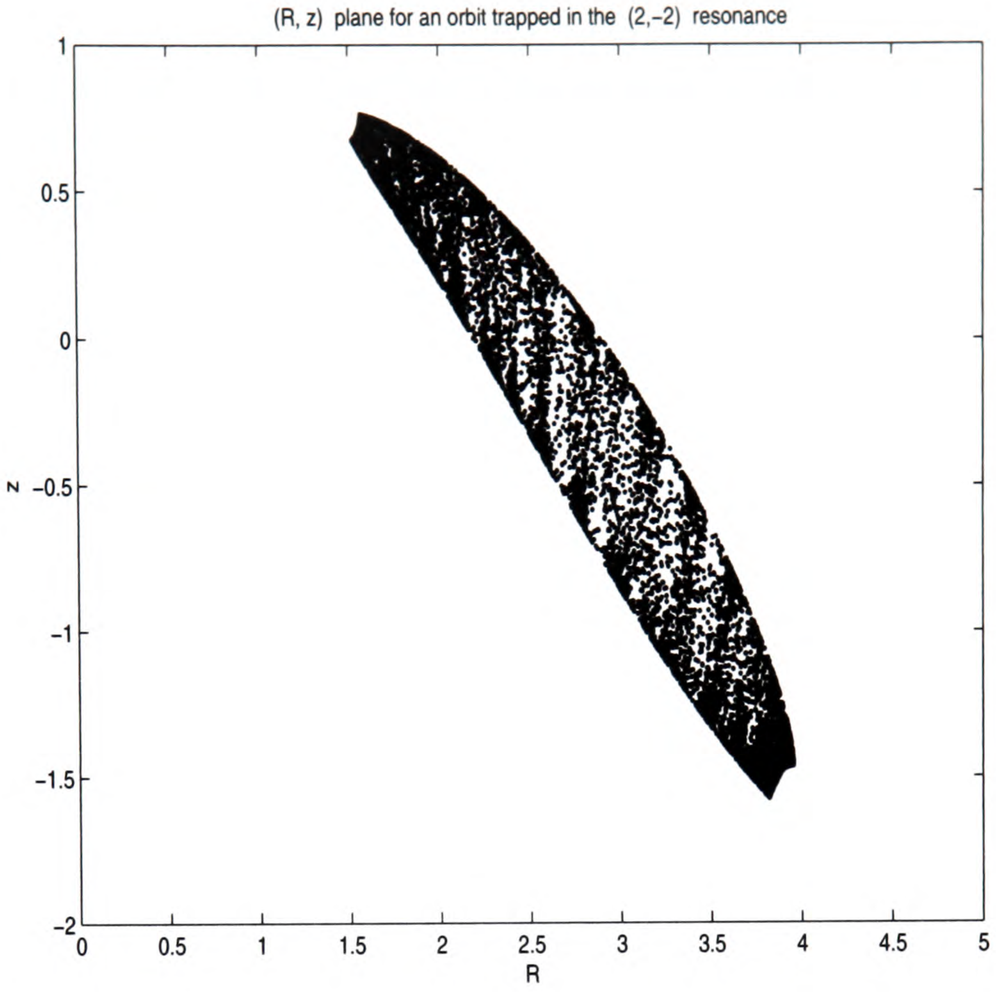

Figure 16: The innermost librating orbit of the capture simulations is shown in the Rz-plane. It executes vertical oscillations at a frequency very close to that of its radial oscillations.

Consider a growing resonance that is moving inward from the zero-velocity torus, for instance as a plot of K_1 against ψ_1. Let true actions and angles (i.e. the three discontinuous sets of actions generated by areas under circulation tori and areas within libration tori) be denoted by (I, ϕ). The true action of the circulation torus immediately below the separatrix I_- gives the area (barring a factor 2π) of the lower circulation region and satisfies $\dot{I}_- < 0$ since this area decreases as the resonance moves down. The resonance itself is growing so that for I_+ the action of a circulation torus immediately above the separatrix, the libration region satisfies $\dot{I}_+ - \dot{I}_- > 0$. The area of the upper circulation region is given by $I_{max} - I_+$, with $I_{max} = K_1^{max}$, the action of the zero-velocity torus. Because the time-mapping is canonical and adiabatic invariance guarantees that orbits starting in the zero-velocity torus will remain in it at some later time, $\dot{I}_{max} = 0$.

As the resonance grows and moves, the lower manifold of the separatrix encounters particles. As observed in Fig. 10, some of these particles will be captured, others flung out into the region above the resonance. Exactly what happens to a given particle depends on the details of its phase (Henrard and Lemaitre, 1983), but the capture probability averaged over initial angles can be obtained far more simply. Let ρ be the uniform density of particles in the lower circulation region. Then the number of particles encountering the resonance between times t and $t + \delta t$ is $-\rho \dot{I}_-(t)\delta t$. In that time, the libration region has grown by $(\dot{I}_+ - \dot{I}_-)\delta t$ and the upper circulation region by $-\dot{I}_+\delta t$. Then the probability of capture is

$$P_{capture} = \frac{(\dot{I}_+ - \dot{I}_-)\delta t}{(\dot{I}_+ - \dot{I}_-)\delta t - \dot{I}_+\delta t} = 1 - \frac{\dot{I}_+}{\dot{I}_-}. \tag{95}$$

Our estimator for $P_{capture}$ is $\hat{P}_{capture} = 1 - \frac{\frac{d}{dt}\langle J_+\rangle_\phi}{\frac{d}{dt}\langle J_-\rangle_\phi}$. When applied to the resonant capture simulation (a graphical depiction of the motion of the resonance is provided in Fig. 11), it is found that the probability of capture starts at 0.45 and decreases to 0.3 over the duration of the flattening. This result is in agreement with the plot of Fig. 10, which tells us (for instance merely by counting dots) that there were 84 orbits in the lower circulation region, of which 25 were captured: a capture probability of 0.29. The discrepancy arises from those points so close to the zero-velocity torus that the resonance was too narrow when it passed them for capture to occur. This band of points is easily seen in Fig. 10 as the points with smallest index for which none of the four initial phases permits capture. If we discount these points, we get a capture fraction of 0.37, well within the predicted range.

4 Capture into resonance and perturbation theory

In this chapter we test the performance of first-order perturbation theory by integrating orbits initially in the zero-velocity torus of the spherical isochrone potential as we flatten the potential. We then model the invariant actions of the quasi-integrable Hamiltonian with perturbation theory and compare them with averaged actions. We also confirm that we are able to observe pendulum dynamics in the resonant action-angle plane. We consider how perturbation theory can be used in an *adaptive dynamics*: we take a suitable parameter space for a toy potential and use the scatter of resonant orbits in the action plane and the distance in perturbed frequency ratio from resonance as diagnostics of the Hamiltonian.

In the Poincaré section, the 1:1 resonance starts infinitely close to the zero-velocity curve as axisymmetry is introduced and moves inwards as q decreases further. The zero-velocity orbit corresponds to initial velocity phase space conditions with $z = \dot{z} = 0$ such that $\frac{\partial\Phi(z=0)}{\partial z} = 0$ and the orbit is confined to the equatorial real-space plane $z = 0$. As $t \to \infty$ this non-closed orbit traces out a rosette and densely covers an annulus in the plane, whose pericentre and apocentre r_{min} and r_{max} can be obtained analytically (Eq. 167). In simulating the effect of the resonance, we choose an initial point in the annulus $(x, y, z) = (r_0, 0, 0), r_0 \in (r_{min}, r_{max})$, and choose E and L_z such that the velocity is determined by

$$(\dot{x}, \dot{y}, \dot{z}) = \left(\sqrt{2(E - \Phi(r_0, 0)) - \frac{L_z^2}{r_0^2}},\ \frac{L_z}{r_0},\ 0\right). \tag{96}$$

From here we proceed as before.

4.1 Performance of first-order perturbation theory

Computationally, the perturbation theory is performed as follows. We integrate around the flattened invariant torus I, $\dot{I} = 0$ and calculate spherical actions $J(t_i)$ for the i-th time-step t_i. Since

$$h(J,\theta) = \sum_{a \in \mathbb{Z}^2 - \{0\}} h_a(J) e^{ia\cdot\theta}, \tag{97}$$

we estimate the Fourier transform of h by setting up a grid $\theta = \frac{2\pi}{N}(k_1, k_2)$, $k_1, k_2 \in \mathbb{Z}$ and evaluating

$$\hat{h}_a(J) = \frac{1}{N^2} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} h\left(J, \theta = \frac{2\pi}{N}(k_1, k_2)\right) e^{-\frac{2\pi i}{N} a\cdot k}. \tag{98}$$

See Appendix B for an exact expression for $h_a(J) - \hat{h}_a(J)$. From Eq. 20, we now have an approximation to the generating function

$$\hat{S}(J(t_i), \theta'(t_i)) = J(t_i)\cdot\theta'(t_i) - i\sum_{a\neq 0} \frac{\hat{h}_a(J(t_i))}{a\cdot\omega_0(J(t_i))} e^{ia\cdot\theta'(t_i)}, \tag{99}$$

such that

$$J'(t_i) = J(t_i) + \sum_{a\neq 0} \frac{a\hat{h}_a(J(t_i))}{a\cdot\omega_0(J(t_i))} e^{ia\cdot\theta'(t_i)}. \tag{100}$$

At $t = t_i$, we can substitute $J(t_i)$ and $e^{ia\cdot\theta(t_i)} \approx e^{ia\cdot\theta'(t_i)}$ into Eq. 100 to yield $J'(t_i)$.

4.1.1 Small actions and effects near pericentre

In the upper lefthand plot of Fig 17 we see an untrapped orbit projected into the (J_r, θ_r) plane. We graph both the spherical action $J_r(\theta_r(t))$ (blue) and the perturbed action $J_r'(\theta_r'(t))$ (red) on the same axes, and compare the resulting plots with the true radial action I_r as estimated by spherical radial action averaged over true orbits[31] $I_r \approx \langle J_r \rangle_t$. The agreement between averaged and perturbed radial action is good away from orbital pericentre, $\theta_r = 0, 2\pi$. Near pericentre, J_r' deviates significantly from I_r, because the high-frequency terms of the Fourier decomposition of the generating function are not sufficiently accurate to produce the large $J_r - J_r'$ needed in this regime. In the lower lefthand plot we see that throughout the domain of θ_ϑ the majority of the points lie near the true radial action. In the lefthand plot of Fig. 18 we see the orbit in $\mathbb{T}^2$ mapped into a plane. The torus is densely covered, confirming that the orbit is non-resonant. The righthand plot summarizes the limitations of perturbation theory. The true orbit is the fibre above the intersection of the green and black solid lines in the action plane. Perturbation theory (in red) produces actions that wander over time, clustering around the true radial action, but spread almost uniformly in latitudinal action across a domain of $O(J_\vartheta)$. For purposes of illustration, a large $\dot{\vartheta}(t=0)$ was chosen. For an orbit closer to the zero-velocity torus, the scale of the latitudinal action is much smaller than the scale of the radial action. In this case, the clustering effect in radial action is sufficient to make perturbation theory useful. Fortunately, for the purposes of the next section, the value of the latitudinal action for untrapped orbits is irrelevant.

Fig. 19 shows a trapped orbit in the various action-angle planes. In the angle-angle plot we see that the orbit is trapped in a 1:1 resonance. The thickness of the bands of dots corresponds to the amount of libration around the exact resonance condition. In the $J_r(\theta_r)$ plot we see that averaged actions provide a good estimator for the exact resonance condition. The width of the band of points is not a failure on the part of perturbation theory, but a

---

[31] Recall that $\langle J_i \rangle_t = \langle J_i \rangle_\theta$ only for circulation tori.

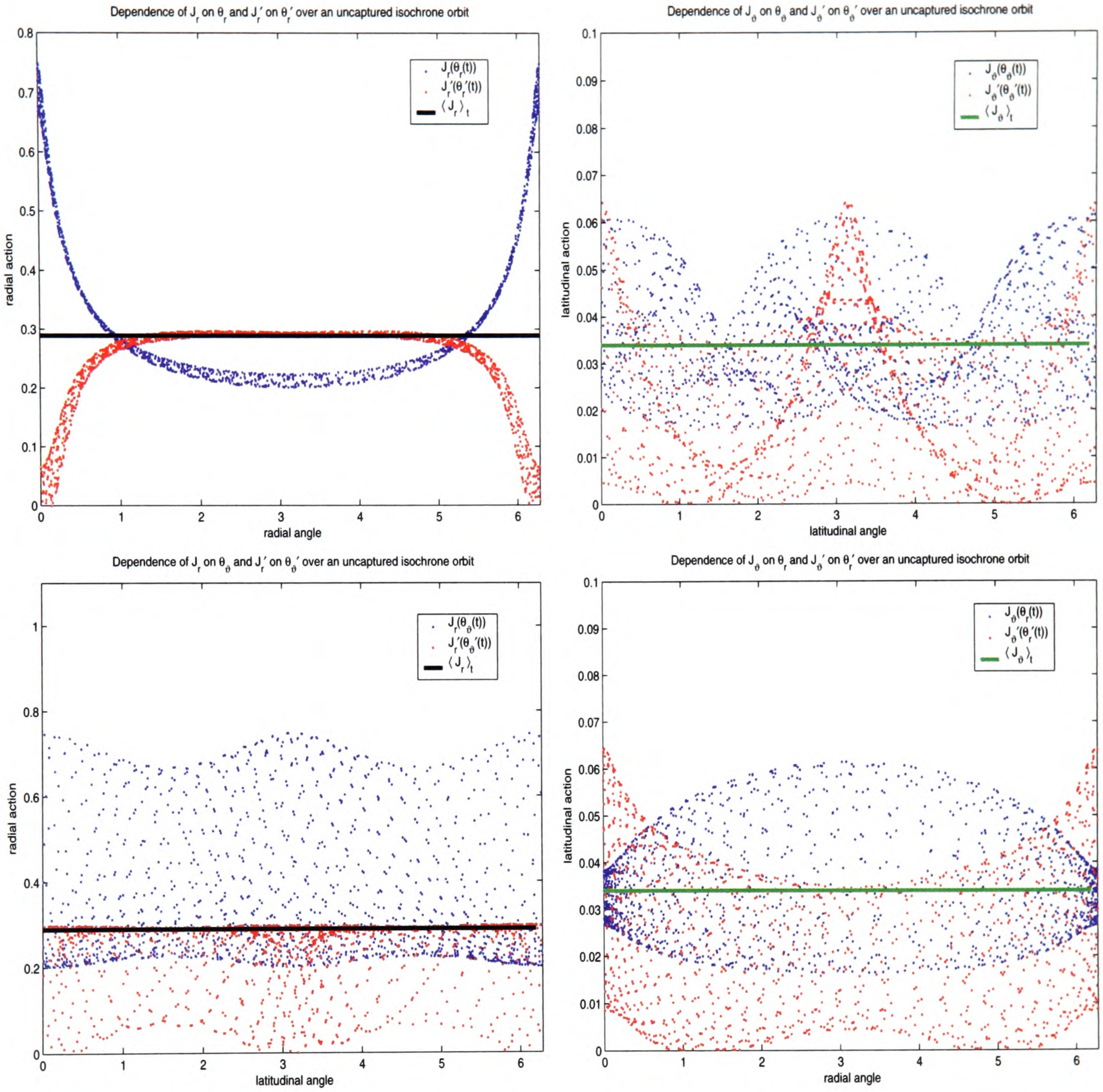

Figure 17: Orbits with initial conditions near the zero-velocity torus of a spherical isochrone potential are integrated forwards in time while flattening is gradually introduced into the potential. At an axis ratio $q(t_{\mathrm{final}}) < q_{\mathrm{res}}$ the flattening is switched off, and the orbit is integrated at fixed axis ratio $q(t_{\mathrm{final}})$. The graphs show projections of the orbit in spherical and perturbed quantities in the action-angle planes. True actions as estimated by $\langle J_r \rangle_{t>t_{\mathrm{final}}}$ and $\langle J_\vartheta \rangle_{t>t_{\mathrm{final}}}$ are graphed alongside.

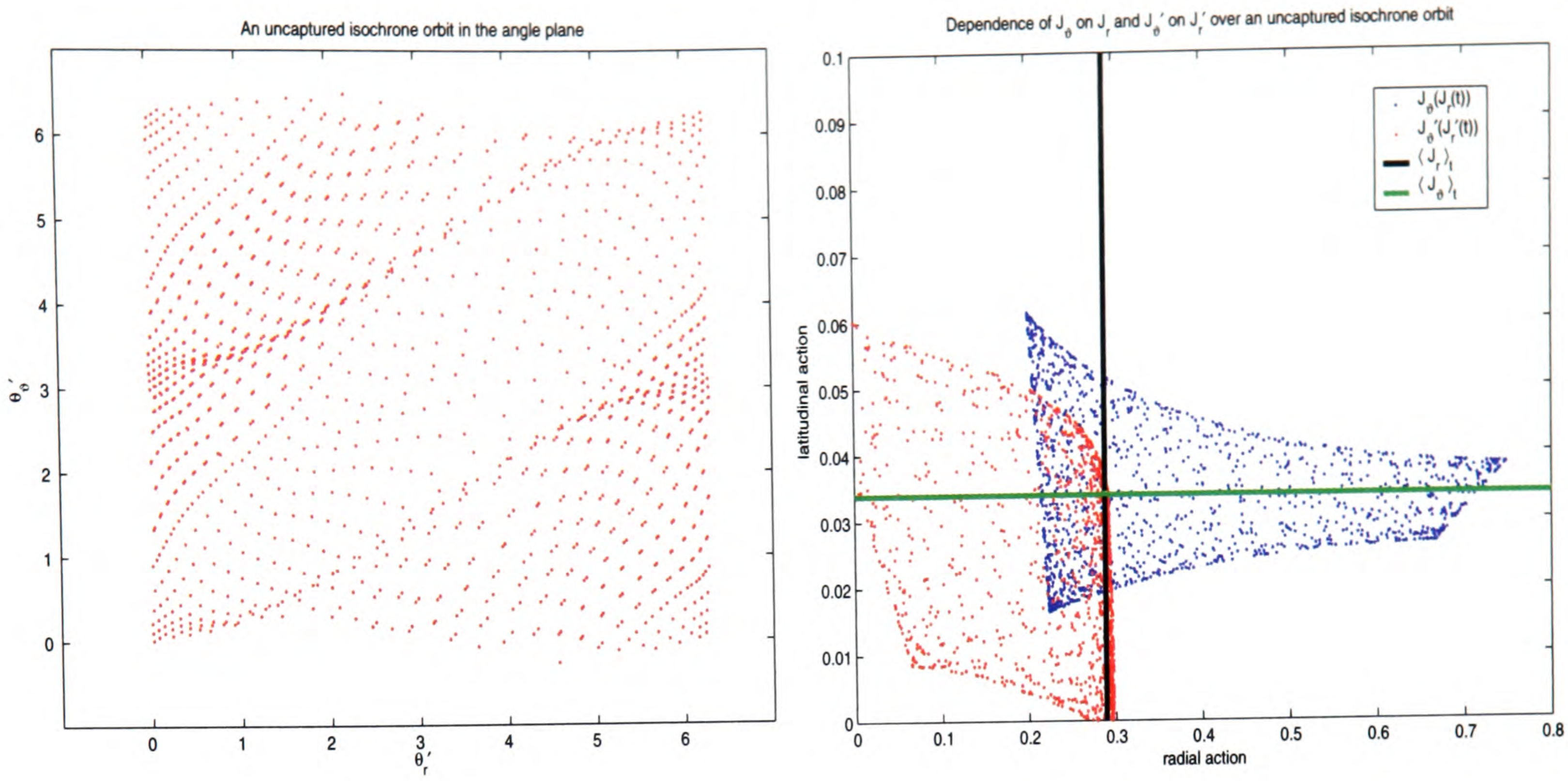

Figure 18: An untrapped orbit in the perturbed angle-angle plane (left) and a summary of the limitations of perturbation theory as seen in the action-action plane (right). The domain in spherical actions available to the orbit is seen in blue; the true actions of the invariant torus are approximated by the intersection of the black and green lines, and the scatter in perturbed action is seen in red.

representation of the libration. Near pericentre, perturbation theory again performs badly. The $J_\vartheta(\theta_\vartheta(t))$ plot reveals that perturbation theory is unable to significantly improve our estimate of the exact resonance condition in latitudinal action.

Recall from Eq. 67 and Eq. 69 that resonant action-angle coordinates are defined by $\psi = U\theta'$ and $K = (U^T)^{-1}J'$. For the 1:1 resonance

$$U = \begin{pmatrix} 1 & -1 \\ a & b \end{pmatrix}, \quad a + b = 1, \tag{101}$$

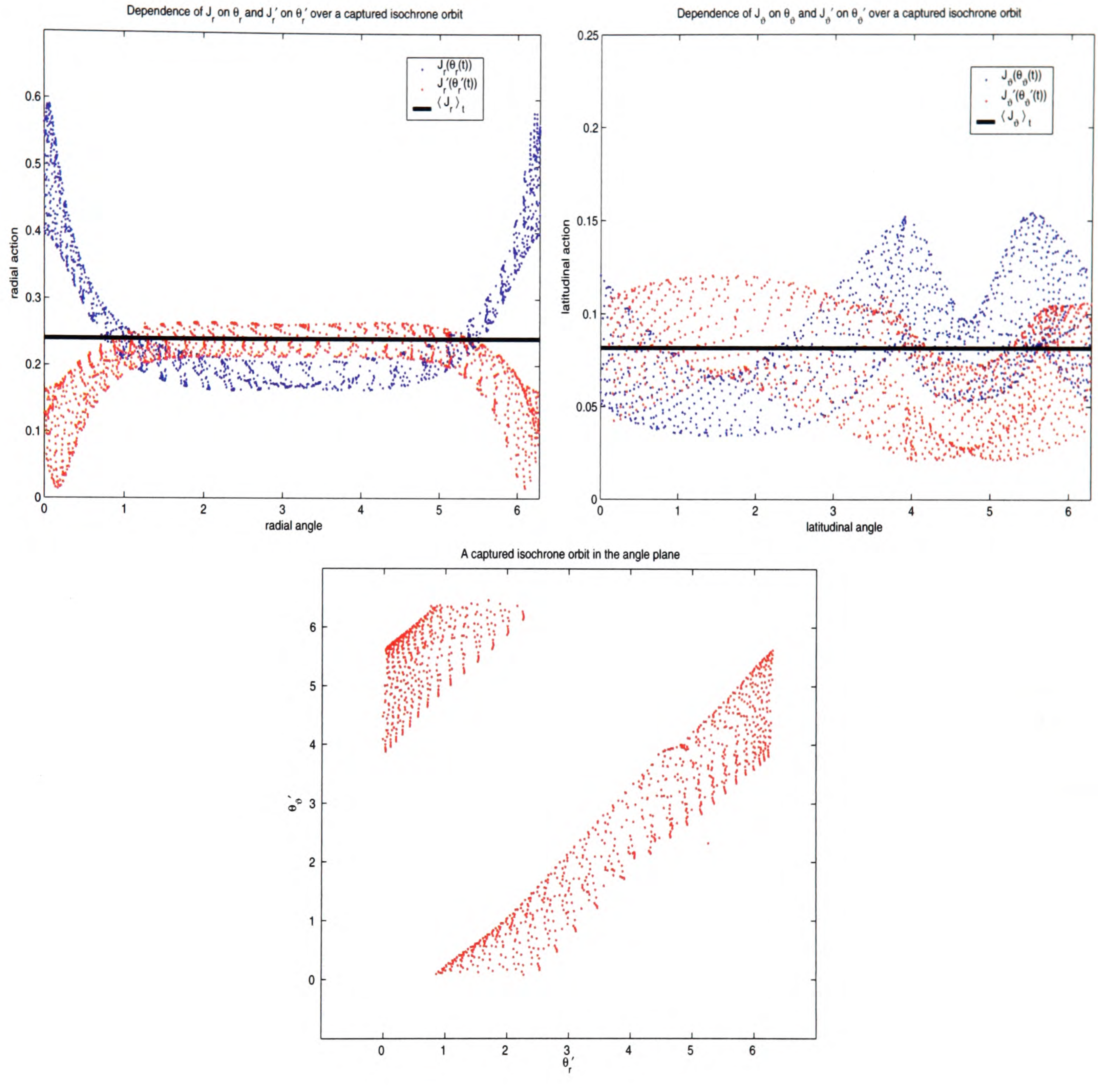

Figure 19: A trapped orbit in two action-angle planes and in the angle-angle plane.

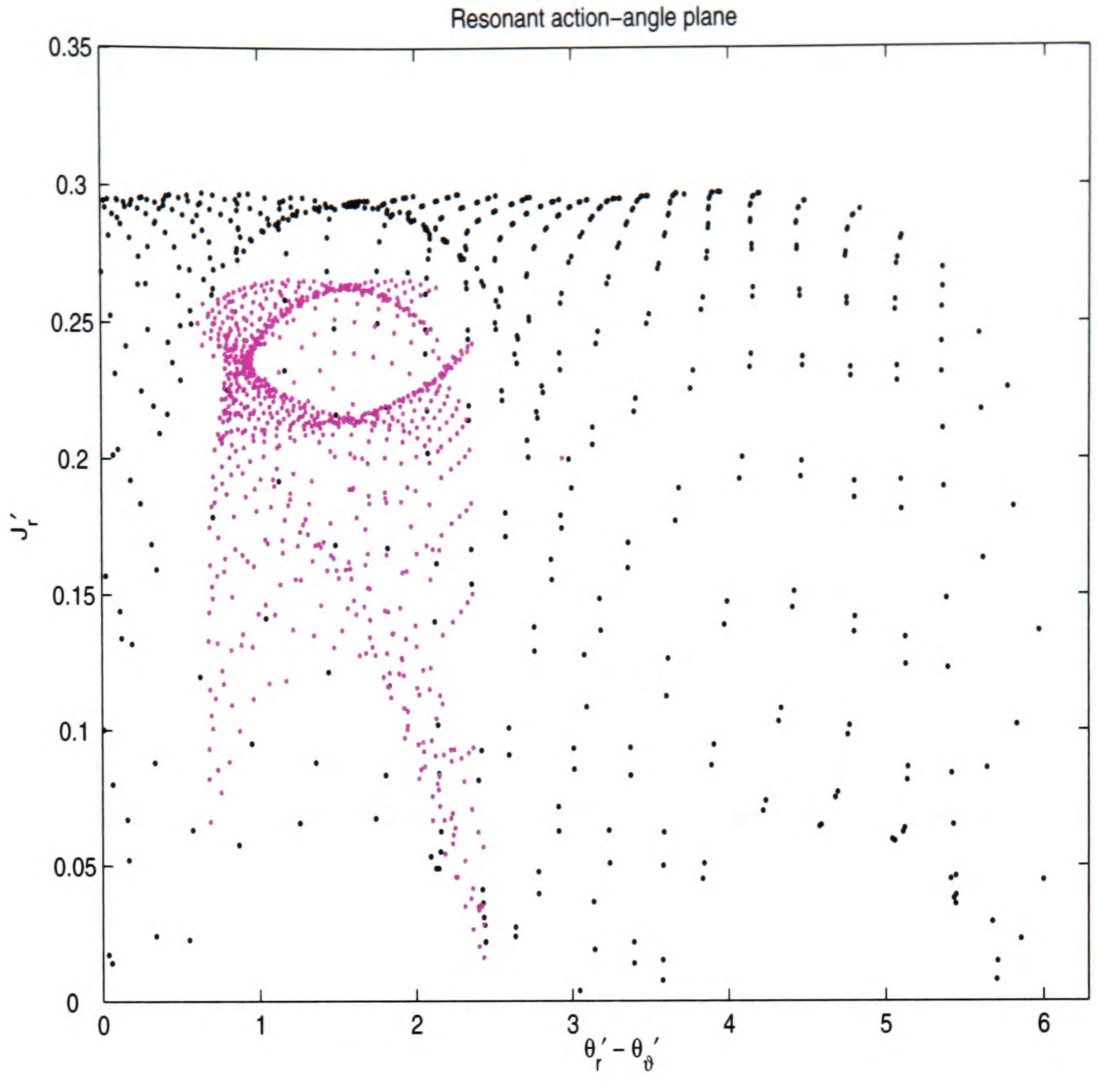

Figure 20: The untrapped (black) and trapped (magenta) orbits in perturbed actions in the resonant action-angle plane. Aside from those points near pericentre for which the perturbation theory is ineffective, we recover the expected pendulum dynamics, with the black orbit circulating and mostly constant at its upper bound, and the trapped orbit librating around the exact resonance condition.

so

$$\psi_1 = \theta_r - \theta_\vartheta \tag{102}$$

$$\psi_2 = a\theta_r + b\theta_\vartheta \tag{103}$$

$$K_1 = bJ_r' - aJ_\vartheta' \tag{104}$$

$$K_2 = J_r' + J_\vartheta'. \tag{105}$$

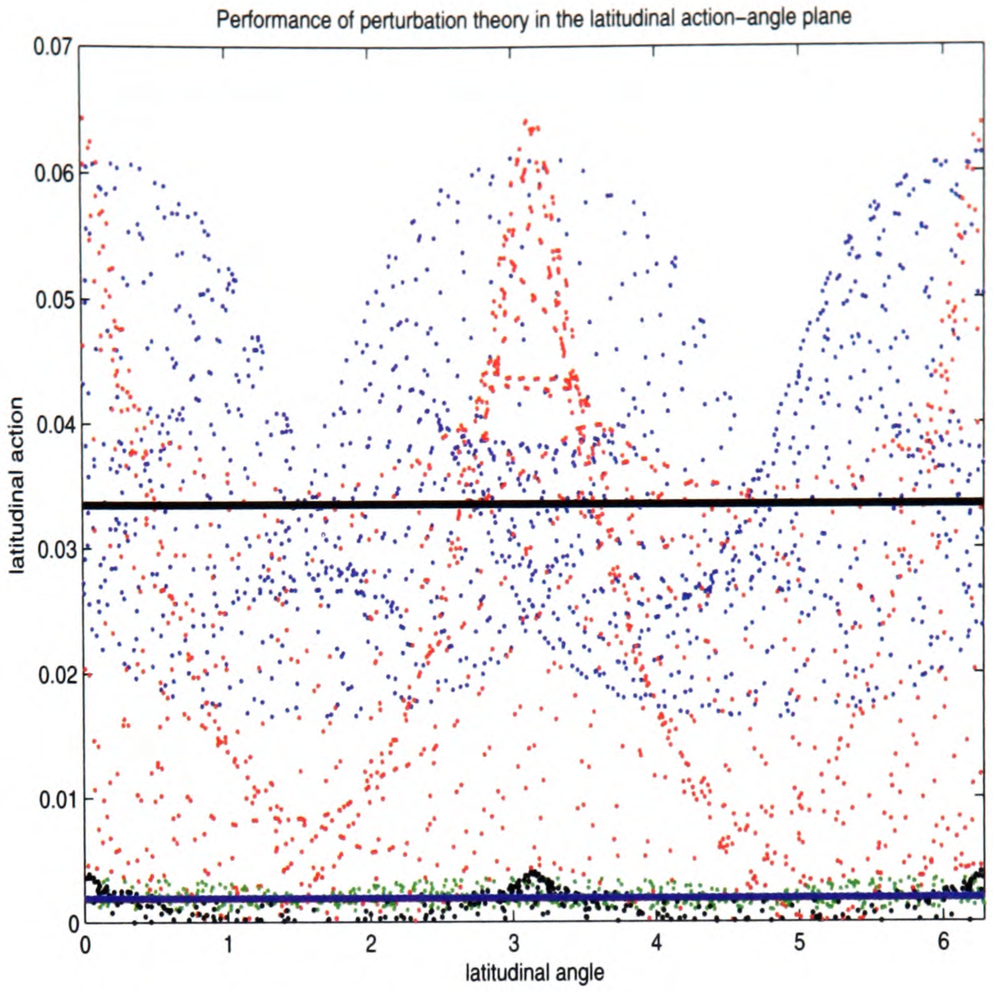

Figure 21: The error in the perturbed latitudinal action remains $O(J_\vartheta)$ as $\dot{\vartheta}(t=0)$ is reduced. The coefficients of the Fourier decomposition of the generating function are all found to scale according to the same relation. The blue and the green points are $(J_\vartheta(t), \theta_\vartheta(t))$ for two values of $\dot{\vartheta}(t=0)$. The red and black points are the corresponding plots of $(J'_\vartheta(t), \theta'_\vartheta(t))$. The black and the blue solid lines are $\langle J_\vartheta \rangle_t$ in each case.

In Fig. 20 we see the untrapped orbit of Fig. 17 (black) graphed alongside the trapped orbit of Fig. 19 (magenta) in the resonant action-angle plane[32] $(J'_r, \theta'_r - \theta'_\vartheta)$, defined by $a = 0$, $b = 1$. The libration of the trapped orbit is apparent, as is the circulating behaviour of the untrapped orbit. The wake of points beneath the ellipse is due to the behaviour near pericentre for which J'_r is significantly underestimated.

In Fig. 21 we examine the performance of the perturbation theory in the small action

---

[32] To be precise, we use J'_r etc. as calculated from the resonant normal form rather than the Birkhoff normal form. Since a computational implementation of the Birkhoff generating function always contains a finite number of terms, the only difference between the two generating functions is the resonant module. In practice, therefore, there is little difference.

J_ϑ as we make the orbit execute smaller vertical motions. We notice that the perturbation theory estimates the order of the latitudinal action correctly, but that its error stubbornly remains $O(J_\vartheta)$. Indeed, a thorough analysis of the coefficients in the Fourier decomposition of the generating function reveals that they all scale with the same factor. Further, it is found that in systems in which $J_\vartheta \approx J_r$ (for instance by decreasing L_z), the perturbation theory performs well in J_ϑ.

Why does a system in which $J_2 \ll J_1$ prove problematic for first-order perturbation theory? Looking back at Eq. 100, there can be only one explanation: the replacement of $e^{ia\cdot\theta(t_i)} \approx e^{ia\cdot\theta'(t_i)}$. The error is only $O(\epsilon^2)$ because $\hat{h}$ is itself $O(\epsilon)$, but the size of ϵ is determined by the largest action, in this case by $J_1' - J_1$. If $\frac{J_2}{J_1} \sim O(\epsilon^2)$, then the error in the perturbation theory $J_2' - J_2 \sim O(J_2)$.

4.2 Perturbation theory as a diagnostic of dynamical models

So far, we have considered a fixed initial orbital energy. We can get a more complete picture of the performance of first order perturbation theory by choosing the initial orbital energy to extend over some range. Orbits are launched as before in the zero-velocity torus of the spherical potential. The system is flattened and then integrated at fixed axis ratio from $t = 0$ until time $t = T$. In Fig. 22 we examine perturbed and averaged actions in the action plane. The spherical actions are averaged over $[0, T]$ to produce the averaged actions $(\langle J_r \rangle_t, \langle J_\vartheta \rangle_t)$ plotted in blue; the values of the perturbed actions $(J_r'(T), J_\vartheta'(T))$ at time T are plotted in red. For the perturbed Hamiltonian $G(J') = E$ and perturbed frequency ratio

$$f \equiv \frac{\omega_r'(J')}{\omega_\vartheta'(J')} = \left(\frac{\partial G}{\partial J_r'}\right)_{J_\vartheta'} \left(\frac{\partial J_\vartheta'}{\partial G}\right)_{J_r'} = -\left(\frac{\partial J_\vartheta'}{\partial J_r'}\right)_G, \tag{106}$$

so that $\left(\frac{\partial J_\vartheta'}{\partial J_r'}\right)_G = -1$ at resonance. For this axis ratio ($q = 0.4$), $f \approx 0.9$ for untrapped orbits, and so isoenergetic curves are approximately as shown. Recall that neither method claims to evaluate the invariant actions of trapped orbits correctly. Averaged actions nevertheless accurately measure the exact resonance condition, i.e. the action of a hypothetical circulation torus at the exact resonance condition. Averaged actions reveal that the exact resonance condition exhibits a strictly linear dependence $J_\vartheta^*(J_r^*(E))$. In Figs. 22 and 23 this is visible in a collimated line of blue points at latitudinal actions significantly above the cluster of untrapped orbits.

We observe that perturbation theory is able to classify orbits unambiguously as trapped or untrapped. For untrapped orbits, the accumulation of red points at low radial action is a manifestation of the difficulty with perturbation theory near pericentre. The inherent problem with applying perturbation theory to J_ϑ is not visible here because the true action of untrapped orbits is found to be distributed within a significant range. This is a manifestation of the breakdown of adiabatic invariance: because all initial conditions are chosen near the zero-velocity torus, the untrapped orbits have all been affected by the passage of the resonance. The perturbed action plot of Fig. 22 could in principle be produced from an astrophysical catalogue, by choosing a band in L_z and radius R and generating perturbed actions for a model Hamiltonian. The parameters of a trial Hamiltonian could for instance be the axis ratio q and the isochrone parameter b. When the perturbation theory is performed with incorrect values of these parameters, the scatter around the true actions

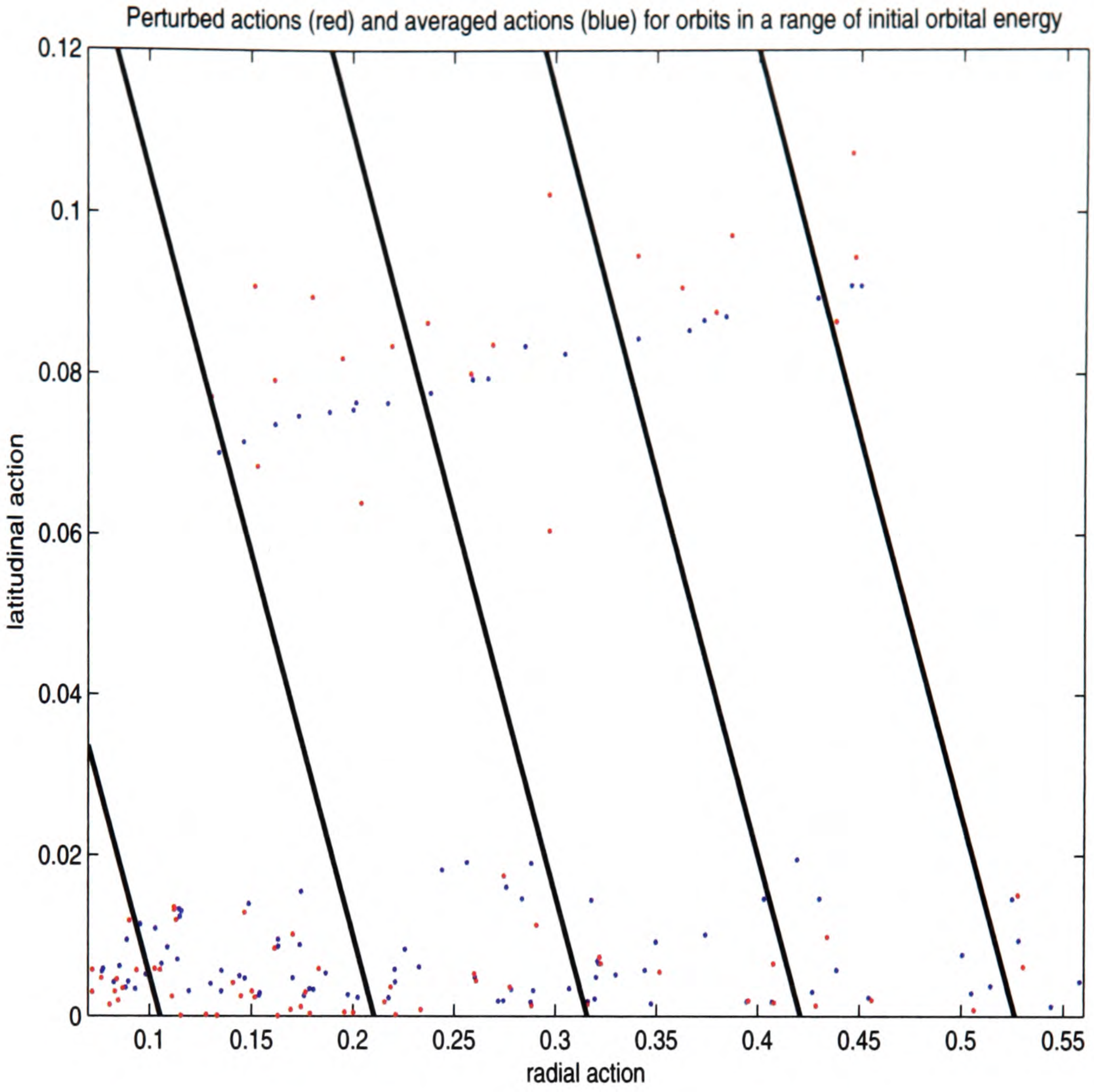

Figure 22: Averaged (blue) and perturbed (red) actions of a flattened isochrone potential. The initial orbital energy lies in a chosen range. Contours $G(J') = E$ run approximately as shown in black.

of the resonantly trapped orbits may be expected to be greater than when the parameters are chosen correctly. This is seen in Fig. 23. We have omitted points deemed too close to pericentre to be effectively modelled by perturbation theory. Such points could similarly be removed from a catalogue.

We observe in Fig. 23 that a horizontal cut $J_\vartheta(J_r) = \text{const}$ is sufficient to separate the resonantly trapped population from the untrapped population, even when the chosen perturbation parameters deviate significantly from the true parameters of the potential. Since

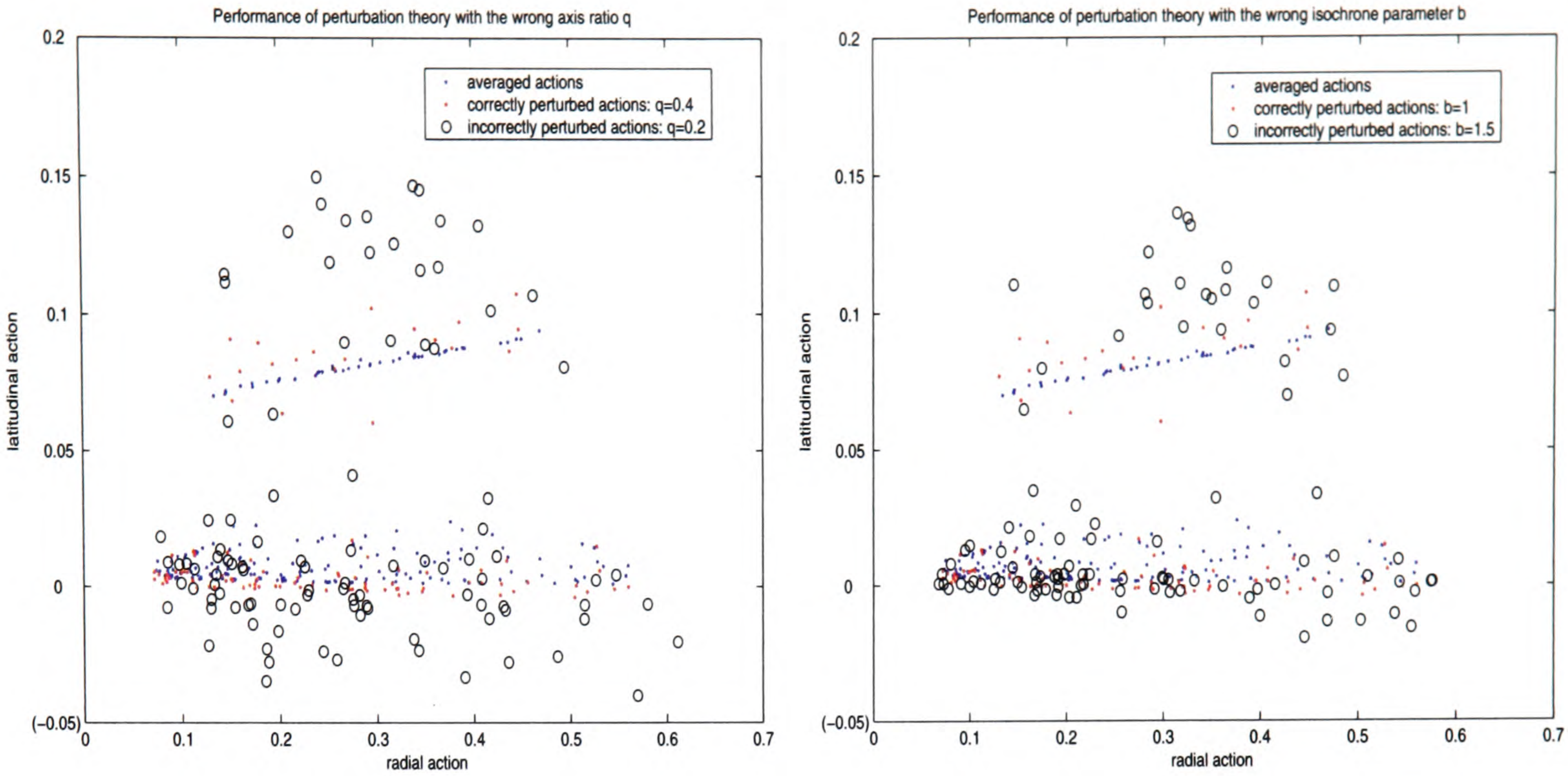

Figure 23: The performance of perturbation theory with incorrectly chosen axis ratio and isochrone parameters. Blue points are time-averaged actions, red points are correctly perturbed actions, and black circles are actions perturbed under the assumption of an incorrect q (left plot) and b (right plot). Points close to orbital pericentre, where perturbation theory is least successful, have been omitted.

we have reason to believe and are able to observe that $J'_\vartheta(J'_r(E))$ is very well approximated by a line, we can quantify the scatter produced by perturbation theory in the parameter space (b, q) by separating the points into subsets of resonantly trapped and untrapped orbits, performing a linear regression for the trapped orbits,[33] and calculating the average absolute error $\frac{1}{n}\sum_{i=1}^{n}|J_\vartheta^{i'} - (a + bJ_r^{i'})|$, for i the index of points. We see the results of this analysis in Fig. 24. The average absolute deviation in action from a linear regression of the resonantly trapped data points is given for a variety of values of the parameters (b, q). The error generated by correctly perturbing the spherical actions is emphasized. While this error is the smallest numerical value in the table, this could easily be an artefact of

---

[33] We know from our analysis of the performance of perturbation theory that untrapped orbits will not lend themselves to this analysis, the latitudinal action being too unconstrained.

	b=0.6	b=0.7	b=0.8	b=0.9	b=1.0	b=1.1	b=1.2
q=0.6	83	90	91	84	86	72	85
q=0.55	72	98	85	119	84	79	100
q=0.5	104	80	77	88	128	107	133
q=0.45	83	73	86	85	102	147	158
q=0.4	95	80	82	107	**71**	103	144
q=0.35	84	104	119	128	154	126	168
q=0.3	97	102	127	121	191	152	142

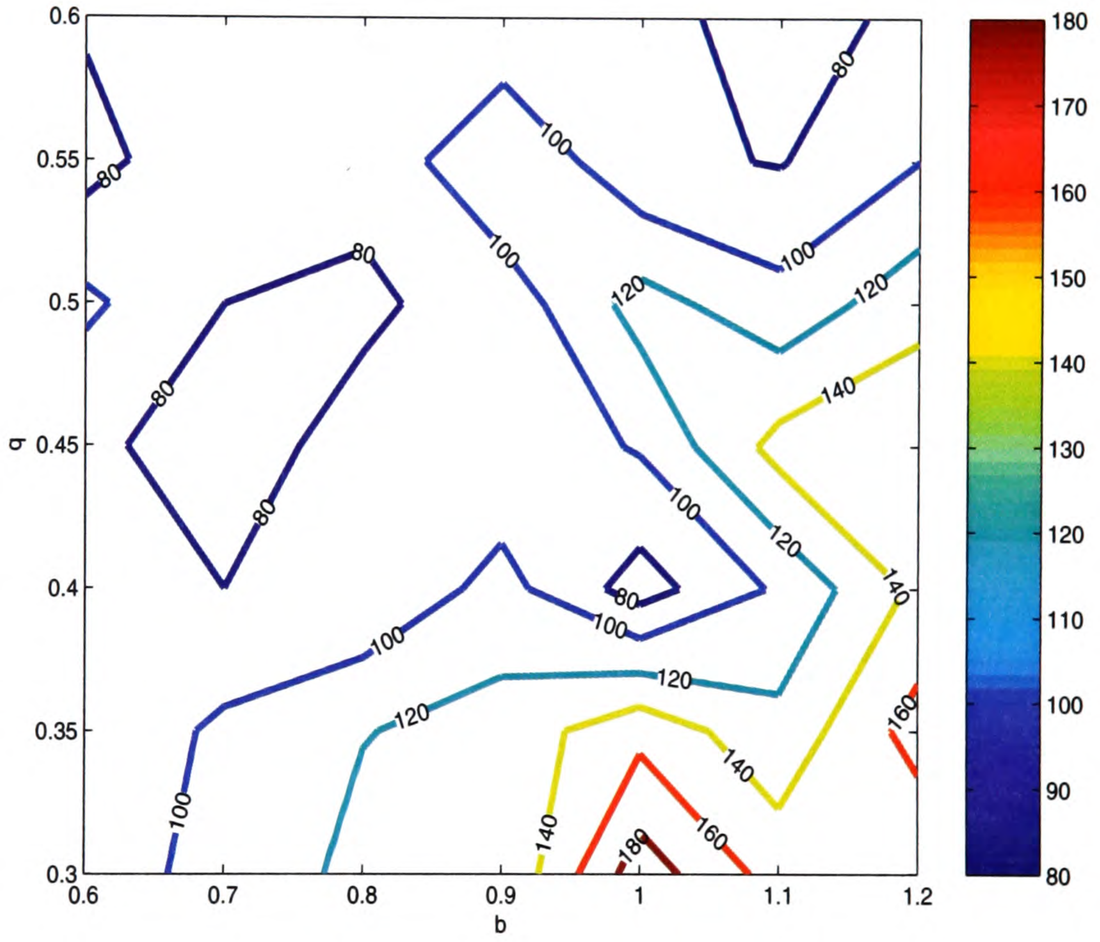

Figure 24: The application of perturbation theory to resonantly trapped orbits in order to constrain the input parameters of dynamical models. Entries in the table are the differences in perturbed latitudinal action J'_ϑ between the fit produced by linear regression and the data-points for trapped orbits. The units are 10^{-4} units of action. The error generated by correctly perturbing the spherical actions is in bold-face. The associated contour map is shown below. For the chosen range of orbital energy, pericentre and apocentre for the most eccentric orbit lie in [1.2,1.9] and [3.5,12.7] respectively.

the discretization, and the absolute minimum may well lie elsewhere. The desired effect is that correctly perturbed actions produce a prominent local minimum. Fig. 24 provides numerical evidence that this is indeed the case.[34] Perturbed frequencies provide another diagnostic of the quality of the model potential. Consider the average absolute distance in perturbed frequency ratio from resonance

$$\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\omega_r'(J^{i'})}{\omega_\vartheta'(J^{i'})}-1\right|, \tag{107}$$

for $J^{i'}$ the perturbed actions of the i-th point. For the same parameter space as above, the results are shown in Fig. 25 with the resulting contour map.

We see that the perturbed frequency ratio of resonantly trapped orbits is a good diagnostic of the correct axis ratio q, while the scatter of resonant orbits in the action plane is a useful diagnostic of the correct isochrone parameter b.

If we treat the two diagnostics as independent, then we can combine them into a single test simply by taking products of the quantities plotted. The resulting contour plot is shown in Fig. 26. We observe that we are able to use resonant dynamics to constrain the parameters of the Hamiltonian. We can make a weak claim about its isochrone parameter b and a strong claim about the true axis ratio q of the density profile for the potential.

We now have in place an adaptive method of using resonant dynamics to probe the Galactic potential:

1. Construct a set of model Hamiltonians, based upon the best available estimates of

---

[34] Averaged actions predict $a = 0.0629$, $b = 0.0641$, with $\frac{1}{n}\sum_{i=1}^{n}|J_\vartheta^{i'} - (a + bJ_r^{i'})| = 5.7 \times 10^{-4}$. Perturbation theory with the correct parameters produces $a = 0.0684$, $b = 0.0609$. The other local minimum, produced by perturbation with $b = 0.7$, $q = 0.45$, yields a very different linearity: $a = 0.0702$, $b = 0.0394$. Unfortunately, time-averaged actions will, in general, not be available to tell us what the 'right' a, b are.

	b=0.6	b=0.7	b=0.8	b=0.9	b=1.0	b=1.1	b=1.2
q=0.6	0.089	0.101	0.110	0.122	0.129	0.142	0.148
q=0.55	0.061	0.070	0.078	0.089	0.097	0.102	0.108
q=0.5	0.032	0.043	0.051	0.054	0.063	0.067	0.071
q=0.45	0.012	0.012	0.014	0.016	0.027	0.033	0.036
q=0.4	0.024	0.024	0.019	0.021	**0.017**	0.013	0.015
q=0.35	0.056	0.057	0.056	0.056	0.047	0.051	0.045
q=0.3	0.084	0.083	0.089	0.090	0.079	0.095	0.096

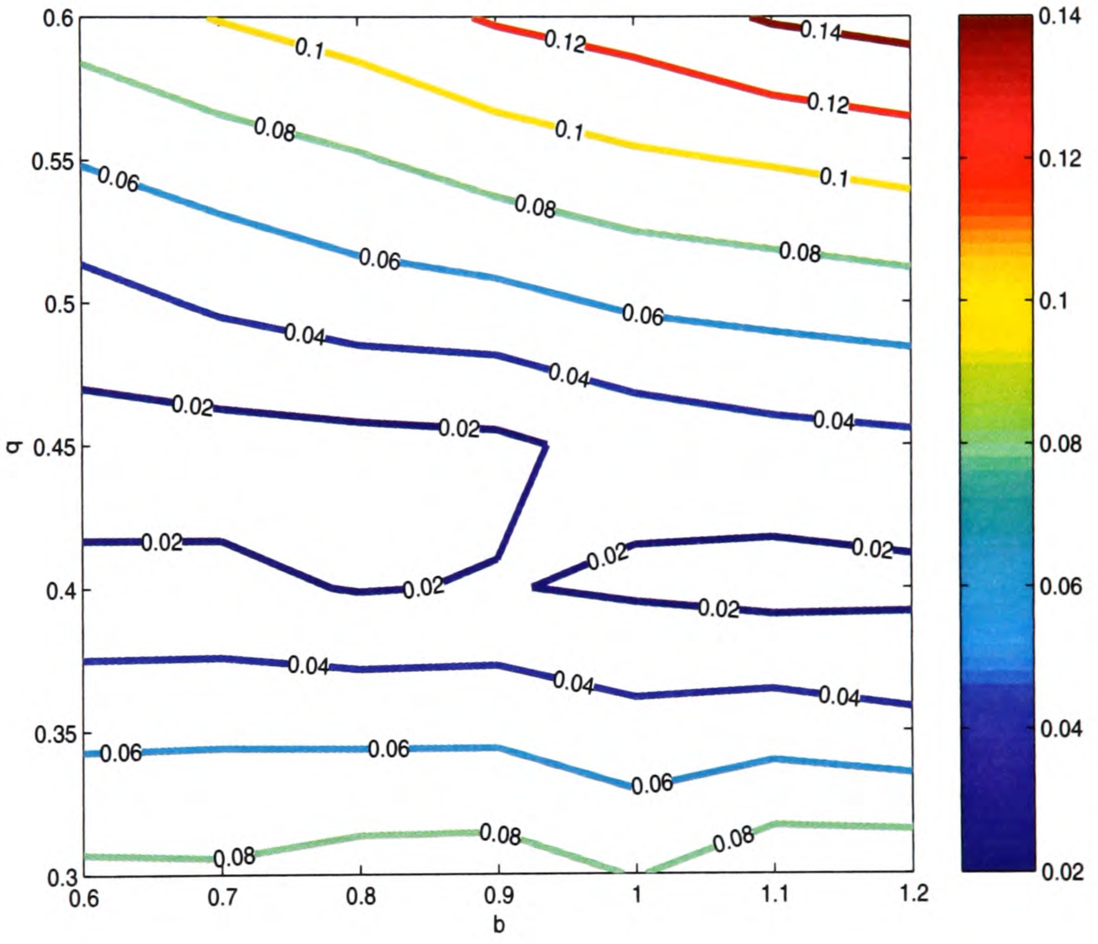

Figure 25: The deviation of the incorrectly perturbed frequency ratio for trapped orbits from the resonance. The deviation in correctly perturbing the spherical actions is in boldface. The contour map of the average absolute deviations in perturbed frequency ratio from the resonance for trapped orbits is shown below. Orbits are perturbed correctly by the choice of parameters $(b, q) = (1.0, 0.4)$.

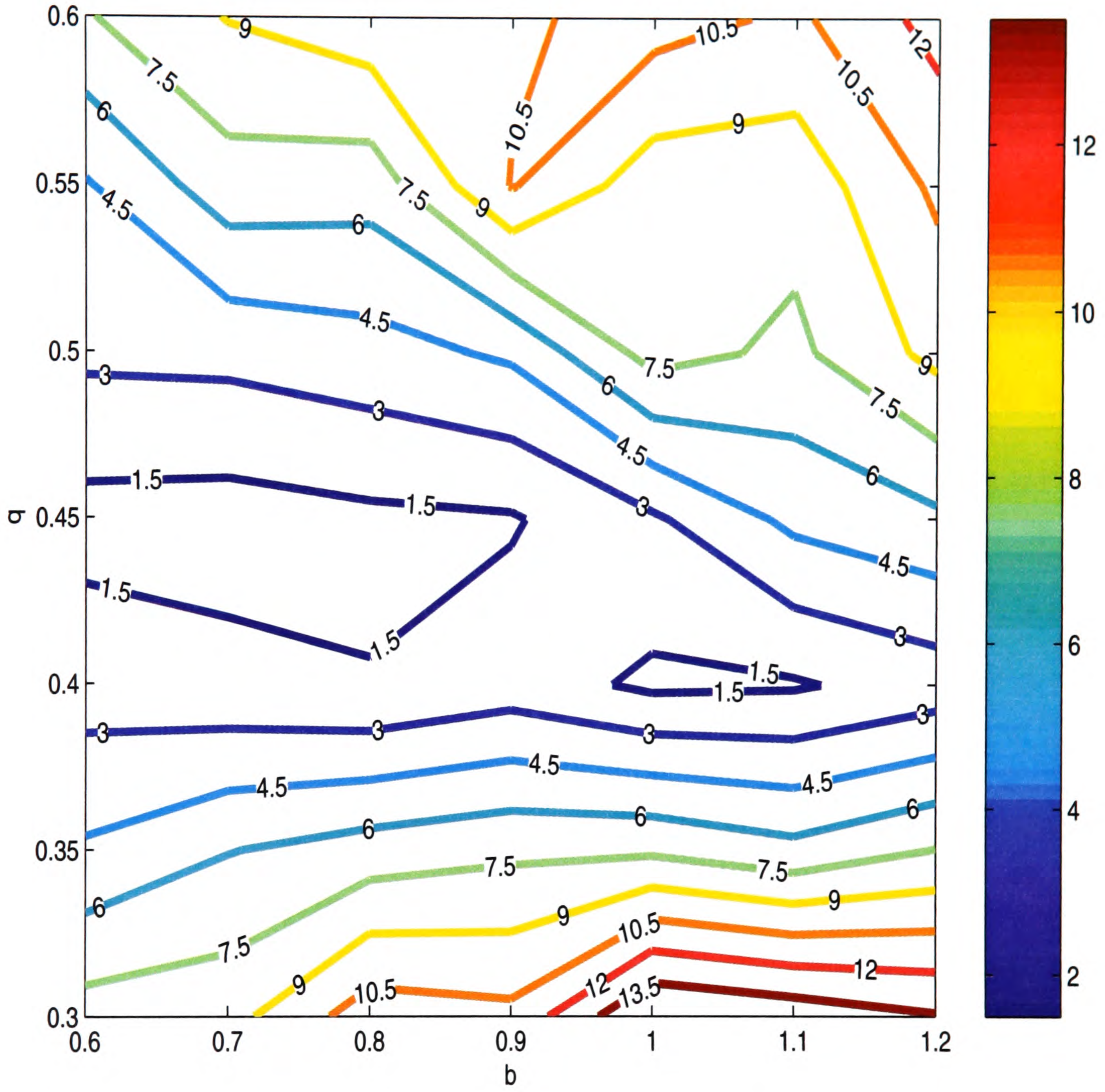

Figure 26: Contour map combining the diagnostic tests of Fig. 24 and Fig. 25. Contours of $\frac{1}{n^2}\sum_{i=1}^{n}|J_\vartheta^{i'} - (a + bJ_r^{i'})| \sum_{i=1}^{n}\left|\frac{\omega_r'(J^{i'})}{\omega_\vartheta'(J^{i'})} - 1\right|$ are plotted against (b, q). Orbits are perturbed correctly by the choice of parameters $(b, q) = (1.0, 0.4)$.

the Galactic potential, on a grid in a parameter space.

2. Investigate their spectra and construct their Laskar frequency maps to determine dynamically important resonances that are expected to trap particles.
3. In each case, plot the observational phase space data in the action space of the model Hamiltonian and identify a significant clustering of data points due to resonant trapping.
4. Simulate the resonant trapping process for the low-order resonance in question. Calculate perturbed actions for all the particles in the simulation, whether trapped or untrapped, and identify the subset of trapped particles as a cluster in action space. Confirm the accuracy of the perturbed actions by comparison with a time-averaged set of available actions.
5. For the observational data, calculate a measure of the mean deviation in perturbed frequency from resonance in the manner of Fig. 25, and find a measure of the scatter in perturbed action as in Fig. 24.
6. Select the model Hamiltonian that minimizes both these quantities. Iterate the process on a finer parameter grid.

5 The null hypothesis

In this chapter we leave aside Hamiltonian dynamics and consider the null hypothesis that a non-parametric approach based on adaptive kernel density estimation is able to resolve resonant structures in velocity phase space. We find that kernel smoothing produces density estimates that fail even to recover the global potential, let alone reveal resonant structures. Results from the numerical analysis community underpin our observations. We find we can comfortably reject the null hypothesis.

5.1 Adaptive kernel density estimation

Density estimation in several dimensions is notoriously difficult, both in terms of computational obstacles and in terms of accuracy of the estimates achieved. There are several traditional methods, notably the maximum penalized likelihood estimator (MPLE) and the kernel method. In the case of partial velocity phase space information, penalized likelihood estimators seem to outperform other methods: Dehnen (1998) was able to resolve a wealth of fine structure in the solar neighbourhood, modeling 4-dimensional data by the MPLE method. As Merritt and Tremblay (1994) point out, however, in an analysis and comparison of the kernel and MPLE methods in one and two dimensional systems, the MPLE method suffers from some drawbacks relative to kernel estimators. The MPLE has a strongly non-linear dependence on the data, such that its statistical properties are less well-known than those of the kernel estimator, and the MPLE method is potentially computationally costlier, involving the solution of a non-linear optimization problem. The principal disadvantage of the kernel method is its behaviour when there are significant gaps in the data. Weighing these pros and cons in the analysis of the specific cases of the Plummer law, the de Vaucouleurs law, and an isotropic Michie-King model, these authors find that both the MPLE and kernel methods are able to provide equally good non-parametric fits, with the MPLE tending to oversmoothed and the kernel method tending to noisy estimates. This view on the relative worth of methods of density estimation is shared by the numerical analysis community, e.g. Thompson and Tapis (1990) and Silverman (1986). One can assume from the work of these authors that the merit of the test of our null hypothesis shall not be method-dependent. The study presented here is carried out with error-free 6-dimensional data. The kernel method and an anisotropic Plummer model[35] are chosen for the non-parametric analysis.

Consider a d-dimensional Cartesian space $\mathbb{R}^d$ and n particles in that space with a configuration $X' \in \mathbb{R}^{d \times n}$. We denote the location of the i-th particle by X'_i. The problem ostensibly solved by the kernel method is how to recast the raw density profile

$$f(x') = \frac{1}{n}\sum_{i=1}^{n} \delta(x' - X'_i), \tag{108}$$

for $x', X'_i \in \mathbb{R}^d$, into a smooth, differentiable form by convolution with appropriate kernels. A kernel is a function $K(x')$ satisfying a normalisation condition $\int_{\mathbb{R}^d} K(x')\, d^d x' = 1$. The multi-dimensional kernel method has two main variants: kernels separable on a rectangular coordinate system (product kernels) and kernels with spherical symmetry in some suitable

---

[35] The first obstacle to a non-parametric implementation is the choice of model. Experimentation with Osipkov-Merritt models revealed that the adaptive kernel method can raise significant difficulties relating to a suitable choice of window width (i.e. smoothing parameter) in distributions with large central densities and rapidly tailing wings, an observation supported by Merritt and Tremblay, who found the kernel method more suited than the MPLE to tackling Plummer and Michie-King models, and the MPLE more suited than the kernel method to modeling de Vaucouleurs profiles. Once it had been decided to pursue the kernel method, therefore, it was imperative to choose a type of model most suited to the kernel approach.

space. Since the Cartesian coordinates are convenient rather than unique to the system, it is preferable to use spherically symmetric kernels.

Any method for modeling a general density distribution must be adaptive, i.e. the amount of smoothing must be greater in regions of lower density. Spherically symmetric d-dimensional kernels are to be chosen, so a general distribution must first be made isotropic. For a hyperellipsoidal distribution function, this is achieved by mapping the space onto one in which the covariance matrix of the points is a unit matrix, a process known as *whitening*. In the new frame, the covariances of the points are 0 and the variances, the diagonal elements of the covariance matrix, are 1. For X, $X' \in \mathbb{R}^{d \times n}$, the matrices of the coordinates X_i and X'_i in the new and old spaces respectively,

$$X_i = D^{-\frac{1}{2}} C^T X'_i, \tag{109}$$

where D is the diagonal matrix of eigenvalues and C the matrix of normalised eigenvectors of the covariance matrix of the points, S':

$$S' = \frac{1}{n} \sum_{i=1}^{n} X'_i X'^T_i. \tag{110}$$

When C and D are chosen in this way, the covariance matrix in the new space is

$$S = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T = \frac{1}{n} \sum_{i=1}^{n} D^{-\frac{1}{2}} C^T X'_i (D^{-\frac{1}{2}} C^T X'_i)^T = D^{-\frac{1}{2}} C^T S' C D^{-\frac{1}{2}} = I \tag{111}$$

as required (since $C^T S' C = D$).[36]

---

[36] The transformation chosen above is not unique in obtaining a unit covariance matrix. In particular,

In the new frame, points are heavily constrained to lie in a d-dimensional cube of side length 2 centred on the origin, and the density of points varies isotropically. In the non-iterative, non-adaptive Fukunaga method (Fukunaga, 1972), this transformation is performed temporarily, spherically-symmetric kernels are chosen, and the transformation is inverted, such that the kernels have the appropriate ellipsoidal symmetry in the original space. The basic principle of the Fukunaga method is thus to choose the symmetry of the kernels to be the same as that of the overall distribution of points.

In our case, adaptivity is required, however, so the overall best choice of method is the adaptive kernel method applied to the whitened data. Only slightly different from Fukunaga, here one transforms the points into the whitened frame and applies the method outlined below for the required number of iterations, before transforming back.

Once in the whitened frame, the adaptive kernel method proceeds as follows. One first finds a pilot estimate of the distribution function $\tilde{f}$, satisfying $\int_{\mathbb{R}^d} \tilde{f}\, d^d x = 1$ and $\tilde{f} \geq 0$, $\forall x$. The adaptive kernel is said to be insensitive to the choice of pilot estimate (Silverman, 1986, including the principles that follow), so cross-validation or other more complex methods are not necessary in its choice. A simple fixed kernel pilot estimate is used here:

$$\tilde{f}(x) = \frac{1}{n h_{\mathrm{fix}}^{d}} \sum_{i=1}^{n} K\left(\frac{1}{h_{\mathrm{fix}}}(x - X_i)\right). \tag{112}$$

Two choices need to be made regarding the estimate: the type of kernel K and the size of

---

a distribution can be *sphered* by the transformation $X_i = S'^{-\frac{1}{2}} X_i'$. Noting that $S'^{-\frac{1}{2}} = C D^{-\frac{1}{2}} C^T$, this is just a mapping of the distribution used here by C. Because in general $\det C \neq 1$, this is not a proper rotation of the whitened distribution, but a rotation and a scaling.

the optimal *smoothing length*, or *window width* h_{fix} of the kernel. The Epanechnikov kernel

$$K(x) = \begin{cases} \frac{1}{2V_d}(d+2)(1-x^Tx), & \{x : x^Tx \leq 1\} \\ 0 & \{x : x^Tx \geq 1\} \end{cases} \tag{113}$$

with V_d the volume of the unit d-ball, is the optimal kernel in terms of minimisation of the mean integrated square error between the estimated and actual distribution functions (Epanechnikov, 1969). For whitened data, there exists an optimal smoothing length in the non-adaptive fixed kernel method $h_{\text{fix}}(K, n, d)$. For the Epanechnikov kernel,[37] $h_{\text{fix}} = 2.49n^{-\frac{1}{d+4}}$.

The method now proceeds to find the geometric mean of the pilot estimates of the density at the sample locations and to scale h_{fix} by factors derived from the ratio of the local density at each point and the geometric mean. These *bandwidths* λ are given by $\lambda_i = \left(\frac{\tilde{f}(X_i)}{g}\right)^{-\alpha}$, for $\log g = \frac{1}{n}\sum_{i=1}^{n} \log \tilde{f}(X_i)$. The quantity $\alpha \in [0, 1]$ is the *sensitivity parameter*. For $\alpha = \frac{1}{d}$, there would be an equal number of particles in each kernel. For $\alpha = 0$, we recover the fixed kernel method. According to Silverman and other authors (Silverman, 1986 and references therein) a choice of sensitivity parameter $\alpha = \frac{1}{2}$ tends to be best. There is no mechanical means of choosing α, but Abramson (1982) showed that a value of $\alpha = \frac{1}{2}$ produces an estimate with a bias smaller than that of the fixed kernel estimate. See later for a discussion on the sensitivity parameter.

In the next step of the adaptive method, the bandwidths chosen above are used to yield

---

[37] This is not necessarily similar to the optimal smoothing length evaluated in the final adaptive estimate.

the adaptive density estimate

$$\hat{f}(x;h) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{(h\lambda_j)^d} K\left(\frac{x - X_j}{h\lambda_j}\right), \tag{114}$$

where the new smoothing length $h = \hat{h}_{\text{opt}}$ is chosen by *cross-validation* (Silverman, 1986) as follows. The mean integrated square error between the actual density f and the model density $\hat{f}$ is given by the expectation value

$$E\left[\int (\hat{f} - f)^2 \, \mathrm{d}^d x\right] = \int f^2 \, \mathrm{d}^d x + E\left[\int (\hat{f}^2 - 2\hat{f}f) \, \mathrm{d}^d x\right]. \tag{115}$$

In the 1-dimensional fixed kernel case, the expectation value is defined by the convolution $E\hat{f}(x) = \frac{1}{h}K' * f$, for $K'(x) = K(\frac{x}{h})$, i.e. it represents summing the kernels over a continuum of sample points distributed according to the actual distribution function. In the d-dimensional adaptive case, this implies that

$$E\hat{f}(x) = \int_{\mathbb{R}^d} \frac{1}{(h\lambda(y))^d} K\left(\frac{x - y}{h\lambda(y)}\right) f(y) \, \mathrm{d}^d y \tag{116}$$

with $\lambda(y)$ the continuous analogy of the discrete λ_i. The mixed term in the expression for the mean integrated square error can be rewritten as follows, with $\hat{f}_{-n}$ the density estimate with the n-th particle removed:

$$\begin{aligned} E\left[\int \hat{f}(x) f(x) \, \mathrm{d}^d x\right] &= E\left[\int \hat{f}_{-n}(x) f(x) \, \mathrm{d}^d x\right] \\ &= E\hat{f}_{-n}(X_n) \end{aligned}$$

$$= E\frac{1}{n}\sum_i \hat{f}_{-i}(X_i). \qquad (117)$$

The first equality arises from the fact that $E(\hat{f})$ and therefore also the expectation value of the space average of $\hat{f}$ are dependent on the choice of kernel and window width, but independent of sample size. The second equality is a statement that the expectation value over many samples is the same for both a) the space-averaged model density with one particle removed and b) the model density at the location of the removed particle. Without removing one particle, it would not be possible to make this claim: the average density at the location of the particles is clearly not representative of the overall density. The final equality is the statement that the average over all samples of the density at a given point is equivalent to the average over all samples of the average density at many points, as long as these points do not coincide with the locations of particles.

If one defines the score function by

$$M_0(h) = \int \hat{f}^2 \, \mathrm{d}^d x \; - \frac{2}{n}\sum_i \hat{f}_{-i}(X_i), \qquad (118)$$

then the value of h that minimizes $EM_0(h)$ also minimizes the mean integrated square error. Unfortunately, one cannot calculate $EM_0(h)$ since this requires knowing f, but if $\hat{h}_{\text{opt}}$, the value of h that minimizes $EM_0(h)$, is close to that which minimizes $M_0(h)$, then an approximation to the optimal smoothing parameter can be obtained by minimizing the score function. Using Monte-Carlo integration (using $\hat{f}$ to sample $\hat{f}^2$) and writing the summation in terms of $\hat{f}$, the score function is simplified here as follows:

$$\hat{f}_{-i}(x) = \frac{1}{n-1}\sum_{j\neq i}\frac{1}{(h\lambda_j)^d}K\left(\frac{x-X_j}{h\lambda_j}\right)$$

$$\hat{f}_{-i}(X_i) = \frac{1}{n-1}\left[n\hat{f}(X_i) - \frac{1}{(h\lambda_i)^d}K(0)\right]$$

such that

$$M_0(h) \approx \sum_i \left[\left(\frac{1}{n} - \frac{2}{n-1}\right)\hat{f}(X_i;h) + \frac{2}{n(n-1)(h\lambda_i)^d}K(0)\right]. \tag{119}$$

Then $M_0(\hat{h}_{\rm opt}) = \min_h M_0(h)$. Together $\hat{h}_{\rm opt}$ and λ now yield an adaptive kernel estimate of the distribution function. Further iterations are possible by re-running the code with the optimal smoothing length to get a new set of bandwidths and a new score function. The density estimate is scaled by $|\det(D^{-\frac{1}{2}}C^T)|(\tilde{f},\hat{f})$ to invert the process of whitening, where the Jacobian, unsurprisingly, is found to be the determinant of the matrix that performs the whitening process.

In summary, the adaptive kernel method can be outlined as follows:

1. A configuration of particles X'_j, $j = 1,\ldots,n$ with covariance matrix S' produces a raw density profile $f(x') = \frac{1}{n}\sum_i \delta(x' - X'_j)$.
2. Generate whitened space x by $x' \overset{M}{\mapsto} x$, such that the covariance matrix $S' \overset{M}{\mapsto} S = I$.
3. Choose a normalized kernel K, and for $h_{\rm fix}(K,n,d)$ a known function, construct the fixed kernel pilot estimate $\tilde{f}(x) = \frac{1}{nh_{\rm fix}^d}\sum_j K\left(\frac{x-X_j}{h_{\rm fix}}\right)$.
4. Calculate the bandwidths $\lambda_j(\tilde{f})$ and the adaptive estimate $\hat{f}(x;h) = \frac{1}{n}\sum_j \frac{1}{(h\lambda_j)^d}K\left(\frac{x-X_j}{h\lambda_j}\right)$. Use $h = h_{\rm fix}$ on the first iteration.

5. Construct the score function $M_0(\hat{f}(X;h),h)$ and find $\hat{h}_{\text{opt}} : M_0(\hat{h}_{\text{opt}}) = \min_h M_0(h)$. Return to step 4 with $h \to \hat{h}_{\text{opt}}$ for as many iterations as required.

6. Perform $x \overset{M^{-1}}{\mapsto} x'$, $\hat{f}(x) \mapsto \hat{f}'(x')$.

Now we consider the special case $d = 6$. One option, clearly, is to model the velocity phase space density as one 6-dimensional whole. This mechanizes the problem and eliminates the need for dynamical considerations, but there are several important disadvantages that make this likely to fail. First, the kernel method provides us only with the average density of a given number of points in a given volume. For a domain in $\mathbb{R}^d$ of volume L^d and a total number of points n, the kernel contains $N \propto \left(\frac{h}{L}\right)^d n$ particles, so $h \propto \left(\frac{N}{n}\right)^{\frac{1}{d}} L$ grows rapidly as d increases. We are specifically interested in resonant, typically small-scale structures, and the large window widths are likely to make their resolution difficult.

Second, there are geometric considerations. We know that integrable orbits will lie in 3-tori, with orbits in axisymmetric and spherical potentials even more heavily constrained. While the linear algebra of the whitening process cannot take us to this action-angle space, the underlying orbital geometry should be partly visible in the Cartesian phase space. Spherical potentials, for example, admit distribution functions with O(3) symmetry in real space, but O(2) in velocity space, and these sub-symmetries would be less easily noticeable in a full phase-space density estimate. Points nearby in velocity phase space are nearby in the configuration and velocity subspaces also, but the converse is not true in general, so if one wishes to average over physically similar regions, a splitting of the phase space becomes necessary.

Third, computational factors speak against density estimation in a 6-dimensional space.

In the simplest implementation, one establishes cubic cells and locates the particles in adjacent cells over which smoothing needs to take place. The number of nearest neighbour cells will scale prohibitively as 3^d.

Finally, there is the "curse of dimensionality", a well-known phenomenon in multi-dimensional density estimation and a subject to which we will return later. The crux of the matter is the sparsity of points. Consider that the volume ratio of the cube to an inscribed sphere is

$$\frac{V_d}{l^d} = \frac{\pi^{\frac{d}{2}}}{d2^{d-1}\Gamma(\frac{d}{2})} \tag{120}$$

For $d = 3$, over half the cube lies in the sphere. By contrast, for $d = 6$, the ratio is 0.08, the rest of the volume lying in the 2^d corners of the cube. As d increases, the particles in a uniform distribution accumulate in a thinner shell bounding the surface of the sphere. The growth with d in sample size needed to compensate for the increase in dimension was estimated by Epanechnikov (1969) and others to be at least exponential.

Because of these difficulties, we split the distribution into separate configuration and velocity space components, $\rho(x)$ and $\rho_v(\dot{x}, x)$, in which the x dependence of the velocity space density arises from the inseparability of a general distribution function. This is by itself free from dynamical assumptions, but the sample size required grows as the configuration space volume over which one averages the velocity space density is reduced, such that the analysis of the velocity subspace is somewhat subjective.

5.2 Tests of the algorithm

The velocity phase space points are sampled from the anisotropic Plummer model

$$f(E,L) = \begin{cases} \frac{6}{(2\pi)^3}(2E - L^2)^{\frac{3}{2}}, & L^2 \leq 2E \\ 0 & L^2 \geq 2E \end{cases} . \tag{121}$$

In real space, the density and potential of the resulting model are (Dejonghe, 1986)

$$\rho(r) = \frac{3}{4\pi}(1 + r^2)^{-\frac{5}{2}} \tag{122}$$

$$\Phi(r) = (1 + r^2)^{-\frac{1}{2}}. \tag{123}$$

The distribution function is sampled separately for the real and velocity subspaces. To provide a complete velocity phase space density distribution, the velocity space distributions are found by normalised integration of f over velocity space and are as such functions also of r.

In the real space case, the fraction of points contained within radius ϱ is given by

$$\chi(r) = \int_0^r 3\varrho^2(1 + \varrho^2)^{-\frac{5}{2}}\, d\varrho, \tag{124}$$

which is easily inverted to give r as a function of the random variable $\chi \in [0, 1]$:

$$r(\chi) = (\chi^{-\frac{2}{3}} - 1)^{-\frac{1}{2}}. \tag{125}$$

The velocity space case is a little more complicated. The velocity space density is given by the normalised integral of the velocity phase space density over velocity space: $f(r, v_r, v_t) = \rho_v(v_r, v_t; r)\rho(r)$, s.t. $\int \rho_v \, d^3v = 1$, for v_r and v_t the radial and transverse velocities. Rewriting $f(E, L)$ in terms of r, v_r, v_t we find, noting the convention that $E = \Phi - \frac{1}{2}(v_r^2 + v_t^2)$,

$$\rho_v(v_r, v_t; r) = \begin{cases} \frac{6}{(2\pi)^3\rho(r)}(2\Phi(r) - v_r^2 - (1 + r^2)v_t^2)^{\frac{3}{2}}, & \left\{0 \leq v_t \leq \left(\frac{2\Phi(r) - v_r^2}{1 + r^2}\right)^{\frac{1}{2}}, \right. \\ & \left. -\sqrt{2\Phi(r)} \leq v_r \leq \sqrt{2\Phi(r)}\right\} \\ 0 & \text{otherwise.} \end{cases} \tag{126}$$

The velocity space density distribution is sampled analogously to the real-space case. The method is described in Appendix A.

The kernel method requires the calculation of the distances between all combinations of points in the space. For each $j = 1, \ldots, n$, all points X_i satisfying $|X_i - X_j| \leq \lambda_j \hat{h}_{\mathrm{opt}}$ contribute to the density $\hat{f}$. Because this calculation is $O(n^2)$ it would be prohibitively expensive in the case of a sufficiently large dataset. The simplest computational solution is to divide space into cells of side length $\delta = h_{\mathrm{fix}}$ in the fixed-kernel case and $\delta = \hat{h}_{\mathrm{opt}} \max_i(\lambda_i)$ in the adaptive kernel case. Then distances need be calculated only for particles that lie within the relevant cell or within neighbouring cells. For each cell, a linked list data structure is used to enumerate the particles within it.[38]

Having calculated the bandwidths, the values of the score function $M_0(h)$ are found for a range of smoothing lengths and the minimum value is chosen. The minimum value

---

[38] Tukey & Tukey (1981) point out that the covariance matrix of the distribution is not robust to outlying points and recommend the removal of outliers by a method such as the *five percent rule*. More complicated methods of making the covariance matrix robust exist (see Mosteller and Tukey (1977)). Outlying points do not contribute significantly to the density, although their importance does depend on the size of the pilot smoothing length chosen. All the points are nevertheless sampled from the same distribution function and are therefore significant. As such, we restrict ourselves to assigning to points classified as out of bounds a local density due to themselves only, allowing these outliers to contribute to the calculation of the bandwidth factors, and more importantly, maintaining them in the covariance matrix.

obviously only equates to the mathematical minimum if the gradient is indeed zero. The score functions are found to yield clear and reproducible minima, and the validity of the method can be tested by a simple examination of a plot of $M_0(h)$ versus h.

5.3 Results

It was discovered that using h_{fix} as a pilot smoothing length in the initial fixed kernel estimate leads to a score function with no resolvable stationary point. The stationary point of the score function is expected to arise because as h increases, the contribution to the sum from the additional points included in the kernel rises faster than the $\frac{1}{(\lambda_i h)^3}$ factor moderating the summand in Eq. 114. Eventually, as the kernel size increases beyond the minimizing value, insufficient additional points are added to the kernel to offset the increase in this factor. The summation provides an average of this condition across the distribution. (Note that the $[\frac{1}{n} - \frac{2}{n-1}]$ factor is negative, leading to a minimum). When $h_{\mathrm{fix}} \gg h_{\mathrm{opt}}$, however, the pilot estimate is thoroughly oversmoothed, with much of the density detail lost. In this case, the bandwidths λ are much larger than for $h_{\mathrm{fix}} \approx h_{\mathrm{opt}}$. The kernels therefore contain points for significantly smaller h and the minimum of the score function is much closer to $h = 0$. In our case, h_{fix} oversmoothed the data to the point where the minimum of the score function was unresolvable. We were thus unable to use h_{fix} and chose an empirical initial value instead.[39]

Significantly, the minimum of the score function is found to be reproducible with differ-

---

[39] It was also found by calculation of points in the vicinity of the global minimum of the score function that smaller features of no global significance appear. Such features are attributable to statistical fluctuations. In light of this, it would seem to be inappropriate to do a more precise analysis, such as by Newton's method, of the minimizing smoothing length.

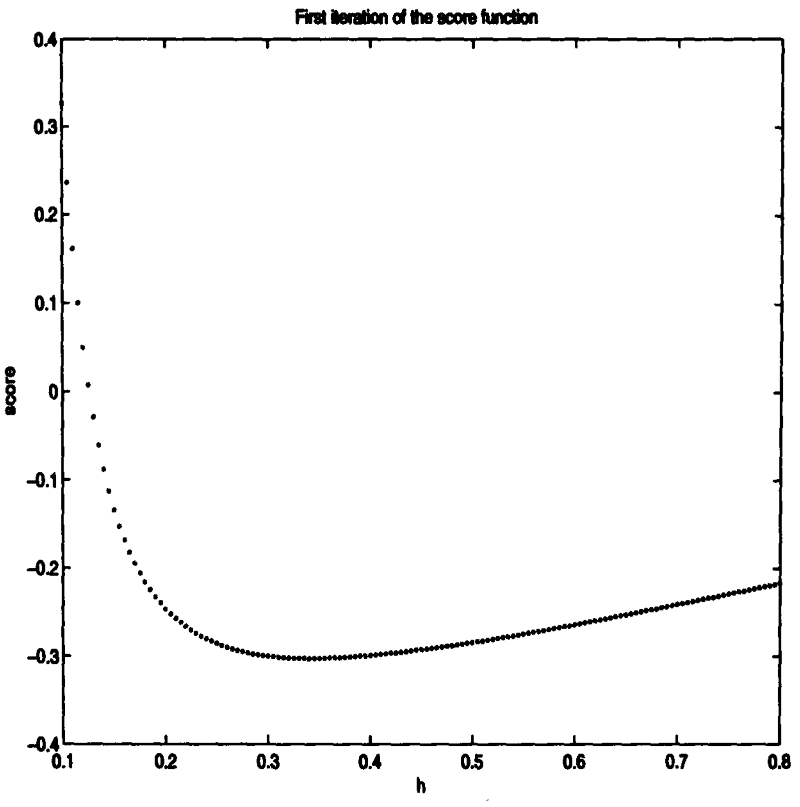

Figure 27: The score function for real-space density estimation in an iteration of the adaptive kernel method. The minima in four interations are found to be smoothing lengths of 0.340, 0.360, 0.365, and 0.3650 respectively.

ent sets of points sampled from the distribution function, and with subsequent iteration of the method, i.e. by substitution of the optimal smoothing length to produce a new fixed kernel estimate. This is seen in Fig. 27. Where there are discrepancies between optimal smoothing lengths in subsequent iterations, the minimum value tends to be an outlier, and the global function has the expected minimum. In the real space case, the scaled density estimates are now graphed (Fig. 28) together with the actual density $\rho(r)$ against r.[40]

Fig. 28, a selection of the results obtained, shows that sometimes the fit is not good for small r, but often the fit is reasonable throughout the domain. The quality of the adaptive approximation is found to improve on the fixed kernel estimate for a suitably chosen sensitivity parameter. For a sensitivity parameter of $\alpha = \frac{1}{2}$, the resulting adaptive

---

[40] Because the integration has not been performed over the sphere, the areas under these curves do not sum to one, but the three graphs are precisely analogous.

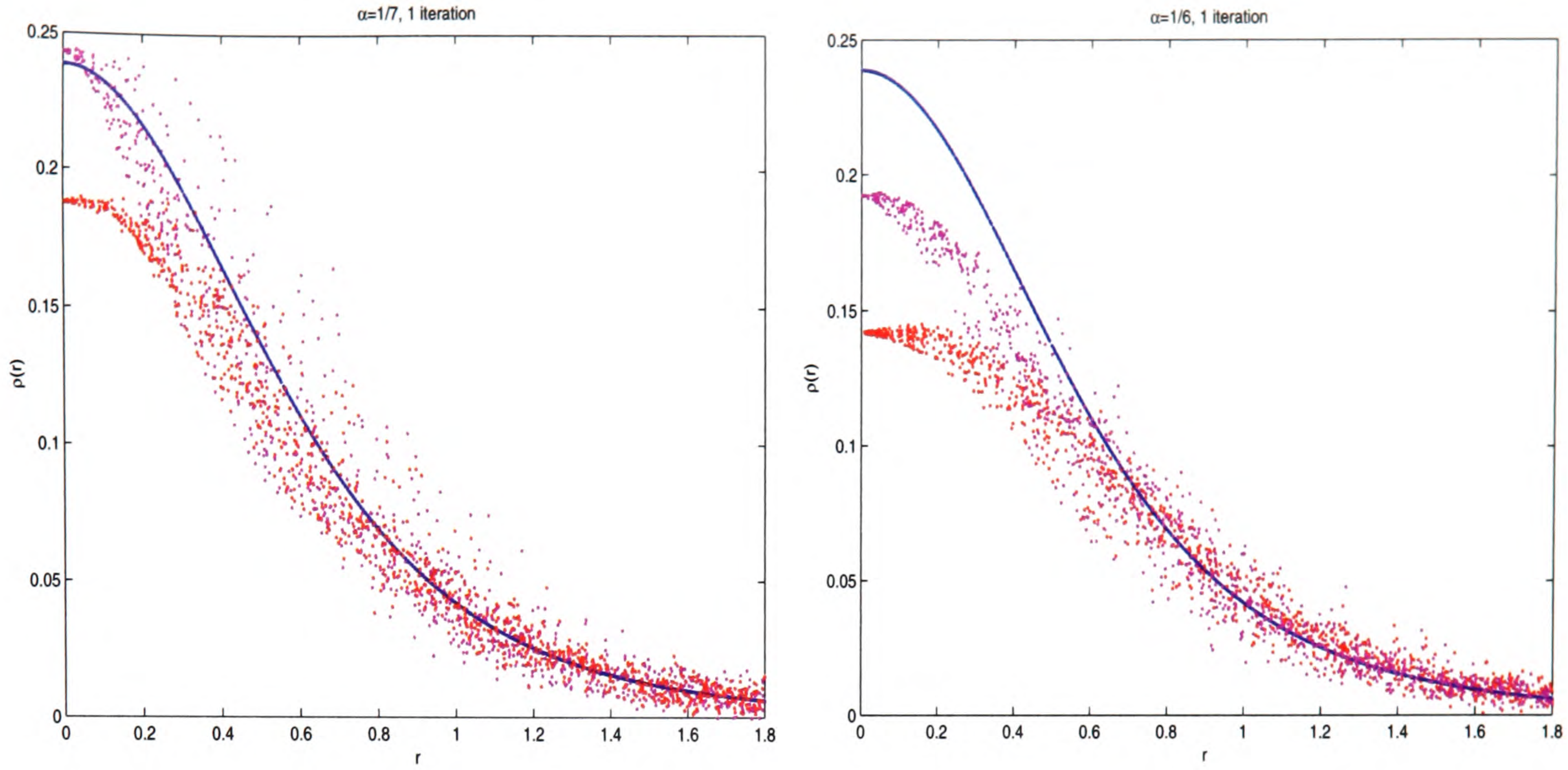

Figure 28: Fixed kernel (red) and adaptive kernel (magenta) estimates of the density profile (blue) for two real space distributions with 1000 points, illustrating a small advantage of the adaptive over the fixed kernel estimates with varying sensitivity parameter. Spherical symmetry was not assumed.

density estimate is essentially useless, fluctuating wildly near the origin. It does not improve with iteration. In contrast to the observations of Silverman, here it is found that a sensitivity parameter of $\alpha = \frac{1}{7}$ yields the most plausible adaptive estimate. Since $h_{\text{fix}} \propto n^{-\frac{1}{7}}$, it is found that $\lambda_i h_{\text{fix}} \propto \left(\frac{\tilde{f}_i n}{g}\right)^{-\frac{1}{7}}$. Comparison of Fig. 29 with Fig. 28 reveals that with larger sample sizes, the advantage of the adaptive over the fixed kernel method diminishes considerably.

The situation in the full velocity phase space model is not satisfying. In velocity space, the score function has a fairly clear minimum, but it is found that taking a different smoothing length (in one instance $\frac{4}{3}\hat{h}_{\text{opt}}$) provides a better estimate (see Fig. 30). For particles at small radii and for particles at large radii, the estimate is not good, and the adaptive approximation is not found to improve the estimate significantly. Without the manipulation of $\hat{h}_{\text{opt}}$, the model fails for all tested values of the sensitivity parameter. Crucially, the

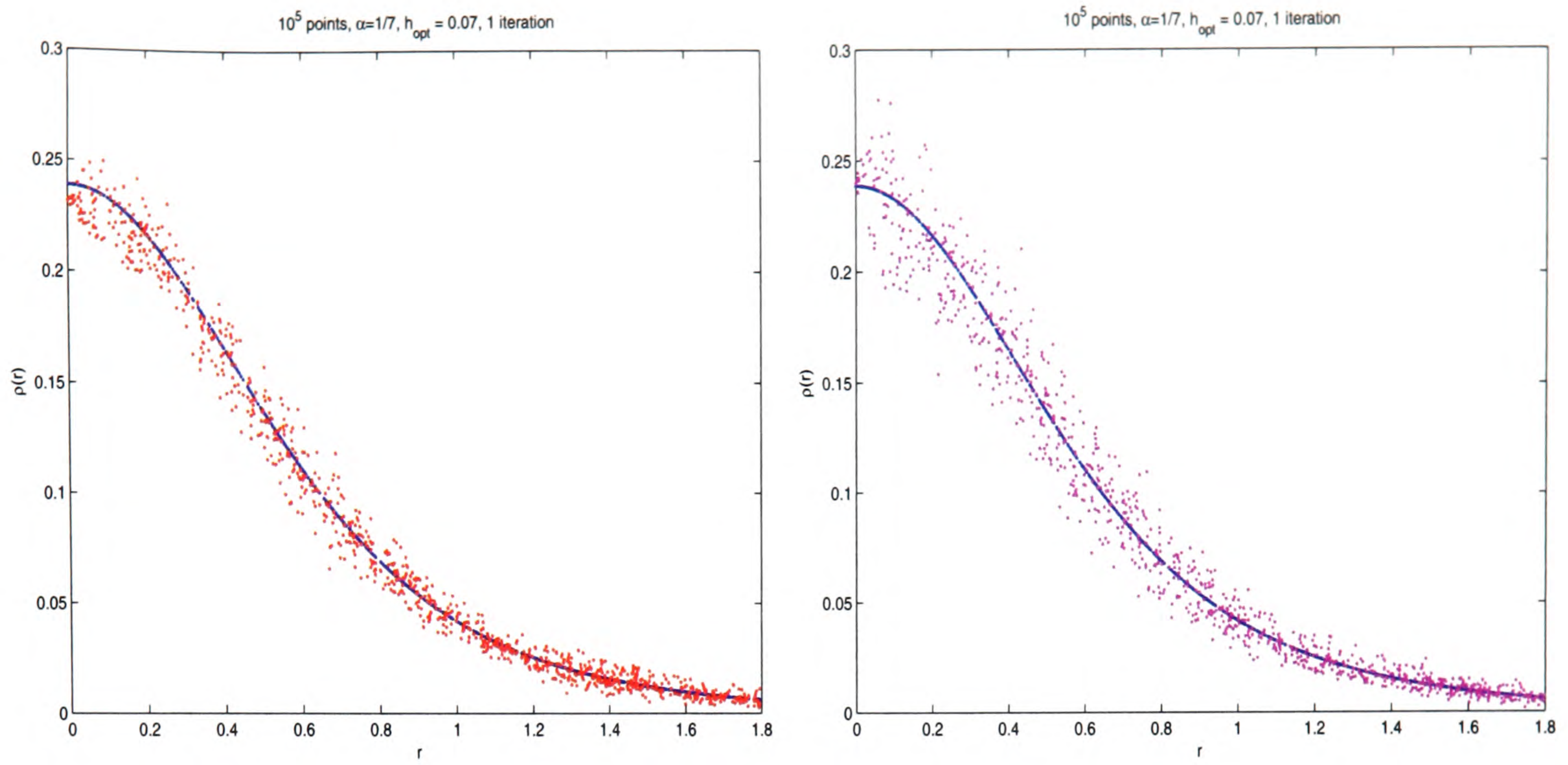

Figure 29: Fixed kernel (red) and adaptive kernel (magenta) estimates of the density profile (blue) for a realisation of the real space distribution function with 10^5 points.

value of $\hat{h}_{\text{opt}}$ corresponds to smoothing over as much as 40% of the data points, so that the approximate accordance of the model and the fit is likely to be statistically insignificant, and we can safely conclude that resonant structure will be unresolvable.

In analysing these results, it is important to remember that a number of sources of inaccuracy cloud the velocity space density estimate. The points only approximately sample the velocity density distribution, since there are significant differences in radius between the points. Second, since the method is not optimized to the specific problem of axisymmetry in velocity space, the statistical fluctuations in v_r, v_ϑ, and v_φ are compounded. Third, sample sizes are small. The graphical evidence suggests a large smoothing parameter of $0.7 < h < 1$ for samples of order 10^3 in velocity space, and smoothing lengths of $0.07 < h < 0.08$ for the real space distribution with 10^5 points.

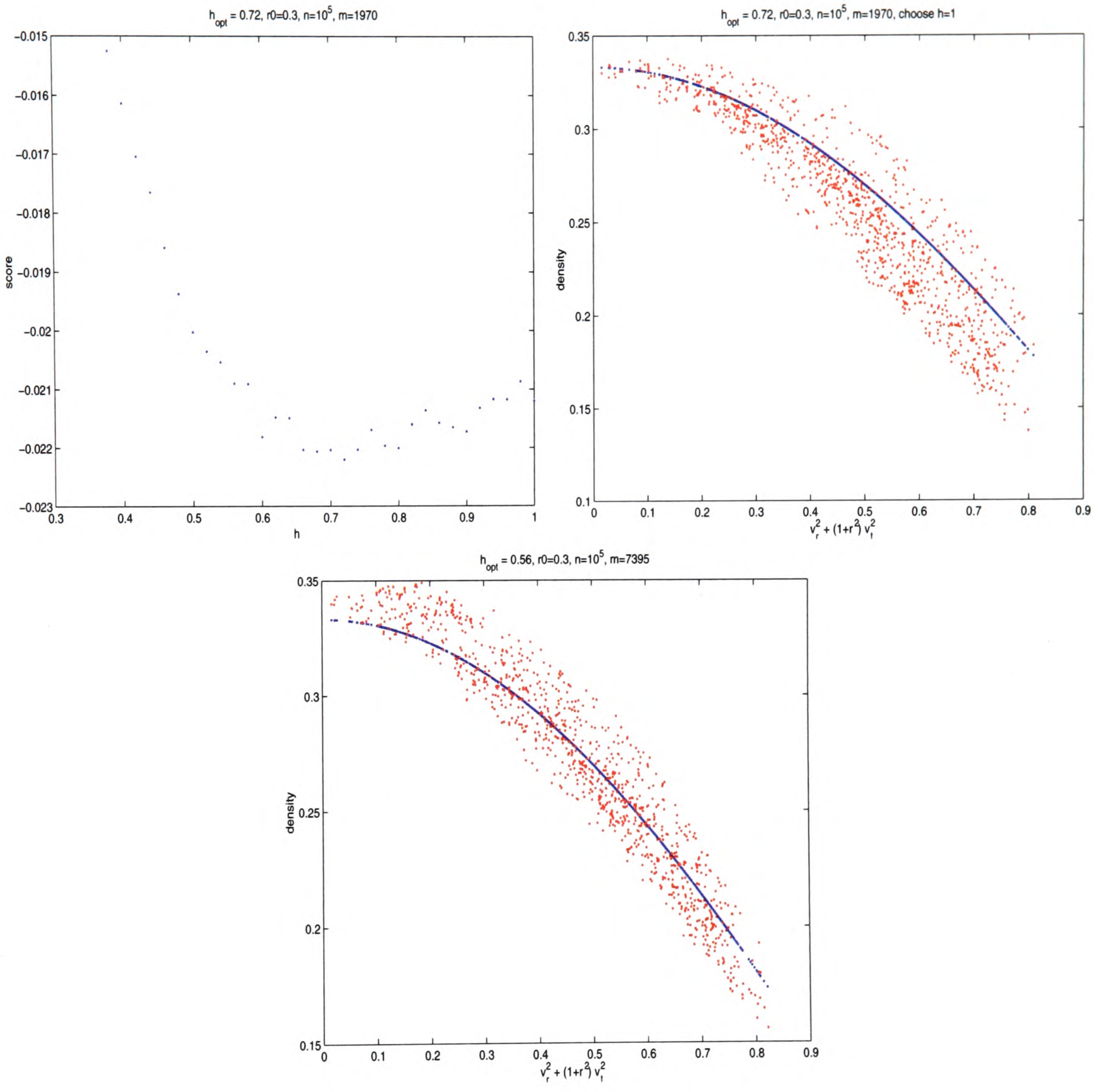

Figure 30: Score function and fixed kernel estimates of the velocity distribution at radius $r = 0.3$, with 1970 points chosen from the sample of 10^5, s.t. $0.2466 < r < 0.3414$, and 7395 points s.t. $0.0094 < r < 0.4619$ respectively.

5.4 The null hypothesis rejected

The distribution function $f(E, L)$ uniquely determines a spherically symmetric gravitational potential $\Phi(r)$. Can our recovered $\hat{f}$ be used to estimate $\Phi(r)$? Under the assumption of the Jeans theorem, one can equate $f(r, v_r, v_t) = f_2(E(\Phi), L)$, $E = \Phi(r) - \frac{1}{2}(v_r^2 + v_t^2)$ to obtain an expression for $\Phi(r)$:

$$\begin{aligned} \frac{\partial f}{\partial r} &= \frac{\partial f_2}{\partial E}\frac{d\Phi}{dr} + \frac{\partial f_2}{\partial L}v_t && (127) \\ \frac{\partial f}{\partial v_r} &= -\frac{\partial f_2}{\partial E}v_r && (128) \\ \frac{\partial f}{\partial v_t} &= -\frac{\partial f_2}{\partial E}v_t + \frac{\partial f_2}{\partial L}r. && (129) \end{aligned}$$

Combining these,

$$\frac{d\Phi}{dr} = \frac{\frac{\partial f}{\partial r} - \frac{v_t}{r}\left(\frac{\partial f}{\partial v_t} - \frac{v_t}{v_r}\frac{\partial f}{\partial v_r}\right)}{\frac{1}{v_r}\frac{\partial f}{\partial v_r}}. \tag{130}$$

Although the distribution function obtained by the kernel method is mathematically smooth, we see plainly in Fig. 29 and Fig. 30 that its derivatives oscillate significantly and uncorrelatedly, so that a direct substitution of the model fit is impossible.[41] One could of course find a least squares fit to the global behaviour of the model and differentiate this as required, but the performance of the density estimation does not warrant such work. The smoothing kernel density estimation technique espoused by Silverman is not up to the task even of finding the global $\Phi(r)$, let alone probing resonant structure.

---

[41] Taking velocity moments of Eq. 130 (i.e. multiplying by $v_r \frac{\partial f}{\partial v_r}$ and integrating by parts) we recover the Jeans equation

$$\frac{d\Phi}{dr} = \frac{1}{r}(\langle 2v_r^2\rangle_r - \langle v_t^2\rangle_r) + \frac{d(\ln\rho)}{dr}\langle v_r^2\rangle_r + \frac{\partial\langle v_r^2\rangle_r}{\partial r}. \tag{131}$$

The results suggest the following: attempting to recover the distribution function from discrete velocity phase space data without making use of dynamical inputs or acknowledging that an integrable system will be restricted in its orbits to a 3-manifold in phase space is a non-optimal method. The extremely low density of points, the concern about ignoring inherent sub-symmetries, and computational considerations are likely to restrict one to splitting the distribution into real and velocity subspaces, and the resulting averaging over narrow configuration space volumes results in a trade-off between highly inaccurate and Poisson noise dominated velocity space distributions.

The adaptive kernel method was found here to be less than satisfying. For a mechanical approach to apply always, there must be a clear and reproducible program of execution. First, the degree of freedom present in the choice of sensitivity parameter is a problem. In the real-space case, non-parametric fitting yielded a good fit. Sensitivity parameters as distinct from the intuitive $\frac{1}{d}$ as $\frac{1}{7}$ were found to apply, however, and for parameters of $\frac{1}{d}$ and $\frac{1}{2}$, as recommended by Silverman, the fit was so poor that the graph displayed significant noise and had no discernible global shape.

Second, there are serious questions about how well-defined a quantity the optimal smoothing parameter is. The optimal fixed kernel smoothing length h_{fix} was found to be completely unrelated to the optimal smoothing length determined by the adaptive method. Cross-validation, as applied here, can fail to find a unique smoothing length, or $\hat{h}_{\text{opt}}$ can be far from the actual minimizer of the mean integrated square error. The degree to which the estimator of the optimum smoothing length $\hat{h}_{\text{opt}}$ approximates the value of the smoothing length h_{opt} that minimizes the integrated square error is measured for example by their ratio and is known as the *asymptotic optimality.* For a symmetric and twice differentiable probability density f, Marron (1986) finds that the dependence of optimality on sample size takes the form

$$n^{\frac{d}{2(d+4)}} \frac{\hat{h}_{\mathrm{opt}} - h_{\mathrm{opt}}}{h_{\mathrm{opt}}} \to N(0, \sigma^2) \tag{132}$$

for a Gaussian random variable N with mean 0 and some dispersion σ independent of n, while the error in the density estimate behaves as

$$\|\hat{f} - f\| \propto n^{-\frac{2}{d+4}} \tag{133}$$

for $\|\cdot\|$ an error criterion such as root mean square error. This leads to a number of interesting conclusions. First, as the dimensionality d increases, it becomes harder to estimate f, with the error declining more slowly as a function of sample size for larger d. Second, and seemingly paradoxically, $\hat{h}_{\mathrm{opt}}$ will tend to be closer to h_{opt} as d increases. In other words, selection of the smoothing parameter becomes easier and more precise, but the quality of the estimate $\hat{f}$ diminishes with dimension. This phenomenon is explained by a reduction in sensitivity to noise around the minimum as d increases. For $d = 3$, the error in the density estimate is $O(n^{-\frac{2}{7}})$; for $d = 6$, we have $O(n^{-\frac{1}{5}})$. A factor 10^4 increase in sample size to incorporate a full treatment of, for instance, the GAIA data-set improves $\|\cdot\|$ by only a factor six. The convergence of $\hat{h}_{\mathrm{opt}}$ improves with dimension, but is poor throughout the range of dimension treated here. Additionally, it can also be shown that for $\mathcal{F}$ a suitable class of functions and c small in some sense, the probability P_f of finding a better convergence for any function of the data $\hat{h}_{\text{opt}}$ is bounded by

$$\sup_{f\in\mathcal{F}} P_f\left[\left|\frac{\hat{h}_{\text{opt}}-h_{\text{opt}}}{h_{\text{opt}}}\right| > cn^{-\frac{d}{2(d+4)}}\right] \to 1 \qquad (134)$$

Marron concludes that 1) least squares cross-validation can be expected to give a smoothing parameter far from optimum, and 2) there will **never** be a substantially better method of smoothing parameter selection.

Many authors in the analysis community instead propose that a subjective approach may be best, i.e. that one might do better to select various h and determine relative quality by eye. Merritt and Tremblay suggest, for example, looking for estimates that closely follow the mean trend of estimates with much smaller smoothing parameters, thus smoothing out noise, without running an exceptional risk of over-smoothing. This amounts to attempting to compromise between bias and variance: in general there exists no universal strategy to achieve this. In any case, such an approach has no hope of success for a multivariate density distribution.

Even the whitening process is not without controversy. Setting the covariance matrix $S = I$ removes all information higher than zeroth order in the data. While the transformation is inverted at the end, there are unresolved questions about whether the kernels are smoothing out important features in the covariance structure of the distribution.

It seems then that the velocity phase space density cannot be accurately modelled by techniques of density estimation only,[42] and one must first reduce the dimensionality of the

---

[42] For the MPLE method, the non-linear nature of its statistics means that there are few rigorous results regarding its behaviour. Merritt and Tremblay (1996) claim that for analysis of the Plummer model, the kernel method is more suitable than the MPLE.

problem. Here too one might consider a model-independent approach. The key techniques offered by numerical analysis are ***principal components analysis*** and ***projection pursuit*** (see Scott, 1992). Principal components analysis is based on the idea that the eigenvalues D_{ii} of the covariance matrix S' yield information on the relative importance of the degrees of freedom. This assumes that a mere rotation of the rectangular coordinate system chosen in the representation of the data points is sufficient to reveal unimportant degrees of freedom. The map to action-angle space is non-linear, and one can reject this approach here.

In projection pursuit analysis, one attempts to find a projection matrix $P : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ for $d' < d$, by maximizing a measure of the non-smoothness of the density. Various such quantities have been proposed (see references in Scott, 1992). It was found, for example, that most random projections of multi-modal distributions down to $\mathbb{R}^1$ led to nearly normal distributions, so that one approach is to look for a projection resulting in the 'least' normal distribution. Since obtaining the most suitable P is a non-linear optimization problem, it is difficult. In many cases (see the discussion in Scott) it fails, and it is safe to say there exists no commonly held view on its efficacy in case of general d and d'. In particular, only $d' = 1, 2$ have been studied in any detail.[43] Projection pursuit is again just linear algebra and hence not suited to mapping into tori. Recently, Delaunay tessellation techniques have shown more promise (Arad et al., 2004), suffering only from long computation times. The so-called FIESTAS technique (Ascasibar and Binney, 2004), involving assigning volume to points according to a binary tree, is considerably faster, but the scope for modelling resonant substructures appears to be limited.

---

[43] ***Neural networks,*** specializing in pattern recognition, can also be adapted to dimensionality reduction. The multi-layer perceptron is essentially a special case of projection pursuit (see Bishop, C.M. 1995, Neural Networks for Pattern Recognition, OUP, Oxford.)

The fact that it was possible to question seriously the validity of non-parametric velocity phase space modelling in this best-case scenario provides a compelling counter-example to a purely numerical approach to resonant dynamics. Scott (1992) sums it up neatly: '*Density estimation is quite difficult if the data cloud...falls (nearly) onto a lower-dimensional manifold.*' It would seem that the optimal way to model global velocity phase space data must involve dynamical considerations. We can reject the null hypothesis.

6 Conclusion

In this chapter, we touch on higher-dimensional systems and reveal the resonant structure in the Laskar frequency map of a triaxial system of some astrophysical relevance. We conclude with a summary.

6.1 Higher-dimensional systems

Many of the results derived in earlier chapters apply to n-dimensional systems. Generating functions for perturbation theory remain the same, as do expressions for the perturbed frequencies. The method of averaging an available set of actions over true orbits remains accurate to $O(|s|^2)$, i.e. to the order of the square of the deviation of the generating function from the identity map. Indeed this method becomes more attractive as the usefulness of the Poincaré section diminishes.

Since isoenergetic surfaces in phase space are $(2n-1)$-dimensional, the codimension of a KAM torus is $n-1$, such that $n=2$ constitutes a special case, for which the KAM tori divide Hamiltonian level surfaces into two separated regions. Since the KAM tori are invariant, orbits with initial phase space conditions not in a KAM torus cannot cross KAM tori and are therefore constrained to orbit in the toric annuli between the nested tori. Large scale instability is therefore impossible for $n=2$. For all other n, these restrictions do not apply, and as n increases, more and more mechanisms for instability appear. Now families of resonant tori are interconnected to form a resonance web called the *Arnold web.*

All resonances in $n=2$ have multiplicity one, but recall from Chapter 1 that in general, a system can have resonances of any multiplicity $1 \leq m \leq n-1$. For $m > 1$, we can no longer map the non-trivial dynamics of a resonance into a single resonant action-angle plane as we did in Chapter 4, and the simple pendulum picture becomes more complicated. In the frequency map of Chapter 3, there were no intersections between resonant lines, because all resonances satisfied $a_1\frac{\omega_r}{\omega_\varphi} + a_2\frac{\omega_\vartheta}{\omega_\varphi} = 0$. For $n = 3$, on the frequency ratio plane $(\frac{\omega_1}{\omega_3}, \frac{\omega_2}{\omega_3})$, a *single* $(m = 1)$ *resonance* condition $a \cdot \omega = 0$ is a line; a *double* $(m = 2)$ *resonance* condition $a^1 \cdot \omega = a^2 \cdot \omega = 0$ is an intersection of two or more single resonance lines (axial resonance orbits require separate consideration, since their frequency ratio is zero or singular). A point at which many resonances intersect is known as a *junction*. As before, a point sufficiently far from resonances on the frequency map represents an $(n-2)$-parameter family of invariant KAM tori (since each point on the map fixes two frequency ratios). If it were possible to follow the time-evolution of isoenergetic orbits with a starting grid of initial conditions, one would then expect to see a regular pattern of points corresponding to these families of KAM tori away from resonance, and chaotic behaviour identifying the location of the resonances. One would further expect to be able to identify single resonances with hyperbolic or other stability by their characteristic signatures on the frequency map. Many examples of such frequency maps exist in the literature: see, for instance, (Binney and Tremaine, forthcoming).

Frequency maps are normally only useful for $n \leq 3$, but we can test their limits and simultaneously study an astrophysically relevant example by introducing a time-varying potential. Of course most triaxial potentials with relevance to the Galaxy are quite weakly triaxial and we cannot hope to find more than small perturbations from axisymmetry.

Consider a disk potential with a Spitzer disk (Binney and Tremaine, 1987) density,

$$\rho_{\text{Spitzer}}(z) = \rho(z_0)\,\text{sech}^2 \frac{z}{2z_0}, \tag{135}$$

with z_0 the scale-height of the disk. Integrating Poisson's equation and imposing the boundary condition $\Phi(0) = 0$, we find that

$$\Phi_{\text{Spitzer}}(z) = 16\pi z_0^2 \rho(z_0) \ln\left(\cosh \frac{z}{2z_0}\right). \tag{136}$$

We make the potential consistent with the surface density of the power-law disk

$$\Sigma(R) = \Sigma(R_0)\left(\frac{R_0}{R}\right)^{1+\beta}, \ \beta \neq 0, \tag{137}$$

by replacing $\rho(z_0) \to \rho(z_0)\Sigma(R)$ and normalizing $\int_{-\infty}^{\infty} \rho_{\text{Spitzer}}(z)\,dz = 1$. The surface density $\Sigma(R)$ is consistent with the potential

$$\Phi_1(R, z=0) = -\frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^{\beta} \tag{138}$$

(see Evans and Read, 1998), where

$$v_\beta^2 = 2\pi\Sigma(R_0)R_0 \frac{\Gamma[\frac{1}{2}(1-\beta)]\Gamma[\frac{1}{2}(2+\beta)]}{\Gamma[\frac{1}{2}(1+\beta)]\Gamma[\frac{1}{2}(2-\beta)]}. \tag{139}$$

We use the bar potential Φ_2 according to Dehnen (2000), so our total potential is of the form

$$\begin{aligned}
\Phi(R,\varphi,z,t) &= \Phi_1(R,z)+\Phi_2(R,\varphi,t), && (140)\\
\Phi_1 &= -\frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^{\beta}+4\pi z_0\Sigma(R_0)\left(\frac{R_0}{R}\right)^{1+\beta}\ln\left(\cosh\frac{z}{2z_0}\right), && (141)\\
\Phi_2 &= A_b\cos(2(\varphi-\Omega_b t))\left[\begin{array}{ll} -\left(\frac{R_b}{R}\right)^3, & R\geq R_b\\ \left(\frac{R}{R_b}\right)^3-2, & R\leq R_b \end{array}\right. . && (142)
\end{aligned}$$

In the above, A_b is the bar amplitude, Ω_b the rotational frequency of the bar, R_b the bar radius, R_0 the radius of the Sun's orbit. Dehnen has the bar amplitude gradually increase from zero, according to

$$\begin{aligned}
A_b &= A_f\left(\frac{3}{16}\xi^5-\frac{5}{8}\xi^3+\frac{15}{16}\xi+\frac{1}{2}\right),\ \xi=2\frac{t}{t_1}-1,\ t\leq t_1\\
A_b &= A_f,\ t\geq t_1. && (143)
\end{aligned}$$

Once we switch on the bar and make the potential triaxial, the Poincaré map is no longer useful, and we must turn to frequency mapping. In Figure 31 we see a representation of the weakness of the bar.[44] The plotted points in the frequency ratio plane come from a 5x5x5 grid of initial conditions in (L_z,θ_v,φ). The grid spans the manifold defined by $\{H=E, R=R_0, \vartheta=\frac{\pi}{2}\}$.[45] Arrows in the figure show the variation in frequency ratio

---

[44] The parameters used in this simulation followed the suggestions in Dehnen (2000) and were chosen as follows: $R_0 = 8$ kpc, $R_b = 3.3$ kpc, using $\frac{\Omega_b}{\Omega_0} = 1.875$ and $\frac{R_b}{R_{CR}} = 0.8$ for R_{CR} the corotation radius and Ω_0 the circular frequency at R_0. We also chose $A_f = 1.2\cdot 10^{-5}$ kpc^2/Myr^2, $t_1 = 4\frac{2\pi}{\Omega_b}$, $\Sigma(R_0) = 5\cdot 10^7$ $M_\odot$/kpc^2, $z_0 = 0.3$ kpc, and $\beta = 0.05$.

[45] Frequency maps are graphed at a chosen $H = E$. Since all orbits go through the plane $z = 0$, we choose the initial $\vartheta = \frac{\pi}{2}$. We use $R = R_0$ and a grid in polar velocity launch angle $\vartheta_v \in (0,\frac{\pi}{2})$ and $L_z \in [L_{min}, R\sqrt{2(E-\Phi)}]$. A lower cutoff $L_z \geq L_{min}$ is necessary to filter out highly eccentric orbits that do not observe a third integral.

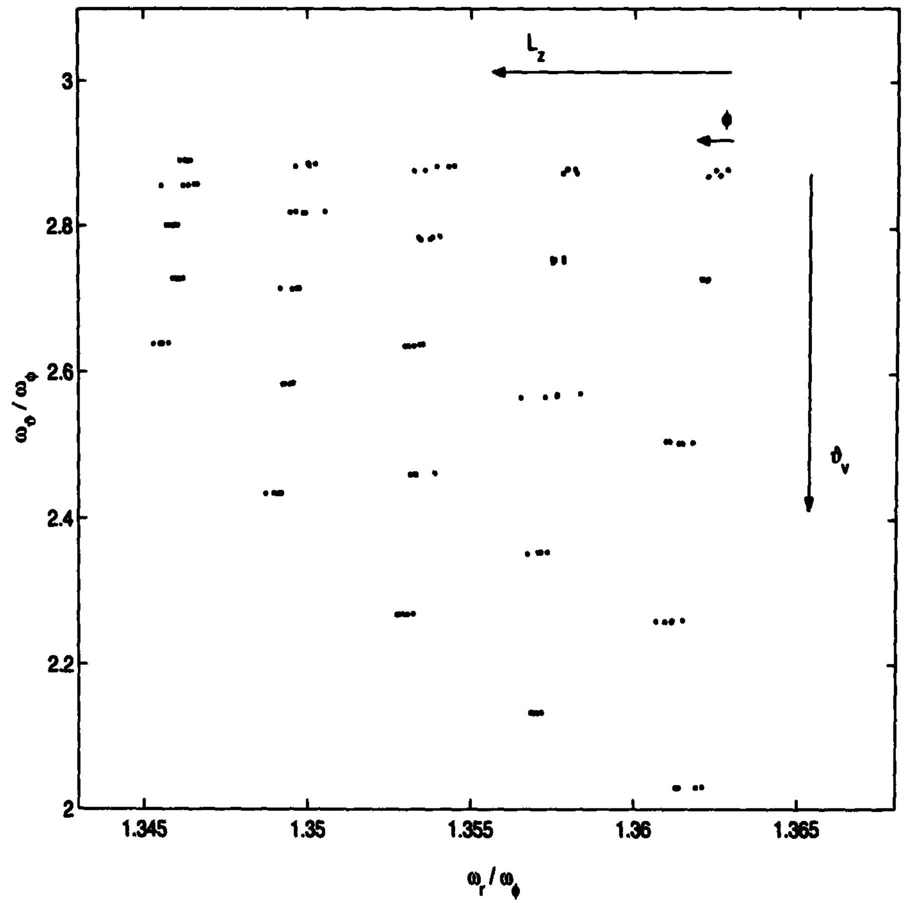

Figure 31: A 5x5x5 grid in initial conditions in $(L_z, \vartheta_v, \varphi)$, for $H = E$, $\vartheta = \frac{\pi}{2}$ and a suitable choice of R, produces this grid of points in the frequency ratio plane. Arrows indicate the effect of varying the initial conditions. The deviation from axisymmetry, as measured by the splitting caused by varying φ, is seen to be weak.

caused by moving along the grid. We see immediately that the potential is predominantly axisymmetric. The splitting caused by varying $\varphi \in [0, 2\pi]$ is small compared to the variation in frequency ratio as one changes L_z. Resonances involving all three frequencies are likely to be very narrow. The resulting Laskar frequency map[46] is shown in Fig. 32. As predicted, the only prominent resonances are axisymmetric. In particular, the $a = (3, -2, 0)$ and $a = (2, -1, 0)$ resonances, for $a \cdot (\omega_r, \omega_\vartheta, \omega_\varphi) = 0$, stand out. Each of these leaves a trail of captured orbits, but also affects the dynamics of a neighbourhood in the frequency ratio plane by depopulating it. If we look more closely, or we increase the final bar amplitude

---

[46] Since the bar rotates, the ω_φ found from the $\tilde{x}$ spectrum should in fact be modified to $\Omega_b - \omega_\varphi$. The axisymmetric resonances are, however, not as clear in the resulting diagram (only the $(2, -1, 0)$, $(3, -2, 0)$ and $(5, -3, 0)$ can be discerned with confidence), and the triaxial resonant structure is not significantly enhanced.

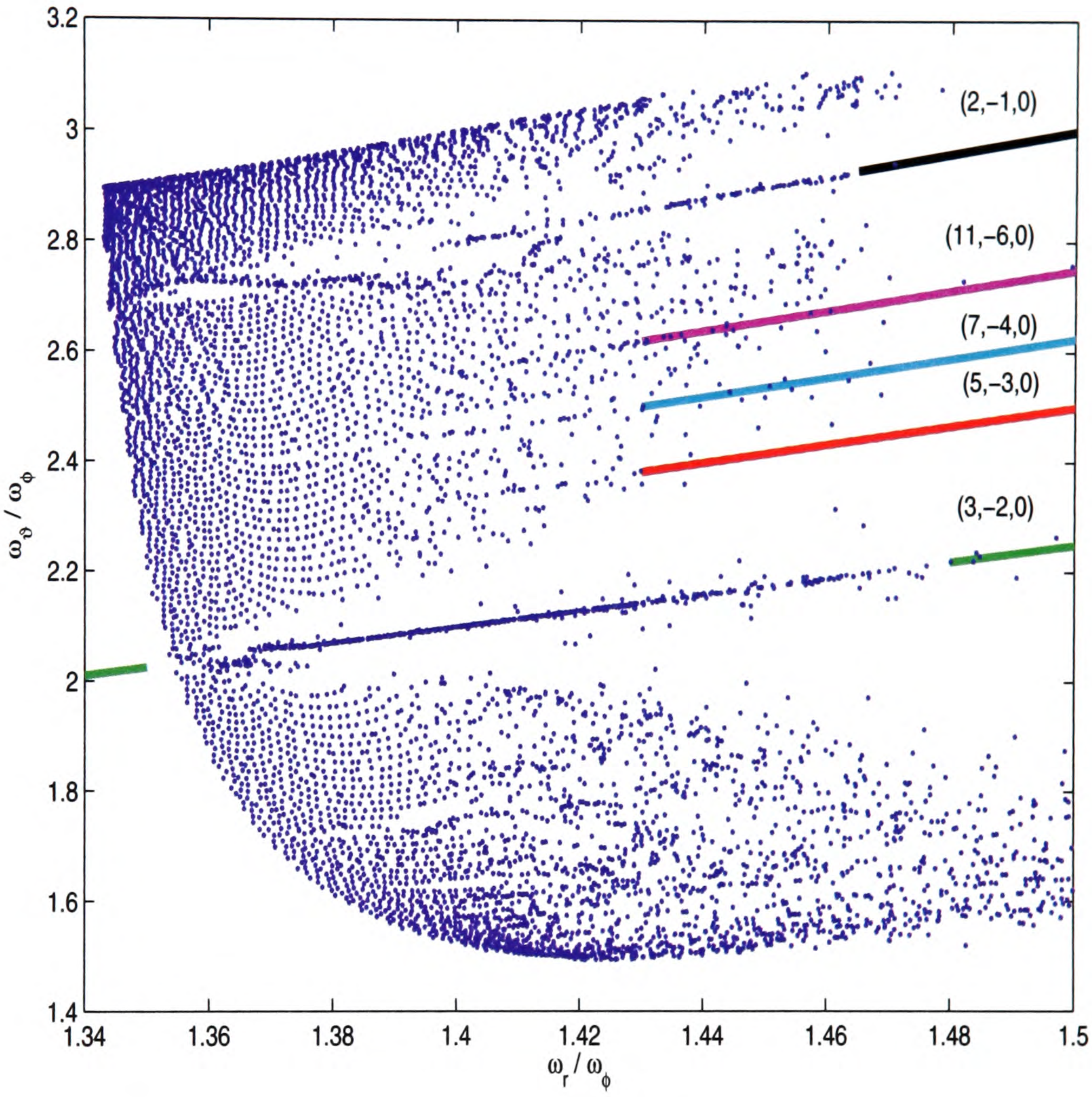

Figure 32: The Laskar frequency map for the weakly-barred potential of Eq. 142. A number of resonances can be seen, including several at surprisingly high order. These are all due to the axisymmetric component of the potential. The map is produced for $E = -0.06$.

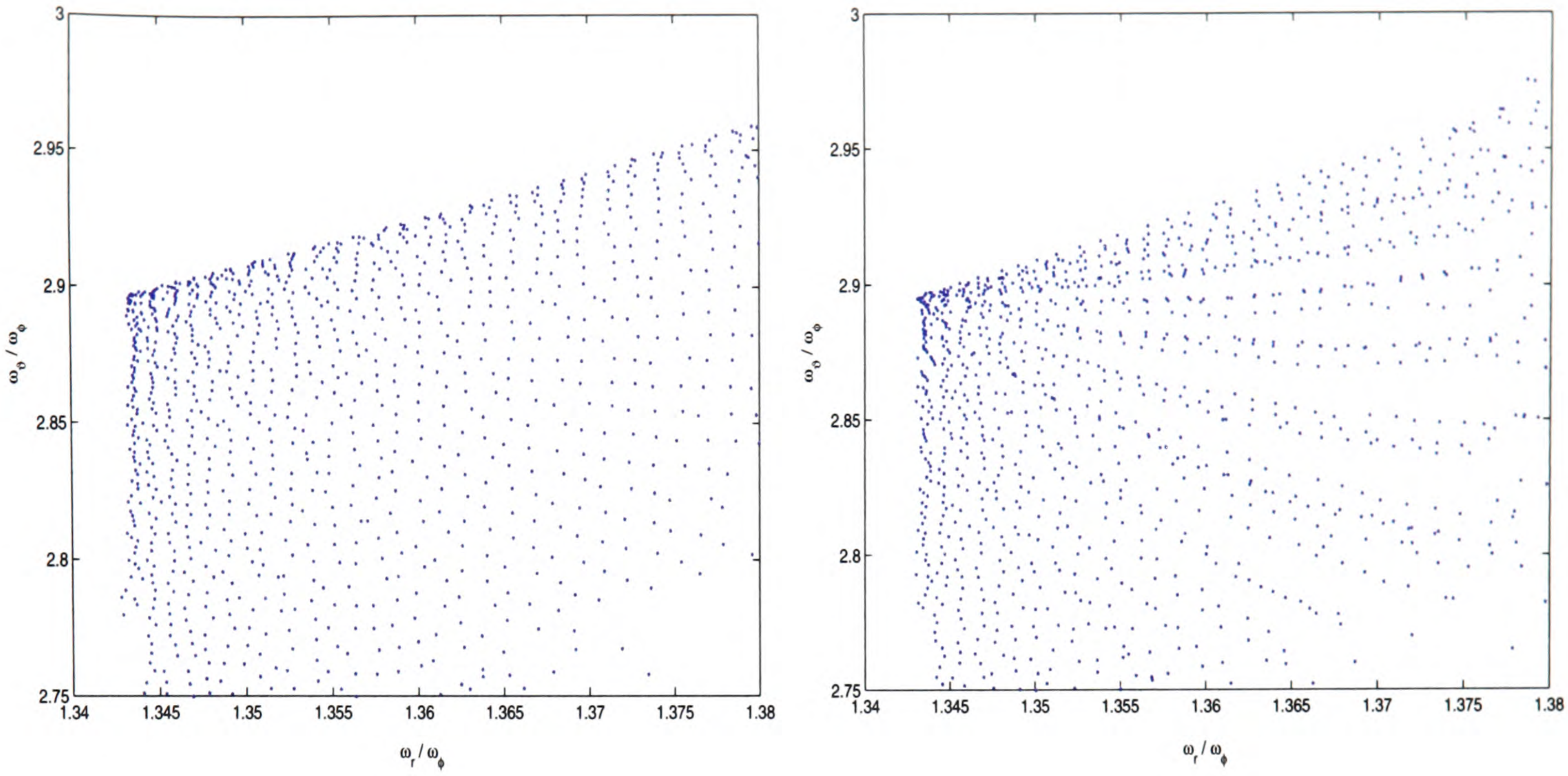

Figure 33: Detail of the frequency map for a maximum bar amplitude $A'_f = 0$ (left) and $A'_f = 10A_f$ (right), for A_f the maximum amplitude of the Dehnen bar. Dynamical features associated with the time-variation of the potential appear in the second diagram.

A_f, we can see effects due to the time-variation of the potential. In Fig. 33 we examine a detail of the frequency map for the cases $A'_f = 0$ and $A'_f = 10A_f$, with A_f the final bar amplitude as recommended by Dehnen. When the bar is more pronounced, we do in fact see an effect on the dynamics.

6.2 Afterword

Numerical techniques form the backbone of the study of resonant dynamics. We have shown that traditional methods of density estimation do not resolve resonant substructures in the velocity phase space, and we are compelled to treat the problem in a momentum-configuration space. The Hamiltonian formulation provides us with action-angle space, the natural symplectic coordinates for the analysis of regular orbits. For quasi-integrable

Hamiltonians, one can derive generating functions for first-order perturbation theory that regularize near-integrable orbits and provide expressions for their perturbed frequencies. We have shown, however, that the success of perturbation theory is strongly dependent on the quality of the available action-angle coordinates. Kaasalainen (1994) generated action-angle coordinates by means of the *torus machine*. In this case, first-order perturbation theory performed extremely well. We find, however, that when only a set of spherical actions is available, perturbation theory underperforms near orbital pericentre. Perturbation theory is also unsatisfactory when the ratio of two actions is $O(\epsilon^2)$ or smaller, for ϵ the deviation of the Hamiltonian from integrability.

Where numerical orbit integrations are available, Poincaré maps to suitable phase planes offer an effective overview of the resonant dynamics of 2-dimensional systems. Traditionally, actions are computed by finding areas enclosed by inferred curves through the consequents in each Poincaré section, but we have found that time-averaging an available set of actions (in practice that means the analytic actions of a spherical potential) can be as effective as numerical line integration. Averaged actions can therefore be used to counter-check actions obtained by perturbation theory. In systems with three degrees of freedom, averaged actions are particularly useful, as the Poincaré section does not provide closed paths in tori around which to integrate.

The averaging technique is also intrinsically useful. Although neither perturbation theory nor the averaging technique produces the correct invariant actions for librational tori, they can tell us both about the location of the exact resonance condition and show us the size of the resonance. The averaging technique was found to be useful in theoretically predicting capture probabilities for orbits trapped in resonance by a time-varying potential.

It also helped identify that the scatter around a linear regression in the action plane would be a useful diagnostic test of a trial Hamiltonian. Perturbation theory also provides us with estimates of the circulation frequencies of non-resonant orbits. We saw in Chapter 3 that frequency ratios can to some extent be used to identify resonantly trapped orbits, even though the method does not yield the librational frequencies of trapped orbits, but in Chapter 4 we demonstrated that perturbed frequencies are essential in constraining trial Hamiltonians.

Adaptive dynamics is the key to this. Spectral dynamics and the Laskar frequency map reveal the important resonances of a model Hamiltonian. The model also provides the action space in which to observe the clustering of resonantly trapped stars. Once a prominent resonance has been identified in the data, it should be straightforward to constrain the potential. We have developed two complementary tests for this purpose: perturbed frequency and scatter around the exact resonance condition. We can also use the averaging technique as an audit of the accuracy of perturbed actions.

The research presented in this thesis is only a very small step towards the application of resonant dynamics to the Galaxy, and the task ahead is both daunting and urgent, since the GAIA data-set will be upon us within a decade. While we have revealed the failure of traditional methods of density estimation and the vulnerabilities of perturbation theory, adaptive dynamics seems to provide a real hope, not only of exploring resonant structures, but also of the wider task, central to Galactic dynamics, of constraining the Galactic potential. It is to be hoped that the demands on the Guardians of Socrates' Republic to chart the velocities and positions of the stars and infer their orbits will then finally be met.

A Null hypothesis: sampling the velocity-space distribution

The velocity space density distribution is given by

$$\rho_v(v_r, v_t; r) = \begin{cases} \frac{6}{(2\pi)^3\rho(r)}(2\Phi(r) - v_r^2 - (1+r^2)v_t^2)^{\frac{3}{2}}, & \left\{0 \leq v_t \leq \left(\frac{2\Phi(r)-v_r^2}{1+r^2}\right)^{\frac{1}{2}},\right. \\ & \left. -\sqrt{2\Phi(r)} \leq v_r \leq \sqrt{2\Phi(r)}\right\} \\ 0 & \text{otherwise.} \end{cases} \tag{144}$$

The conditions on v_r and v_t define an ellipsoidal volume in $(v_\vartheta, v_\varphi, v_r)$ space, $2\Phi(r) = v_r^2 + (1+r^2)v_\vartheta^2 + (1+r^2)v_\varphi^2$ and a semi-elliptical region in the (v_r, v_t) plane that forms the domain of ρ_v. The elliptical boundary of integration is determined by r only, and geometrically similar semi-ellipses enclosed entirely by the boundary of the domain are given by $v_r^2 + (1+r^2)v_t^2 = 2\Phi(r)p^2$ with $p \in (0,1]$. So the fraction of points contained within the bounding ellipse identified by p (i.e. integrating also over the velocity cylindrical polar angle ψ) is

$$\Xi(p) = \frac{6}{(2\pi)^2\rho(r)} \int_{-p[2\Phi(r)]^{\frac{1}{2}}}^{p[2\Phi(r)]^{\frac{1}{2}}} \int_0^{\left(\frac{2\Phi(r)p^2-v_r^2}{1+r^2}\right)^{\frac{1}{2}}} (2\Phi(r) - v_r^2 - (1+r^2)v_t^2)^{\frac{3}{2}} v_t \, dv_t \, dv_r, \quad p \in (0,1]. \tag{145}$$

The integral can be evaluated analytically and yields, the r dependence dropping out,

$$\Xi(p) = \frac{32}{5\pi}\left(-\frac{5}{6}(1-p^2)^2 + \frac{5}{24}(1-p^2) + \frac{5}{16}\right)p(1-p^2)^{\frac{1}{2}} + \frac{2}{\pi}\arctan\frac{p}{(1-p^2)^{\frac{1}{2}}}. \tag{146}$$

Clearly there is not the dimmest hope of inverting this analytically to get $p(\Xi)$. To proceed numerically, we choose the random variable $\Xi \in [0,1]$ and a range of test values of p. The most accurate of these test values is then used as the initial estimate in MATLAB's inbuilt iterative zero-finder. The method is found to work extremely well, and calculates values of p that are accurate to seven significant figures or better, far beyond the necessary precision.

We now have a way of populating the ellipses. We then choose the azimuthal coordinates by noting that in the transformed coordinate system $(v'_\vartheta, v'_\varphi, v_r) = ((v_\vartheta, v_\varphi)(1+r^2)^{\frac{1}{2}}, v_r)$ the boundary of integration is a sphere and the density distribution becomes spherically symmetric. Then for $v'^2 = v_r^2 + v'^2_\vartheta + v'^2_\varphi = 2\Phi(r)p^2$,

$$\begin{aligned} v_r &= v' \cos\eta && (147) \\ v_\vartheta &= v' \sin\eta \cos\psi \, (1+r^2)^{-1/2} && (148) \\ v_\varphi &= v' \sin\eta \sin\psi \, (1+r^2)^{-1/2}, && (149) \end{aligned}$$

where $\cos\eta$ is randomly distributed in $[-1,1]$. Finally, ψ is chosen randomly in $[0,2\pi]$.

B Aliasing

Perturbation theory calls for an expansion of $h(J, \theta)$ in a discrete Fourier series

$$h(J,\theta) = \sum_{a \in \mathbb{Z}^2 - \{0\}} h_a(J) e^{i a \cdot \theta}, \tag{150}$$

the solution to which is

$$h_a(J) = \frac{1}{4\pi^2} \int_{[0,2\pi]^2} h(J,\theta) e^{-i a \cdot \theta} \, d^2\theta. \tag{151}$$

For computational purposes, we use the estimator

$$\begin{aligned}
\hat{h}_a(J) &= \frac{1}{N^2} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} h\left(J, \theta = \frac{2\pi}{N}(k_1, k_2)\right) e^{-\frac{2\pi i}{N} a \cdot k} && (152) \\
&= \frac{1}{N^2} \sum_{b \in \mathbb{Z}^2 - \{0\}} \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} h_b(J) e^{\frac{2\pi i}{N}(b-a)\cdot k} \\
&= \frac{1}{N^2} \sum_{b \in \mathbb{Z}^2 - \{0\}} h_b(J) \prod_{j=1}^{2} \frac{e^{2\pi i (b_j - a_j)} - 1}{e^{\frac{2\pi i}{N}(b_j - a_j)} - 1} \\
&= \sum_{b \in \mathcal{B} - \{0\}} h_b(J) && (153)
\end{aligned}$$

for $\mathcal{B}$: $b = (c_1 N + a_1, c_2 N + a_2)$, $c \in \mathbb{Z}^2$. To see this last step, consider what happens as $b_j - a_j \to c_j N$, treating $\alpha \equiv \frac{b_j - a_j}{N}$ as a continuous variable:

$$\lim_{\alpha \to c_j} \frac{e^{2\pi i N \alpha} - 1}{e^{2\pi i \alpha} - 1} = \lim_{\alpha \to c_j} \frac{2\pi i N e^{2\pi i N \alpha}}{2\pi i e^{2\pi i \alpha}} = N. \tag{154}$$

Then we have

$$\begin{aligned}\hat{h}_a(J) = h_a(J) \quad &+ \quad h_{a_1+N,a_2}(J) + h_{a_1-N,a_2}(J) + \cdots \\ &+ \quad h_{a_1,a_2+N}(J) + h_{a_1,a_2-N}(J) + \cdots \end{aligned} \qquad (155)$$

i.e. the error in the estimator arises from Fourier coefficients shifted down or up by multiples of the size of the grid. This phenomenon is called *aliasing.* The remedy is simple: doubling N eliminates precisely half the error terms.

C Reference: actions and angles of the spherical isochrone potential

The actions of the spherical isochrone potential $\Phi(r) = -\frac{1}{b+\sqrt{b^2+r^2}}$ with angular momentum L, special component of angular momentum L_z and orbital energy E are

$$
\begin{aligned}
J_r &= \frac{1}{\sqrt{-2E}} - \frac{L}{2} - \sqrt{\frac{L^2}{4} + b} && (156)\\
J_\vartheta &= L - |L_z| && (157)\\
J_\varphi &= |L_z|, && (158)
\end{aligned}
$$

where Eq. 158 is valid for all axisymmetric potentials, and Eq. 157 is valid for the subset of spherical potentials. With frequencies $\omega = \frac{\partial H}{\partial J}$ and the following definitions,

$$
\begin{aligned}
a &\equiv \frac{1}{-2H} - b && (159)\\
e^2 &\equiv 1 - \frac{(J_\vartheta + J_\varphi)^2}{a}\left(1 + \frac{b}{a}\right) && (160)\\
s &\equiv 1 + \sqrt{1 + \frac{r^2}{b^2}} && (161)\\
\cos\eta &\equiv \frac{1}{e}\left[1 - \frac{b}{a}(s-2)\right] && (162)\\
\cos i &\equiv \frac{J_\varphi}{J_\vartheta + J_\varphi}, && (163)
\end{aligned}
$$

the angles can be written as

$$
\begin{aligned}
\theta_r &= \eta - \frac{ea}{a+b}\sin\eta && (164)\\
\theta_\vartheta &= \arcsin\left(\frac{\cos\vartheta}{\sin i}\right) + \frac{\omega_\vartheta}{\omega_r}\theta_r - \arctan\left(\sqrt{\frac{1+e}{1-e}}\tan\left(\frac{\eta}{2}\right)\right)\\
&\quad - \frac{1}{\sqrt{1+4b/J_\vartheta^2}}\arctan\left(\sqrt{\frac{1+e+2b/a}{1-e+2b/a}}\tan\left(\frac{\eta}{2}\right)\right) && (165)\\
\theta_\varphi &= \theta_\vartheta - \varphi + \arcsin\left[\cot\vartheta\frac{J_\varphi}{\sqrt{(J_\varphi+J_\vartheta)^2 - J_\varphi^2}}\right]. && (166)
\end{aligned}
$$

Pericentre r_- and apocentre r_+ for the zero-velocity orbit of the isochrone potential are given by

$$r_\pm = \sqrt{\frac{B \pm \sqrt{B^2 - 4E^2C}}{2E^2}} \tag{167}$$

with $B = 2Eb + EL^2 + 1$, $C = L^2(b + \frac{1}{4}L^2)$. The Hamiltonian for the spherical isochrone potential is

$$H(J) = -\frac{1}{2}\left(J_r + \frac{L}{2} + \sqrt{\frac{L^2}{4} + b}\right)^{-2}, \tag{168}$$

and $\omega_r = \frac{\partial H}{\partial J_r} = (-2H)^{\frac{3}{2}}$. The other two frequencies are

$$\omega_\vartheta = \omega_\varphi = \frac{1}{2}\left(1 + \frac{L}{\sqrt{L^2 + 4b}}\right)\omega_r. \tag{169}$$

For further details please see Saha (1989), McGill and Binney (1990), and Kaasalainen (1994).

Bibliography

Abramson, I. S. 1982, Ann. Statist., 10, 1217

Arad, I., Dekel, A., Klypin, A. 2004, MNRAS, 353, 15

Arnold, V. I. 1963, Siberian Math. J., 4, 471.

Arnold, V. I. 1989, Mathematical Methods of Classical Mechanics,
Springer-Verlag, New York

Ascasibar, Y., Binney, J. 2004, MNRAS, 356, 872

Binney, J. 1978, Comments Ap., 8, 27

Binney, J. Modelling the Galaxy for GAIA, in
The Three-Dimensional Universe with GAIA, eds. Turon, C., Perryman, M., ESA, 2005.

Binney, J., Gerhard, O. E., Hut, P. 1985, MNRAS, 215, 59

Binney, J., Spergel, D. 1982, ApJ, 252, 308

Binney, J., Spergel, D. 1984, MNRAS, 206, 159

Binney, J., Tremaine, S. 1987, Galactic Dynamics, Princeton University Press, Princeton

Binney, J., Tremaine, S. Forthcoming edition of Galactic Dynamics

Chirikov, B. V. 1979, Physics Reports, 52, 265

Contopoulos, G. 1960, Zeitschrift für Astrophysik, 49, 273

Contopoulos, G. 1963, ApJ, 138, 1297

Contopoulos, G. 1983, A & A, 117, 89

Dehnen, W. 1998, AJ, 115, 2384

Dehnen, W. 2000, AJ, 119, 800

Dejonghe, H. 1986, Physics Reports, 133, 217

Epanechnikov, V. I. 1969, Theory Probab. Appl. 14, 153

Evans, N. W., Read, J. C. A. 1998, MNRAS, 300, 83

Famaey, B., Jorissen, A., Luri, X., Mayor, M., Udry, S., Dejonghe, H., Turon, C. 2005, A & A, 430, 165

Fukunaga, K. 1972, Introduction to Statistical Pattern Recognition, Academic Press, New York

Hénon M., Heiles C. 1964, Astrophys. J., 69, 73

Henrard, J., Lemaitre, A. 1983, Celest. Mech., 30, 197

Kaasalainen, M. K. J. 1994, D.Phil. Thesis, University of Oxford

Kolmogorov, A. N. 1954, Dokl. Akad. Nauk SSSR, 98, 527;
English translation in *Lecture notes in physics* 93, Springer-Verlag, 1979, Berlin.

Laskar, J. 1993, CeMDA, 56, 191

Lichtenberg, A.J. and Lieberman, M.A. 1991, *Regular and Chaotic Dynamics.* Springer-Verlag, New York

McGill, C., Binney, J. 1990, MNRAS, 244, 634

Marron, J. S. 1986, Function Estimates, Contemporary Mathematics, 59, 169

Merritt, D., Tremblay, B. 1994, AJ, 108, 514

Morbidelli, Alessandro. 2002, Modern Celestial Mechanics:
Aspects of Solar System Dynamics, Taylor & Francis, London

Mosteller, F., Tukey, J.W. 1977, Data Analysis and Regression, Addison-Wesley, Cambridge, MA

Nekhoroshev, N. N. 1977, Russ. Math. Surveys, 32, 1

Saha, P. 1989, D.Phil. Thesis, University of Oxford

Scott, D. 1992, Multivariate Density Estimation, Wiley, New York

Schwarzschild, M. 1979, ApJ, 232, 236

Silverman, B.W. 1986, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London de Simone, R., Wu, X., Tremaine, S. 2004, MNRAS, 350, 627

Sridhar, S., Touma, J. 1996, MNRAS, 279, 1263

Thompson, J.R., Tapis, R.A. 1990, Nonparametric Function Estimation, Modeling, and Simulation, Philadelphia: SIAM

Tukey, P.A., Tukey, J.W. 1981, Interpreting Multivariate Data (V. Barnett, ed.), Wiley, New York

de Zeeuw, T., Peletier, R., Franx, M. 1986, MNRAS, 221, 1001