CONDITIONAL ENTROPY
A Tool to Explore the Phase Space

P. CINCOTTA*

Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque, 1900 La Plata, Argentina

C. SIMÓ†

Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain

Abstract. In this paper we show that the Conditional Entropy of nearby orbits may be a useful tool to explore the phase space associated to a given Hamiltonian. The arc length parameter along the orbits, instead of the time, is used as a random variable to compute the entropy. In the first part of this work we summarize the main analytical results to support this tool while, in the second part, we present numerical evidence that this technique is able to localize (stable) periodic and quasiperiodic orbits, 'aperiodic' orbits (chaotic motion) and unstable periodic orbits (the 'source' of chaotic motion). Besides, we show that this technique provides a measure of chaos which is similar to that given by the Lyapunov Characteristic Number. It is important to remark that this method is very simple to compute and does not require long time integrations, just realistic physical times.

Key words: Chaos – Lyapunov Characteristic Number – Entropy

1 Introduction

It is well known that the Lyapunov Characteristic Number (LCN) provides a measure of stochasticity. By definition, the LCN is an asymptotic value for $t \to \infty$. For practical applications in Galactic Dynamics, reasonably asymptotic results are obtained for motion times of the order of $10^5 - 10^6 T_D$, where T_D is a characteristic period of motion of the system, while the Hubble time or the lifetime of a galaxy is of the order of $10^2 T_D$, $10^3 T_D$ as much. Therefore, the motion time needed to compute the LCN is very unrealistic for these applications (see, for example, [8] and [10]). Also, such long computing times turn this tool unsuitable to carry out a detailed study of the orbital structure of a given potential. Besides, since the LCN is a time–average magnitude for large t, (the mean rate of exponential divergence of nearby orbits), relevant information about the dynamics is missed. One alternative is the spectra of stretching numbers or Lyapunov numbers for finite times (see [2] and references there; [11]). Another powerful technique is the Frequency Map Analysis ([6], [7]; [12] for applications to realistic galactic models). In this work we propose the 'Conditional Entropy' of nearby orbits as an effective tool to investigate the

*Present address (until December 1998): Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain. E-mail: pablo@zeus.maia.ub.es–pmc@fcaglp.fcaglp.unlp.edu.ar

†E-mail: carles@maia.ub.es

phase space structure of a given Hamiltonian in short motion times, i.e., $t \lesssim 10^3 T_D$. The latter concept, that was introduced first in [9], differs from the standard entropy in the fact that the arc length parameter of the orbit under consideration is used as a random variable. In that work the authors presented numerical evidence that this technique is efficient to separate ordered and stochastic regions of the phase space in relatively short times.

This work is divided in two parts. The fist one deals with some theoretical considerations about the Conditional Entropy that lack in [9]. However a more detailed theory behind this method will be addressed in a separate paper. In the second part, we illustrate the use of this tool using a simple 2D system. As it was shown in [9], this technique seems to work very well also in 3D systems.

2 Set Up, Definitions and Notation

Let us consider the Hamiltonian:

$$H(\mathbf{p},\mathbf{q}) = \frac{\mathbf{p}^2}{2} + \phi(\mathbf{q}), \qquad \mathbf{p},\mathbf{q} \in \mathbb{R}^N, \tag{1}$$

where ϕ is a smooth potential. For the sake of simplicity we write:

$$\mathbf{x} = (\mathbf{p},\mathbf{q}) \in \mathbb{R}^{2N}, \qquad \mathbf{v} = (-\partial H/\partial \mathbf{q},\ \partial H/\partial \mathbf{p}) = (-\boldsymbol{\nabla}\phi, \mathbf{p}) \in \mathbb{R}^{2N}. \tag{2}$$

Then the equations of motion are:

$$\dot{\mathbf{x}} = \mathbf{v}(\mathbf{x}). \tag{3}$$

Let M_h be the manifold (energy surface): $M_h = \{\mathbf{x} : H(\mathbf{p},\mathbf{q}) = h\}$, so $\phi(\mathbf{q}) \le h$. Throughout the present work we consider that the motion is bounded in phase space, that is, M_h is compact. Let T_D be a dynamical time–scale associated with the Hamiltonian (1) and $\gamma \subset M_h$ be an arc of an orbit of $\mathbf{v}$:

$$\gamma = \{\mathbf{x}(t;\mathbf{x}_0) : \mathbf{x}_0 \in M_h,\ 0 \le t \le T < T_*\}, \tag{4}$$

where we suppose that $T_* \sim 10^3 T_D$. Since T is finite, the length of γ:

$$L_\gamma = \int_\gamma ds, \tag{5}$$

is a finite quantity. In (5), s is the arc length parameter associated with the orbit γ[1]. Let $\psi(\mathbf{x})$ be any scalar function of the phase–space coordinates. We define the average value of ψ along the orbit γ as:

$$\langle\psi\rangle_\gamma \equiv \frac{1}{L_\gamma}\int_\gamma \psi(\boldsymbol{\chi}(s))ds, \tag{6}$$

where $\boldsymbol{\chi}(s(t)) = \mathbf{x}(t)$. Besides:

$$\int_\gamma \psi(\boldsymbol{\chi}(s))ds = \int_0^T \psi(\mathbf{x}(t))|\dot{\mathbf{x}}(t)|dt = T\langle\psi|\mathbf{v}|\rangle_T, \tag{7}$$

[1] Clearly, s depends on the metric; in this case $ds^2 = dx_1^2 + ... + dx_{2N}^2$.

where $\langle . \rangle_T$ denotes the time–average but over a finite time, $|.|$ is the usual Euclidean norm and we have made use of (3). From (5), (6) and (7) we readily see that:

$$L_\gamma = T\langle |\mathbf{v}| \rangle_T, \qquad \langle \psi \rangle_\gamma = \frac{\langle \psi |\mathbf{v}| \rangle_T}{\langle |\mathbf{v}| \rangle_T}. \tag{8}$$

Set $\psi = \ln |\mathbf{v}(\mathbf{x})|$. We call *entropy of the orbit* γ, $S(\gamma)$, the magnitude given by:

$$S(\gamma) = -\langle \ln |\mathbf{v}| \rangle_\gamma + \ln L_\gamma. \tag{9}$$

The term $\ln L_\gamma$ is introduced just for convenience. Indeed, from (8) and (9) it is not difficult to show that $S(\gamma)$ can be written in the form:

$$S(\gamma) = -\int_0^T \rho_\gamma(t) \ln \rho_\gamma(t) dt, \qquad \rho_\gamma(t) = \frac{|\mathbf{v}(\mathbf{x}(t))|}{L_\gamma} \geq 0. \tag{10}$$

From (10) we see that $\int_0^T \rho_\gamma dt = 1$. Formally, the definition of $S(\gamma)$ resembles the familiar continuous entropy:

$$S_c(X) = -\int_a^b \rho(x) \ln \rho(x) dx, \tag{11}$$

for the set $X = \{x \in \mathbb{R},\ a < x < b\}$ and the distribution density $\rho(x)$. The latter may be considered as the continuous limit of:

$$S_d(X) = -\sum_{i=1}^{m} \mu(x_i) \ln \mu(x_i) \geq 0, \tag{12}$$

for the set $X = \{x_i, i = 1, ..., m,\ a < x_i < b\}$, where μ is the probability associated to x_i, i.e., the normalised measure of the elements of a given partition in X. Therefore, we can follow [3]. Let us summarize the main results.

Consider a measurable space M provided with a normalised measure (probability) μ. Let $A = \{a_i, i = 1, ..., n\} \subset M$. The entropy of the set A, $S(A)$, is defined as in (12). Consider another set $B = \{b_j, j = 1, ..., m\} \subset M$. The conditional entropy of A relative to B is defined by:

$$S(A|B) = \sum_{j=1}^{m} \mu(b_j) S(A|b_j) = -\sum_{j=1}^{m} \mu(b_j) \sum_{i=1}^{n} \mu(a_i|b_j) \ln \mu(a_i|b_j),$$

where $\mu(a_i|b_j)$ is the conditional probability of a_i for a given b_j. Since $\mu(a_i|b_j) = \mu(a_i, b_j)/\mu(b_j)$, with $\mu(a_i, b_j)$ the joint probability of a_i and b_j, then one readily finds:

$$S(A|B) = -\sum_{i,j} \mu(a_i, b_j) \ln \mu(a_i, b_j) + \sum_j \mu(b_j) \ln \mu(b_j) = S(A, B) - S(B),$$

where $S(A, B)$ is the joint entropy of both sets. Introducing $I(A, B)$ as the symmetric conditional entropy: $I(A, B) \equiv \frac{1}{2}[S(A|B) + S(B|A)]$ we get:

$$I(A, B) = S(A, B) - \frac{1}{2}[S(A) + S(B)]. \tag{13}$$

It is not difficult to verify that if, for example, $n = m$ and $\mu(a_i) = \mu(b_i)$ for all i, then $S(A,B) = S(A)$ and I reduces to zero. On the other hand if A and B are independent sets (in a probabilistic sense), then $S(A,B) = S(A) + S(B)$ and $I = [S(A) + S(B)]/2 > 0$. In general, it can be proved that $I(A,B) \geq 0$ for any sets A and B, while $I(A,B) = 0$ if and only if $\mu(a_i) = \mu(b_i)$ (see [1] for a formal presentation of the subject and proofs). Since $\mu(a_i|b_j)$ is a measure of the correlation between the sets A and B we can state the former mathematical results in 'statistical language' as follows: *if the sets A and B are strongly correlated, then $I(A,B) \approx 0$, while $I(A,B) \approx [S(A) + S(B)]/2$ whenever A and B are uncorrelated.*

When we deal with a continuous case, like in (11), we extend the former definitions but the probability, $\mu(a_i)$, has to be replaced by a continuous probability density $\rho(a)$, such that $d\mu(a) = \rho(a)da$ and $S(A) = -\int_A d\mu(a) \ln \rho(a)^2$. Proceed in a similar way for $S(B)$ and $S(A,B)$. Hence, since I is a relative entropy, we conserve the property $I \geq 0$ that holds in the discrete case (for further details see [5]).

Let us return to the Hamiltonian (1), the orbit γ defined in (4) and identify $\gamma \to A$. Now consider another orbit γ' similar to γ, on the same or a nearby, $M_{h'}$, level of energy but for a slightly different initial condition: $\mathbf{x}'_0 = \mathbf{x}_0 + \boldsymbol{\delta}_0$. Let us denote as $\dot{\mathbf{x}}' = \mathbf{v}(\mathbf{x}') \equiv \mathbf{v}'$ the associated vector field. In the same way as we did with γ, we can construct $\rho_{\gamma'}$ and $S(\gamma')$. Then we identify $\gamma' \to B$. To define the joint entropy of both, γ and γ', we proceed as follows. Let $\Gamma = \gamma \times \gamma' \subset M_h \times M_{h'}$ be the curve defined through the vector field $\mathbf{V} = (\mathbf{v}, \mathbf{v}')$ such that $|\mathbf{V}|^2 = |\mathbf{v}|^2 + |\mathbf{v}'|^2$. Then $\rho_\Gamma = |\mathbf{V}(\mathbf{x}(t), \mathbf{x}'(t))|/L_\Gamma$. Thus we can define $I(\gamma, \gamma')$ in the same way as in (13):

$$I(\gamma, \gamma') = S(\Gamma) - \frac{1}{2}[S(\gamma) + S(\gamma')]. \tag{14}$$

Suppose that the Hamiltonian (1) is a near–integrable one, where the associated phase space is shared between regular and chaotic motion. Let us denote with $\Sigma_r, \Sigma_c \subset M_h$ those regions of the energy surface where the motion is regular and chaotic respectively. Among others, the main feature that distinguish both components is the following: if $\mathbf{x}_0 \in \Sigma_r$, then γ and γ' will remain close one another, diverging in mean at a linear rate (in fact, any subexponential rate will do). On the other hand, if $\mathbf{x}_0 \in \Sigma_c$, then γ and γ' will diverge at an exponential rate. Using the results given above –in the statistical sense– we can state: *Let $\mathbf{x}_0 \in \Sigma_r$ and T finite. Then γ and γ' will be strongly correlated and $I(\gamma, \gamma') \approx 0$; on the other hand, if $\mathbf{x}_0 \in \Sigma_c$ then, for $T > T_c$, γ and γ' will be uncorrelated and $I(\gamma, \gamma') > 0$, where T_c is some critical time.* In other words, for very short times ($T < T_c \ll T_*$) we expect the same behaviour of I for both, regular and chaotic orbits ($I \approx 0$). This is just a consequence of the fact that, for this time interval, the divergence of initially nearby orbits is very small and thus γ and γ' are very similar. As the time (T) increases, the correlation between the two orbits evolves. But for $T > T_c$, this evolution will be rather different if $\mathbf{x}_0$ belongs to a regular or to a chaotic region of the phase space. It is clear that T_c can be interpretated as the time needed to pass close to an hyperbolic object with non coincident separatrices. If the initial condition lies in a regular region, we expect that γ and γ' will lose its correlation very slowly and then $I \ll 1$ for all T. On the other hand, if $\mathbf{x}_0$ lies

[2] Actually it should be added a constant, say '$S_0 = \infty$'. So in the continuous limit, we lose the positivity of the entropy.

in a chaotic region, γ and γ' will lose its correlation very fast (due to the exponential divergence), and I will take then a larger value. The 'distributions' $\rho(t)$ contain all the information about the flow and should reveal this rather different behaviour.

In the next paragraphs we shall summarize the main analytical results for I. By definition, for any t, we have $\dot{\mathbf{x}}'(t) = \dot{\mathbf{x}}(t) + \dot{\boldsymbol{\delta}}(t)$. Denote with $x(t) \equiv |\dot{\mathbf{x}}(t)|$, $\delta(t) \equiv |\boldsymbol{\delta}(t)|$, $d(t) \equiv |\dot{\boldsymbol{\delta}}(t)|$ and $\xi(t) = d(t)/x(t)$, which is assumed to be small. Then, up to second order in ξ, an elementary but tedious computation shows that:

$$\rho_{\gamma'} \approx \rho_\gamma \left(1 + a_1 + a_2 - k_1 a_1\right), \qquad \rho_\Gamma \approx \rho_\gamma \left(1 + a_1/2 + a_2/4 - k_1 a_1/4 + a_3\right), \tag{15}$$

where:

$$a_j(t) = \frac{\Delta_j(t)}{x(t)} - \frac{\langle \Delta_j \rangle_T}{\langle x \rangle_T} \quad j = 1, 2, 3, \qquad k_1 = \frac{\langle \Delta_1 \rangle_T}{\langle x \rangle_T}, \tag{16}$$

with:

$$\Delta_1 \equiv \Delta x = \frac{\dot{\boldsymbol{\delta}}.\dot{\mathbf{x}}}{x} = d \cos\zeta, \qquad \Delta_2 = \frac{1}{2}\frac{d^2}{x}\sin^2\zeta, \qquad \Delta_3 = \frac{d^2}{8x}. \tag{17}$$

In (17), $\zeta(t)$ is the angle between $\dot{\mathbf{x}}(t)$ and $\dot{\boldsymbol{\delta}}(t)$. From (15), it is straightforward to verify that $\rho_{\gamma'}$ and ρ_Γ are normalised. As we see from (15)–(17), $\rho_{\gamma'}$ and ρ_Γ differ from ρ_γ –at $\mathcal{O}(\xi)$– by the instantaneous fluctuation of $\Delta x(t)$, the first variation of $|\mathbf{v}(\mathbf{x}(t))|$. After some algebra we obtain for I:

$$I(\gamma, \gamma') \approx \frac{\langle x a_1^2 \rangle_T}{8\langle x \rangle_T}. \tag{18}$$

For the moment being we shall assume that $\langle x \rangle_T$ depends *in a mild way* on T (see, however, the remark at the end of this section). Then we write $\langle x \rangle_T \approx X$, independent of T, and we can approximate I by:

$$I(\gamma, \gamma') \approx \frac{1}{8X^2}\left(\langle \Delta_1^2 \rangle_T - \langle \Delta_1 \rangle_T^2\right) \tag{19}$$

As we note the dependence of I on T is given by the variance of Δ_1 between $t = 0$ and $t = T$. From (17), Δ_1 is expected to have a very small average since it depends linearly on $\cos\zeta$. However this is not the case for Δ_1^2. Therefore, to obtain I we must be able to compute the time–mean square value of Δ_1. From the second equality in the first of (17), the second in (2) and recalling that:

$$\dot{\delta}^k = \frac{\partial v^k}{\partial x^l}\delta^l, \qquad v^k \equiv (\mathbf{v})^k, \tag{20}$$

(the sum over repeated indexes is understood) then it is straightforward to get:

$$\Delta_1 \approx \frac{1}{X}(\phi_k \phi_{kl} - \phi_l)\delta q^l = \frac{1}{X}\Phi_l \delta q^l, \tag{21}$$

where the subscripts in ϕ denote derivatives respect to the position, $\Phi_l = \phi_k\phi_{kl} - \phi_l$ and $\boldsymbol{\delta}\mathbf{q}$ is the displacement from $\mathbf{q}$ at the time t. The latter is given by the variational equations (20). However, unless we are near to an equilibrium point, these equations cannot be solved analytically. Therefore, to make some progress in the theoretical approach, we have to introduce further assumptions. Let us consider the simplest case: regular motion. In this case we know that $\delta q^l(t) \approx \delta_0 \lambda_l t$ is the expected law for the time evolution of δ, where $\delta_0 \ll 1$ is the initial displacement and λ_l is the time–rate of linear divergence in a neighbourhood of the initial point $(\mathbf{x}_0)$ in the l–direction. Let $\lambda(\mathbf{x}_0) = \max\{\lambda_l\}$, so $\lambda_l = \lambda b_l$, where $b_l \leq 1$. Then, introducing this in (21) we have:

$$\Delta_1(t) \approx \frac{\delta_0 \lambda t}{X} \Psi\left(\mathbf{q}(t)\right), \tag{22}$$

where $\Psi = \Phi_l b_l$ depends only on the position through the derivatives of the potential. If $\boldsymbol{\omega}$ is the frequency vector associated with the invariant torus where the motion proceeds, then we can expand $\Psi\left(\mathbf{q}(t)\right)$ in Fourier series and, provided that $(\boldsymbol{\omega}.\mathbf{k})T$ is large enough for all integer vectors $\mathbf{k} \neq 0$, then from (19) and (22) we finally obtain for $I(\gamma, \gamma')$:

$$I(\gamma, \gamma') \approx \frac{\delta_0^2 \lambda^2 A}{24 X^4} T^2, \tag{23}$$

where A is a positive constant (the average value of Ψ^2) which is almost independent of T for $T \gg T_D$. As we note, I depends on the initial condition through $\lambda(\mathbf{x}_0)$, so we can write $I(\gamma, \gamma') \equiv I(T; \mathbf{x}_0)$. Therefore, for two different initial conditions in Σ_r (and for the same values of δ_0 and T), we have: $I_1/I_0 = \lambda_1^2/\lambda_0^2$, where the smaller λ appears for $\mathbf{x}_0$ in a neighbourhood of a stable periodic orbit (think about using a Normal Form around this orbit).

At this step, one has to keep in mind that (23) is valid for small $\xi(t)$. Using $d(t) \approx R\delta(t)$ $(R \sim 1)$[3], then we can state that (23) is true for $\lambda T \ll X/\delta_0$. Therefore, $I \sim \mathcal{O}\left(\delta_0^2 T^2\right) \ll 1$. We see then that for regular motion, I behaves in the expected way: I takes a very small value due to the strong correlation between γ and γ'.

Since I depends on T, it will be interesting to know its time–rate in general. Using the analytic formula for the derivative of the entropy (see (28)) and up to second order in ξ we get:

$$\frac{dI}{dT} \approx \frac{1}{8T}\left(a_1^2(T) - 8I(T)\right)\frac{x(T)}{\langle x \rangle_T}.$$

We introduce then the quantity $J(\gamma, \gamma') \equiv d\log I / d\log T$, which is given by:

$$J(\gamma, \gamma') = \frac{T}{I}\frac{dI}{dT} \approx \left(\frac{a_1^2(T)}{8I(T)} - 1\right)\frac{x(T)}{\langle x \rangle_T} \approx \frac{\Delta_1^2(T)}{\langle \Delta_1^2 \rangle_T} - 1, \tag{24}$$

where the last follows for $x(t) \approx \langle x \rangle_T \approx X$ and $\langle \Delta_1 \rangle_T = 0$. This relation shows that J seems to be a more interesting magnitude than I. For example, in the case of regular motion, we get:

$$J(\gamma, \gamma') \approx 2, \qquad T \gg T_D, \tag{25}$$

independent of δ_0, T, λ and any other scale parameter. Besides, it is not difficult to see that J only depends on δ: that is, if δ grows with some power of t, say t^r, then $J \approx 2r$.

[3] This two quantities cannot differ too much since $\delta^2 = |\boldsymbol{\delta}\mathbf{q}|^2 + |\boldsymbol{\delta}\mathbf{p}|^2$ while $d^2 = |\boldsymbol{\delta}\mathbf{p}|^2 + |\boldsymbol{\delta}\boldsymbol{\nabla}\phi|^2$. In fact $d \approx R\delta$, with R of bounded magnitude seems to be a reasonable approximation.

For the chaotic case, the analytical results are not so accurate. We shall say that if δ grows exponentially: $\delta(t) \approx \delta_0 e^{\sigma t}$, then the dominant terms in I and J are:

$$I(\gamma, \gamma') \sim \frac{e^{2\sigma T}}{\sigma T}, \qquad J(\gamma, \gamma') \sim 2\sigma T - J_0, \qquad T \gg T_D. \tag{26}$$

where $J_0 \sim 1$ is a constant independent of σ, the mean–rate of exponential divergence.

As we have already mentioned, the approximate analytical results rest on the assumption $\langle x \rangle_T \approx X$. By definition

$$x^2 = 2[h - \phi(\mathbf{q})] + |\boldsymbol{\nabla}\phi(\mathbf{q})|^2.$$

Since the motion is bounded in the phase space (confined to the energy surface) and if ϕ is smooth enough, then $\langle x \rangle_T$ is a bounded quantity of the order of the size of the system. In the simple case of an N–dimensional harmonic oscillator, $x(t) = \langle x \rangle_T = \sqrt{2h}$.

3 Numerical Examples

For the numerical study of this technique, we considered the Hénon–Heiles model [4], for the energy level $h = 0.118$ ($T_D \lesssim 10$). Within this energy surface, we have restricted the analysis to the region shown in Fig. 1(a). Fixing $p_2 = 0$, for $0.3 \leq q_2 \lesssim 0.508$ we have a large regular component, where $q_2 \approx 0.305$ corresponds to the stable 1–periodic orbit. For $0.508 \lesssim q_2 \lesssim 0.57$ we observe a 5–periodic island with its stochastic layer around the separatrix and bounded by a KAM curve. This KAM curve, which can be seen in this figure, separates different stochastic regions and disappears for $h = 0.119$. For $q_2 \gtrsim 0.57$ we see a highly stochastic region. Certainly, within the whole domain, a large number of small islands is present. In Fig. 1(b) we show the computed LCN for 1000 initial conditions along the q_2 axis, $0.55 \leq q_{2_0} \leq 0.60$ and for $T = 2.5 \times 10^5$. This was done in order to have a rough estimation of the measure of chaos in the stochastic components. Even though the Hénon–Heiles is a very simple model, the structure of the region considered in Fig. 1(a) is always present in almost all near–integrable 2D Hamiltonian systems.

For the explicit computation of I and J we take advantage that from (10), $S(\gamma) \equiv S(T; \mathbf{x}_0)$ can be written in the form:

$$S(T; \mathbf{x}_0) = -\frac{1}{L(T; \mathbf{x}_0)} \int_0^T |\mathbf{v}(t; \mathbf{x}_0)| \ln |\mathbf{v}(t; \mathbf{x}_0)| dt + \ln L(T; \mathbf{x}_0), \tag{27}$$

where $L(T; \mathbf{x}_0)$ is the time–integral of the vector field (3). $S(\gamma')$ and $S(\Gamma)$ satisfy similar formulae. Then we use (14) to obtain $I(T; \mathbf{x}_0)$. Besides, from (27) one readily finds:

$$\frac{dS}{dT} \equiv \dot{S}(T; \mathbf{x}_0) = \rho(T; \mathbf{x}_0)\,(1 - S(T; \mathbf{x}_0)) - \rho(T; \mathbf{x}_0) \ln \rho(T; \mathbf{x}_0). \tag{28}$$

Then $\dot{I} = \dot{S}(\Gamma) - \frac{1}{2}[\dot{S}(\gamma) + \dot{S}(\gamma')]$ and finally, $J = T\dot{I}/I$.

The numerical calculations can be done in three different ways. The alternatives are: (A1) To take two nearby initial conditions, $\mathbf{x}_0$ and $\mathbf{x}_0'$, and to integrate the equations of motion (3) for both initial conditions to get explicitly γ and γ'. Then with the help of (27)

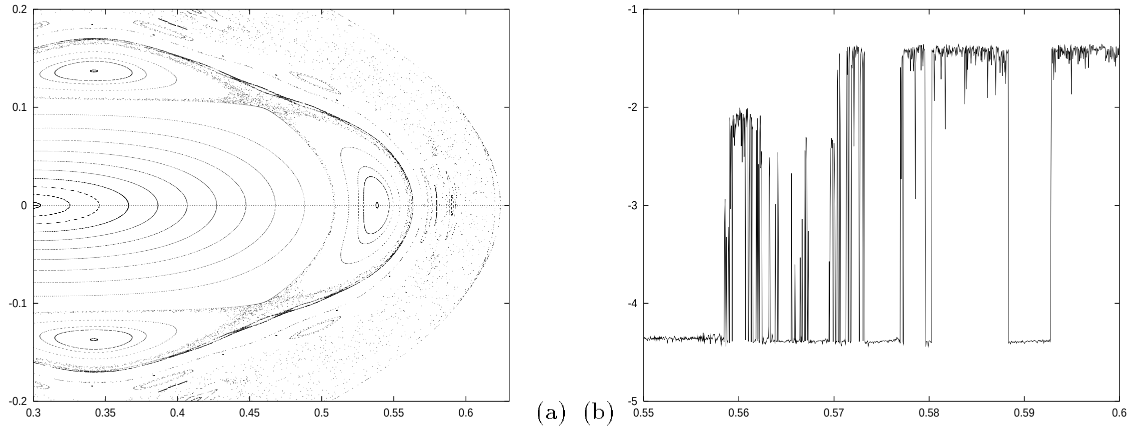


Figure 1: (a) Surface of section (q_2, p_2), $p_1 > 0$ for the Hénon–Heiles model with $h = 0.118$ and several initial conditions: $q_{1_0} = 0$ and $(q_2, p_2)_0$ taken on the line $p_{2_0} = 0$, with $0.3 \leq q_{2_0} \leq 0.62$. (b) log (LCN), after $T = 2.5 \times 10^5$, for $p_{2_0} = 0$ and 1000 values of q_{2_0} in the range $0.55 \leq q_{2_0} \leq 0.60$.

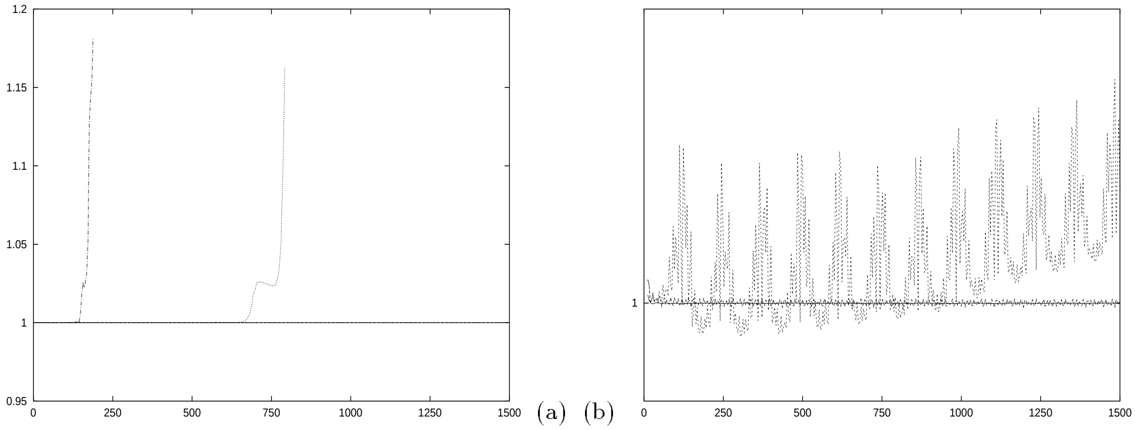


Figure 2: (a) $\rho_{\gamma'}/\rho_\gamma$ vs. T, for $p_{2_0} = 0$, $q_{2_0} = 0.305,\ 0.5,\ 0.5085$ (these three curves look as the same, very close to 1), 0.509 (dotted line) and 0.6 (dotted–dashed line). (b) Zoom in the neighbourhood of $\rho_{\gamma'}/\rho_\gamma = 1$, for the first three curves in (a), window: [0.999995,1.000015] (see text).

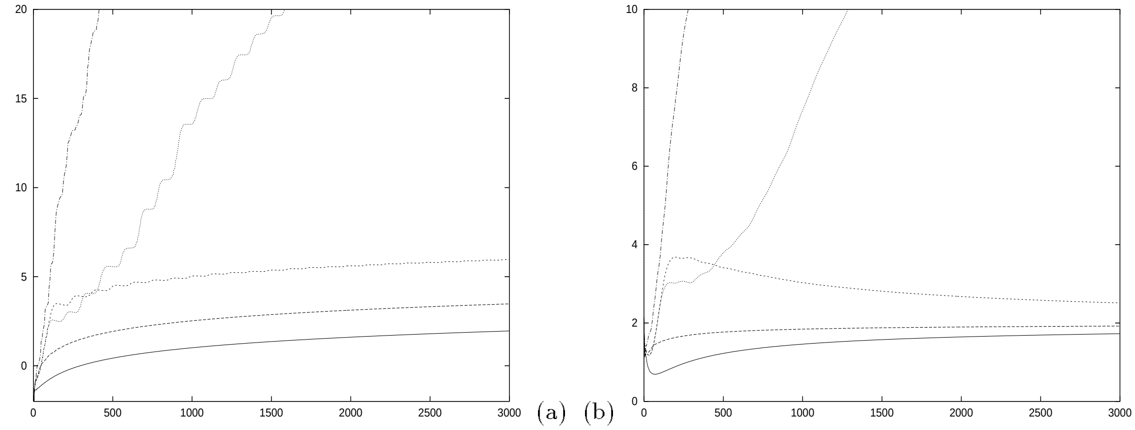


Figure 3: (a) $\log I$ vs. T; (b) J vs. T for the same orbits than in Fig. 2. Starting from the solid curve: $q_{2_0} = 0.305,\ 0.5,\ 0.5085,\ 0,509$ and 0.6.

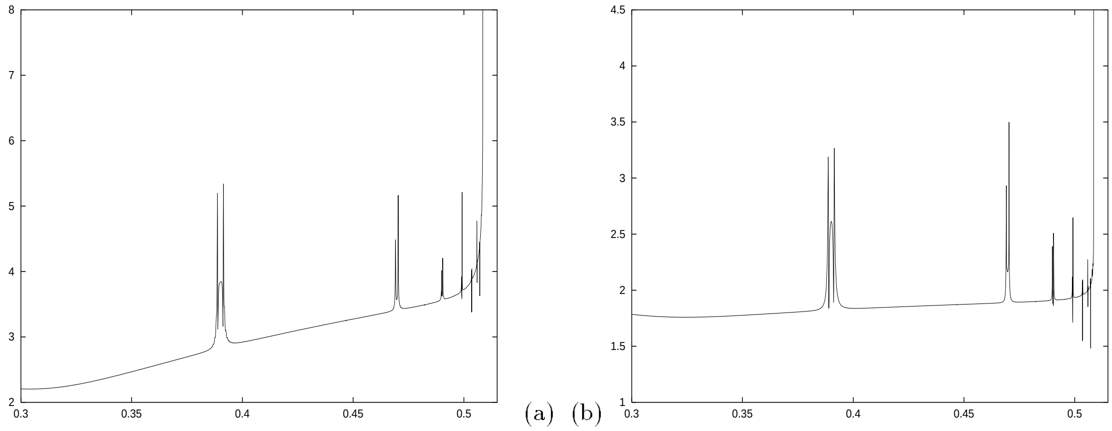
and (28), we obtain I and J. (A2) To approximate $\mathbf{x}'(t) = \mathbf{x}(t) + \boldsymbol{\delta}(t)$ where $\mathbf{x}'_0 = \mathbf{x}_0 + \boldsymbol{\delta}_0$ and $\boldsymbol{\delta}(t)$ is the solution of the variational equations (20). Having computed γ and γ' we proceed then like in (A1). This approximation is justified by the fact that, even though I depends on $\boldsymbol{\delta}^2$, $I \propto \langle a_1^2 \rangle_T$, this term is the square of a term that comes from the first variational equations, while those cuadratic terms that come from the second variationals cancel. (A3) To use the second order formulae given by (18) and by the first in (24). A1 was used in [9] and has the restriction that $\delta(t) = |\mathbf{x}'(t) - \mathbf{x}(t)|$ reaches a saturation value given by the size of the system for the corresponding energy level. A2 is used here for the calculation of $\rho(T)$, with $\delta_0^k = \delta_0/2$, $\delta_0 = 10^{-6}$, $k = 1, ...4$ (note that γ and γ' are in different but very close energy levels). Even though with the A2 procedure $\delta(t)$ does not have a saturation level as in A1, this second alternative has some numerical limitations. Indeed, in both limits, $\delta_0 \to 0$ and/or δ large ($\sim X \sim$ the size of the system), the algorithm becomes numerically unstable. Therefore A2 has the restriction that I, and consequently J, depend on the selected value of δ_0. However, within the range $10^{-7} \leq \delta_0 \leq 10^{-4}$, J is almost independet of δ_0 while for I, the latter behaves as an scale factor (see (23)). A3, that is used here for all the computations of I and J, does not require the explicit calculation of γ' and therefore, is numerically stable. In this case δ_0 is always a scale factor that we set (arbitrarily) equal to one. Therefore, $I \sim \mathcal{O}(T^2) \gg 1$ for the regular component while $I \sim \mathcal{O}(\exp(\sigma T)) \ggg 1$ for the stochastic component. In any case, since J depends explicitly on $\rho(T)$ we averaged the computed J, just performing the sum of the successive values of $J(T_i)$ for $T_i \leq T$ and dividing the sum by T. This was done twice in order to lower as much as possible the effect of fast oscillations that do not affect the aperiodical changes. This is a simple smoothing procedure.

Fig. 2(a) shows the time evolution of $\rho_{\gamma'}/\rho_\gamma$ (averaged once and using A2) for initial conditions in different zones along the q_2 axis: $q_{2_0} = 0.305$ (near to the stable 1–periodic orbit), $q_{2_0} = 0.5$ (quasiperiodic orbit associated with the latter periodic orbit), $q_{2_0} = 0.5085$ (similar to the former but closer to the separatrix), $q_{2_0} = 0.509$ (inside the stochastic layer), $q_{2_0} = 0.6$ (in the highly stochastic sea). We did not plot ρ_Γ because its behaviour is almost the same than $\rho_{\gamma'}$. We see that for regular motion, $\rho_{\gamma'}$ and ρ_γ seem to be identical, while for chaotic motion the behaviour is similar up to certain value of $T = T_c$. For $T > T_c$, $\rho_{\gamma'}$ diverges form ρ_γ in an exponential way (see (15)–(17)). Note that, for the orbit in the highly stochastic sea, this divergence occurs after a few periods of motion while for that in the stochastic layer, this divergence occurs after several periods. Besides, the time–rate of exponential divergence seems to be different in both cases. A magnification around $\rho_{\gamma'}/\rho\gamma = 1$ (Fig. 2 (b) – note the size of the window in the vertical axis) reveals that, for the quasiperiodic orbit ($q_{2_0} = 0.5085$) which is very close to the separatrix, $\rho_{\gamma'}/\rho_\gamma$ exhibits small periodic pulses and a drift. However this effect is not observed for the other regular orbits. For $q_{2_0} = 0.5$, $\rho_{\gamma'}/\rho_\gamma$ oscillates very fast about 1 with a very small amplitude while for $q_{2_0} = 0.305$, $\rho_{\gamma'}/\rho_\gamma = 1$ for the resolution of the figure. This periodical behaviour for the outermost quasiperiodic orbit is due to the interaction between the latter and the 5–periodic hyperbolic orbit (in Fig. 1(a) we just see two of the five 'hyperoblic points'). Indeed, the motion in the vicinity of a hyperbolic point is mainly determined by its associated stable and unstable manifolds (or stable and unstable separatrices). Therefore, while the motion is confined to this small region of the phase space, regular nearby orbits diverge exponentially during certain interval Δt. The latter interval is the interaction time between both orbits. The width of a pulse is then a measure of Δt. The drift is due to the cumulative changes produced by the latter interaction. It is important to remark that while I is an averaged quantity, J is not. Thus I will not 'see' the periodical variations of $\rho_{\gamma'}$, but of course, the drift will do. On the other hand J will be sensitive to both effects.

In Fig. 3 we show the time evolution of I and J for the orbits considered in Fig. 2 and using the A3 procedure (note the scale in I) . We see that $\log I$ has a logarithmic dependence with T for $\mathbf{x}_0 \in \Sigma_r$ (see (23)) while it varies nearly linear for $\mathbf{x}_0 \in \Sigma_c$ (the first in (26)). As we mentioned above, the smallest value of I corresponds to $\mathbf{x}_0$ very close to the stable 1–periodic orbit. Also J behaves in the expected way: for $\mathbf{x}_0 \in \Sigma_r$ and $T \gtrsim 10^2 T_D$, $J \approx 2$ (see (25)) while it depends linearly with T for $\mathbf{x}_0 \in \Sigma_c$ and $T > T_c$ (the second in (26)). These figures confirm that the time–rate of exponential divergence is slower for $\mathbf{x}_0$ in the stochastic layer. Besides, we see that for the quasiperiodic orbit close to the separatrix, $J \to 2$ in a different way than the other regular orbits. We observe that, in this case, J reaches a maximum value and then it decreases asymptotically to 2. This is just a consequence of the effect observed in Fig. 2(b): we are seeing the pulses and the drift. The reason to observe just one instead of several peaks (as in Fig. 2(b)) is due to the fact that J is averaged (in the sense mentioned above) and therefore, only the first peak, for small T, is significant in this case. Note that for $\mathbf{x}_0 \in \Sigma_c$, I and J cover the whole time interval (both curves, $I(T)$ and $J(T)$ escape from Fig. 3) while in Fig. 2(a), these curves for $\rho_{\gamma'}/\rho_\gamma$ end at $T \lesssim 10^3$. Indeed, these figures clearly show the limitation of the A2 procedure (that was used to perform Fig. 2). In the latter case, the integration was stopped when I became negative. As we have already mentioned, I must be a positive quantity (see Section 2). This is true while the orbit γ' is on an energy surface $M_{h'}$ such that $h' \approx h$ (recall that we took γ' in a slightly different energy level than γ). Then for $\mathbf{x}_0 \in \Sigma_c$ and after some time interval (that for which $\delta \gtrsim X \sim$ the size of the system), γ' will be in an energy surface which is far from M_h and then the procedure A2 leads I to decrease, in almost all the cases, monotonically. However both, A2 and A3, provide the same results for regular motion for all T (except for a scale factor in I) and, for stochastic motion, the results agree up to $T \lesssim 10^3$ and, in any case, for $\delta_0 \geq 10^{-7}$ when the A2 procedure is used.

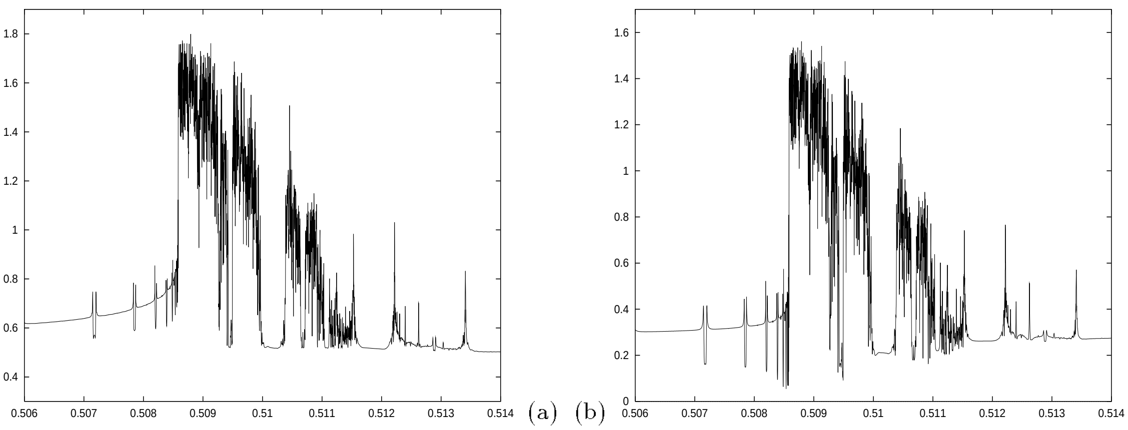
In Fig. 4 we have plotted the final value of $\log I$ for $T = 4000$, and for $\mathbf{x}_0$ along the q_2 axis for several values of q_{2_0} within the 1–periodic island. From Fig. 4(a) we confirm once again that I depends on the initial condition through $\lambda(\mathbf{x}_0)$, its minimum corresponds to the stable 1–periodic orbit. Besides, we observe that I increases very fast as q_{2_0} approaches to the separatrix (recall the logarithmic scale). Fig. 4(b) shows that $J \approx 2$ along the 1–periodic island in accordance with the expected value for regular motion. Both, I and J, are highly sensitive to the presence of small periodic islands. The 'discontinuities' observed in I and J reveal the existance of unstable periodic orbits for the corresponding values of $\mathbf{x}_0$. Fig. 5 is a continuation of Fig. 4 for q_{2_0} across the separatrix of the 5–periodic island. It clearly shows the presence of the stochastic layer as well as the regular regions at both sides of the separatrix ($\log 2 \approx 0.3$). Note that the regular and the stochastic component are separated by several orders of magnitude (see the scale used in the I axis), in accordance with the estimations given above. A simple inspection of (26) shows that Figs. 5(a) and 5(b) should be similar for the stochastic component and this is, in fact, observed.

In Fig. 6(a) we have plotted the final value of $\log J$ for the same set of initial conditions



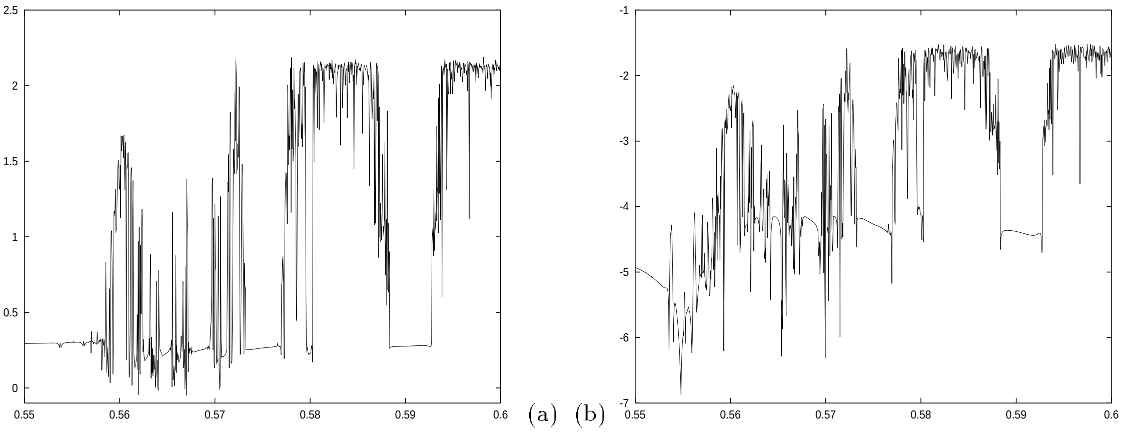
(a) (b)

Figure 4: (a) $\log I$ vs. q_{2_0}, (b) J vs. q_{2_0} for 3000 initial conditions within the 1–periodic island and $T = 4000$.



(a) (b)

Figure 5: The same as Fig. 4 but for q_{2_0} in the neighbourhood of the separatrix of the 5–periodic island for 4000 initial conditions and $T = 4000$. (a) $\log(\log I)$ vs q_{2_0}; (b) $\log J$ vs. q_{2_0}, $\log 2 \approx 0.3$.



(a) (b)

Figure 6: (a) $\log J$ vs. q_{2_0}, (b) $\log \sigma_E$ vs. q_{2_0} for the same set of initial condition than Fig. 1(b) but $T = 7000$ (see text for details).

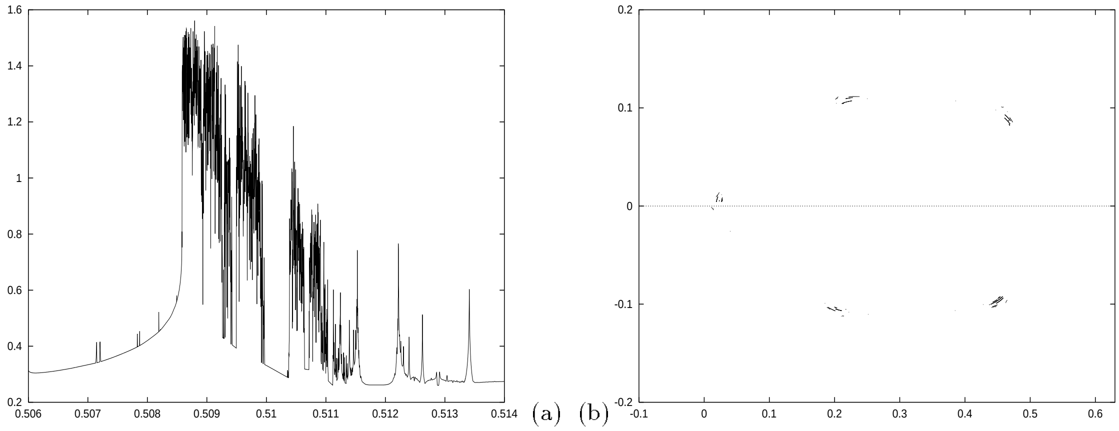


Figure 7: (a) $\log J_{max}$ vs. q_{2_0} in the neighbourhood of the separatrix for 4000 initial conditions. (b) 'Surface of section' $(q_2, p_2)_{T \approx T_m}$ for 1300 initial conditions in the interval $0.506 \le q_{2_0} \le 0.50859$, (see text for details).

than that used for the computation of the LCN but for $T = 7000$. We see again that $J \approx 2$ in the ordered component while J is clearly much larger that 2 in the stochastic regions. In fact, J reaches higher values in the large stochastic sea than in the stochastic layer. Then, because we can separate different stochastic components and, obviously, the regular one, we can state that J provides a measure of chaos. We can easily relate J with the LCN. Indeed, recalling the second of (26) we see that $dJ/dT \sim LCN$. Therefore since J depends nearly linear with T for $T > T_c$ (see Fig. 3(b)), we can fit by least squares the linear part of the curve $J(T)$. This was done for each initial condition of Fig. 6(a). The expected value for the slope is $\sigma \approx 0$ in the regular component (since $J \approx 2$) while $\sigma \sim$ LCN in the stochastic component. As J is the averaged value of the true derivative, the factor 2 in front of the second of (26) compensates in part the averaging procedure. Fig. 6(b) shows the computed value of σ, σ_E. A comparison of this figure with that for the LCN (Fig. 1(b)) reveals a good agreement between both magnitudes. It is important to remark that while σ_E was computed for $T = 7000$, the LCN was for $T = 2.5 \times 10^5$. Actually, $T = 5000$ is enough to get σ_E while the total motion time used to compute the LCN is, perhaps, not sufficient to get a good asymptotic value. Besides, note that Fig. 6(b) shows the structure of the regular component while Fig. 1(b) does not provide any information about it.

Finally, let us consider the quasiperiodic orbits near the separatrix. As we have already shown in Fig. 3(a), J presents a maximum and then it goes asymptotically to 2. In Fig. 7(a) we show the computed maximum, J_{max}, for several orbits in the vicinity of the separatrix. We see that while we move towards the separatrix from the regular side, J_{max} is slightly sensitive to the presence of small periodic islands (compare with Fig. 5(b)). However the 'continuum' is smooth and 'diverges' on the separatrix. On the other side of the separatrix (stochastic), we can still observe nearly the same 'continuum' that in the regular side but the presence of many sharp 'lines', reveals the existance of the stochastic layer. This structure is similar to that observed in Fig. 5(b) for the borders of the small islands, but in that case the motion is almost all regular (stochastic motion, if it is present, is insignificant). We can use the regular part of this curve as a tool to find out the location of the 5–periodic hyperbolic orbit. Indeed, we have already mentioned that the existence of a maximum in J is due to the presence, somewhere, of this unstable orbit. Therefore if we compute the time T_m for which $J(T_m) = J_{max}$, then we can integrate each initial condition (in the regular part) up to $T \approx T_m$, and then plot the final value (q_2, p_2) when $q_1 = 0,\ p_1 \geq 0$. This is shown in Fig. 7(b). A comparison of this figure with Fig. 1(a) confirms that the maximum value observed in J is really due to the interaction between the quasiperiodic orbits and the unstable periodic orbit.

Conclusions

In this paper we showed that the Conditional Entropy, defined through the arc lenght parameter along the orbits, is an efficient tool to explore the phase space in short motion times. It is important to mention that this method provides an easy way to find out the location of unstable periodic orbits. The preliminary analytical results agree with that obtained by numerical simulations (at least for very simple systems). Besides, the mean rate of exponential divergence of nearby orbits can be estimated using this technique, but for motion times which are two or three orders of magnitude less than that for the computation of the true LCN.

However, we would like to remark that there are some topics that still remain to be done: i) a more complete theory; ii) numerical study of 3D models and iii) application to realistic dynamical systems. All of these issues will be addressed in next papers.

Acknowledgements

The first author would like to acknowledge the Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina who supports his visit to the University of Barcelona. The second author has been supported by DGICYT grant PB 94-0215 (Spain). Partial support from the EC grant ERBCHRXCT940460 and the catalan grant CIRIT 1996S0GR-00105 is also acknowledged.

References

[1] Arnold, V. & Avez, A. 1989, *Ergodic Problems of Classical Mechanics*, (New York: Addison–Wesley), 2nd. ed.

[2] Contopoulos, G. & Voglis, N. 1996, *Cel. Mech. & Dynam. Astron.* **64**, 1

[3] Fraser, A. & Swinney, H. 1986, *Phys. Rev. A* **33**, 1134

[4] Hénon, M. & Heiles, C. 1964, *AJ* **69**, 73

[5] Katz, A. 1967, *Principles of Statistical Mechanics, The Information Theory Approach*, (San Francisco: W. H. Freeman & Co.)

[6] Laskar, J. 1990, *Icarus* **88**, 266

[7] Laskar, J. 1993, *Physica D* **67**, 257

[8] Merritt, D. & Valluri, M. 1996, *ApJ* **471**, 82

[9] Núñez, J., Cincotta, P. & Wachlin F. 1996, *Cel. Mech. & Dynam. Astron.* **64**, 43

[10] Udry, S. & Pfenniger, D. 1988, *A&A* **198**, 135

[11] Voglis, N. 1998, this volume

[12] Wachlin, F. & Ferraz–Mello, S. 1998, *MNRAS* **298**, 22