Entropy and irreversibility in dynamical systems

Oliver Penrose
Department of Mathematics and
the Maxwell Institute for Mathematical Sciences,
Colin Maclaurin Building, Heriot-Watt University,
Riccarton, Edinburgh EH14 4AS, Scotland, UK

February 16, 2012

Abstract

A method of defining non-equilibrium entropy for a chaotic dynamical system is proposed which, unlike the usual method based on Boltzmann's principle $S = k \log W$, does not involve the concept of a macroscopic state. The idea is illustrated using an example based on Arnold's 'cat' map. The example also demonstrates that it is possible to have irreversible behaviour, involving a large increase of entropy, in a chaotic system with only two degrees of freedom.

Keywords

irreversibility, entropy, Boltzmann's principle, Arnold map, macrostates, chaotic dynamical system

1 Introduction

It is a part of everyday experience that matter behaves irreversibly: heat flows from hot to cold, never from cold to hot; plates break when we drop them, but never reconstitute themselves, and so on. A way of quantifying this irreversibility is provided by the second law of thermodynamics, which can be encapsulated in the statement that the thermodynamic entropy of a thermally isolated system cannot decrease.

Irreversibility is displayed by most of the partial differential equations we use to model the macroscopic behaviour of matter — the heat equation, the Navier-Stokes equations, and so on (though not the Euler equations). The irreversibility comes from the lack of symmetry of these PDEs under time reversal — the fact that they are not invariant under the transformation $t \to -t, v \to -v$ (reversal of the sign of the time variable and all velocity variables). From such PDEs it is usually possible to derive a result corresponding to the second law of thermodynamics. If the boundary conditions permit no heat to cross the boundary, this result takes the form that some quantity which can be interpreted as the entropy increases with time or stays constant. Likewise, Boltzmann's integro-differential equation for the time evolution of the velocity distribution in a gas is not symmetrical under time reversal, and Boltzmann showed, in his celebrated H theorem, that the entropy-like quantity $-H$ increases with time or stays constant[1].

For the 'microscopic' descriptions used in statistical mechanics to describe the motion of individual molecules in detail, the situation is different. The differential equations of particle dynamics (Hamilton's equations of motion in classical mechanics, Schrödinger's time-dependent equation in quantum mechanics) are symmetric under time reversal. From this symmetry it follows that in particle dynamics there cannot be a dynamical variable that increases with time for every solution of the equations of motion. The reason is that, for every motion on which a dynamical variable increases with time, there must be another motion, obtained from the first one by time reversal, on which the same dynamical variable decreases with time.

What this paradoxical situation reveals is that the microscopic models, even though they contain so much more detail than the PDE models, are incomplete in the sense that their differential equations do not capture the difference between plausible motions (such as heat moving from hot to cold) and implausible ones (such as heat moving from cold to hot). One way of tackling this incompleteness might be to append to the differential equations a criterion for ruling out the implausible motions. This could take the form of a dynamical variable that is equal to the entropy; then motions for which the entropy decreased with time would be recognized as implausible.

An important method for defining entropy in a non-equilibrium system is Boltzmann's principle[2, 3, 4]. To formulate this principle, let us denote by Γ the phase space, consisting of all possible dynamical states of the system, and by Γ_M the set of all dynamical states compatible with a given macroscopic state M. Then the entropy of the system, when it is in the macroscopic state M, can be defined as

$$S = k \log(c\, \mu(\Gamma_M)) \tag{1}$$

where k is Boltzmann's constant, $\mu(\Gamma_M)$ is the measure of the set Γ_M, and c is a constant depending on the number and type of particles in the system, which is necessary in general for consistency with the thermodynamic entropy but can be taken equal to 1 for the very simple systems considered in this paper.

While the definition (1) of entropy has the virtue of being simple in principle, there are some difficulties. One of them is that the definition of a macrostate is quite vague. If one defines macrostates in terms of how many particles are in a particular region of space, for example, then it is not clear what is meant by saying that two values for this particle number are macroscopically identical. Suppose there are a million particles in that region. Would we say that this is macroscopically identical to 999,999? to 999,000?

A second difficulty is that the very notion of 'macroscopic' seems to presuppose that the system consists of a very large number of particles. But later in this paper an example will be given showing that irreversible behaviour is possible in a system with only two degrees of freedom. One would like to have a definition of entropy that can be used for systems of any size, or ones that do not consist of particles at all.

The purpose of the present paper is to propose an alternative to the definition (1) of entropy which can be used for any dynamical system at all, regardless of the number of degrees of freedom and of whether the concept of macroscopic state applies.

2 Dynamical self-correlation and non-equilibrium entropy

In this section a purely mechanical definition is proposed for the "entropy" of a chaotic dynamical system. It defines an entropy for a segment of a trajectory. Very roughly, this entropy is the logarithm of the measure of the part of phase space that can be reached from phase points near the ends of the trajectory segment during the time while it is being traversed.

Let Γ denote the phase space and γ a general point in Γ, and let ϕ_t be the flow or iterated mapping on Γ which defines the time evolution. It will be assumed that ϕ_t has time-reversal invariance. This means that there is an involution operator T with the property that if $\{\gamma(t)\}_{t\in(-\infty,\infty)}$ is a trajectory of the dynamical system (i.e. a solution of its equations of motion), then the set $\{T\gamma(-t)\}_{t\in(-\infty,\infty)}$ is also a trajectory. For example, in Hamiltonian mechanics, T is the operator that reverses all velocities and/or momenta, but leaves the positions invariant. The assumed time-reversal symmetry of the dynamical system can be written

$$T\phi_t(T\phi_t(\gamma)) = \gamma \quad \text{or} \quad T\phi_t(\gamma) = \phi_{-t}(T\gamma) \quad \forall\gamma \in \Gamma \tag{2}$$

We assume further that there is a measure μ on Γ which is preserved by ϕ_t, i.e. that $\mu(\phi_t(A)) = \mu(A)$ for every measurable set A in phase space. In classical Hamiltonian mechanics μ can be the Lebesgue measure on phase space.

Consider now a trajectory segment whose ends are γ_1 (arbitrary) and $\gamma_2 := \phi_{t_{12}}(\gamma_1)$ where t_{12} is arbitrary and may have either sign. That is to say, t_{12} is arbitrary and γ_1, γ_2 are arbitrary subject to the condition $\gamma_2 := \phi_{t_{12}}(\gamma_1)$. Choose any small number ϵ and define the *dynamical self-correlation* of the two end points of the trajectory segment to be

$$C_\epsilon(\gamma_1, \gamma_2) := \frac{\mu(\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2))}{\mu(B_\epsilon)} \tag{3}$$

where $B_\epsilon(\gamma)$ denotes the ball of radius ϵ centred at the phase point γ, and $\mu(B_\epsilon) := \mu(B_\epsilon(\gamma))$, the measure of a ball of radius ϵ, which is the same for all γ.

The dynamical self-correlation has the obvious properties

$$C_\epsilon(\gamma_1, \gamma_1) = 1, \quad 0 \le C_\epsilon(\gamma_1, \gamma_2) \le 1 \tag{4}$$

Moreover, the function $C_\epsilon(\cdot, \cdot)$ is symmetric in its two arguments:

$$C_\epsilon(\gamma_2, \gamma_1) = C_\epsilon(\gamma_1, \gamma_2) \tag{5}$$

This property follows from the fact that the mapping ϕ_t preserves measure, so that the sets $\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2))$ and $\phi_{t_{21}}(\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2))) = B_\epsilon(\gamma_1) \cap \phi_{t_{21}}(B_\epsilon(\gamma_2)$ (where $t_{21} := -t_{12}$) have the same measure and hence the numerators in the definitions of $C_\epsilon(\gamma_1, \gamma_2)$ and $C_\epsilon(\gamma_2, \gamma_1)$ are equal.

It can also be shown, using the symmetry properties of ϕ_t and B_ϵ with respect to the involution T, that the function $C_\epsilon(\cdot, \cdot)$ is invariant under the time reversal involution

$$C_\epsilon(T\gamma_1, T\gamma_2) = C_\epsilon(\gamma_1, \gamma_2) \tag{6}$$

Informally, $C_\epsilon(\gamma_1, \gamma_2)$ as defined in (3) can (for positive t_{12}) be interpreted as the conditional probability that, if the system is started at time t_1 from a phase point chosen at random from the neighbourhood $B_\epsilon(\gamma_1)$ then its phase point at the later time t_2 will lie in the neighbourhood $B_\epsilon(\gamma_2)$. The larger the region of phase space the phase point can stray to within the time interval $t_2 - t_1$, the less likely it is to find its way to $B_\epsilon(\gamma_2)$ at the appointed time, and so we might expect $C_\epsilon(\gamma_1, \gamma_2)$ to be inversely proportional to the volume of phase space the system can reach from phase points near γ_1 during the available time. According to Boltzmann's principle, the entropy associated with this amount of phase space is a constant plus k times the logarithm of its volume, so we may expect to interpret k times the logarithm of $1/C_\epsilon(\gamma_1, \gamma_2)$ as an entropy.

To formulate a quantitative relation between dynamical self-correlation and entropy, consider the behaviour of the dynamical self-correlation when the length of the trajectory segment is very large. According to the informal interpretation just mentioned, $C_\epsilon(\gamma_1, \gamma_2)$ is the probability that, at the time t_2, the phase point will be found in the region $B_\epsilon(\gamma_2)$. For very large t_2, assuming the dynamical system to be chaotic, one may plausibly equate this probability with the equilibrium probability of finding the phase point in $B_\epsilon(\gamma_2)$. If the entire phase space is ergodic (meaning that there are no invariants of the motion, not even an energy invariant), this equilibrium probability is

$$\frac{\mu(B_\epsilon(\gamma_2))}{\mu(\Gamma)} \tag{7}$$

Using the fact that $\mu(B_\epsilon(\gamma)$ is independent of γ we arrive at the following conjecture :

Conjecture 1 *If the entire phase space is ergodic, then*

$$\lim_{t_{12} \to \infty} C_\epsilon(\gamma_1, \gamma_2) = \frac{\mu(B_\epsilon)}{\mu(\Gamma)} \tag{8}$$

Some support for this conjecture is given by the result in section 3 where eqn (8) is shown to hold (albeit for square, not circular, neighbourhoods) in the case of the Arnold 'cat' map.

Using Boltzmann's principle (1) we can use the conjecture (8) to express the equilibrium entropy in terms of dynamical self-correlation :

$$S_{eqm} = k \log \left(\frac{c\, \mu(B_\epsilon)}{\lim_{t_{12} \to \infty} C_\epsilon(\gamma_1, \gamma_2)} \right) \tag{9}$$

for the case where the phase space is ergodic.

Formula (9) gives us the possibility of defining an 'entropy' for any trajectory segment whatever by an analogous formula

$$S_\epsilon(\gamma_1, \gamma_2) := k \log \left(\frac{c\, \mu(B_\epsilon)}{C_\epsilon(\gamma_1, \gamma_2)} \right) \tag{10}$$

In the alternative case where the phase space is not ergodic, the long-time behaviour of C_ϵ as defined in eqn (3) is more complicated, and it seems preferable to work with a revised definition. Consider, for example, the case where there is just one invariant of the motion, the energy. To each value of E there corresponds an 'energy surface', the set $\Gamma_E := \{\gamma : H(\gamma) = E\}$ where $H(\cdot)$ is the Hamiltonian. The invariant measure obtained by restricting the measure μ to Γ_E (the microcanonical measure at energy E) will be denoted by μ_E.

In formulating a definition of dynamical self-correlation, analogous to (3), for this case we have to take account that, although $\mu(B_\epsilon(\gamma)$ is independent of γ, the microcanonical measure of the set $B_\epsilon(\gamma) \cap \Gamma_E$ does depend on γ. At the same time, we would like to preserve the symmetry properties such as (5). These requirements can be met by replacing the $\mu(B_\epsilon)$ in the denominator of formula (3) by the geometric mean of the microcanonical measures of the two relevant neighbourhoods on the energy surface, $B_\epsilon(\gamma_1) \cap \Gamma_E$ and $B_\epsilon(\gamma_2) \cap \Gamma_E$. Thus the 'microcanonical' version of formula (3) is

$$C_\epsilon^{(E)}(\gamma_1, \gamma_2) := \frac{\mu_E(\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2) \cap \Gamma_E)}{\sqrt{\mu_E(B_\epsilon(\gamma_1) \cap \Gamma_E)\, \mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)}} \tag{11}$$

As before, we consider the behaviour of the dynamical self-correlation when the length of the trajectory segment is very large and the conditional probability that the phase point, if started at time t_1 from a point in the ball $B_\epsilon(\gamma_1)$ chosen at random using the microcanonical probability distribution μ_E, will be found in the ball $B_\epsilon(\gamma_2)$ at the later time t_2. This probability is

$$\frac{\mu_E(\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2) \cap \Gamma_E)}{\mu_E(B_\epsilon(\gamma_1) \cap \Gamma_E)} \tag{12}$$

For very large t_2 it is reasonable to equate the conditional probability (12) with the equilibrium probability of finding the phase point in $B_\epsilon(\gamma_2)$, as given by the microcanonical probability measure. This conditional probability is

$$\frac{\mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)}{\mu_E(\Gamma_E)} \tag{13}$$

Thus we expect the ratio of the expressions (12) and (13) to approach the limit 1 as $t_{12} \to \infty$, i.e

$$\lim_{t_{12}\to\infty} \left(\frac{\mu_E(\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2) \cap \Gamma_E)\, \mu_E(\Gamma_E)}{\mu_E(B_\epsilon(\gamma_1) \cap \Gamma_E)\, \mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)} \right) = 1 \tag{14}$$

Combining this formula with the definition (11), we are led to the following analogue of conjecture 1:

Conjecture 2 *If the energy surface Γ_E is ergodic, then*

$$\lim_{t_{12}\to\infty} \left(\frac{C_\epsilon^{(E)}(\gamma_1, \gamma_2)}{\sqrt{\mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)\, \mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)}} \right) = \frac{1}{\mu_E(\Gamma_E)} \tag{15}$$

The analogue of formula (9) for the equilibrium entropy is thus

$$S_{eqm} = \lim_{t_{12}\to\infty} k \log \left(\frac{c\sqrt{\mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)\, \mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)}}{C_\epsilon^{(E)}(\gamma_1, \gamma_2)} \right) \tag{16}$$

and the analogue of (10) for the non-equilibrium entropy is

$$\begin{aligned} S_\epsilon(\gamma_1, \gamma_2) \quad &:= \quad k \log \left(\frac{c\sqrt{\mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)\, \mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)}}{C_\epsilon^{(E)}(\gamma_1, \gamma_2)} \right) \\ &= \quad k \log \left(c\, \frac{\mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)\, \mu_E(B_\epsilon(\gamma_2) \cap \Gamma_E)}{\mu_E(\phi_{t_{12}} B_\epsilon(\gamma_1) \cap B_\epsilon(\gamma_2) \cap \Gamma_E)} \right) \end{aligned} \tag{17}$$

3 Example 1: Arnold's 'cat' map

To illustrate some of the properties of the entropy definition (10), we apply it to two dynamical systems which are simple enough to permit some of the relevant quantities to be calculated exactly.

The first example is a system with discrete dynamics, whose phase space is a square $Q_L := [0, L) \otimes [0, L)$ of arbitrary side length L, with opposite edges identified so as to make it a two-dimensional torus. The dynamical rule is the Arnold 'cat' map [5] obtained by multiplying the column vector $[p, q]^T$ representing the phase point γ by the matrix

$$\mathbf{A} := \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \tag{18}$$

and then projecting into the square Q_L using the projection $\mathbf{P}_L$ defined by $\mathbf{P}_L(p, q) := (p \bmod L, q \bmod L)$ in which, for example, $p \bmod L := p - L[p/L]$ where $[p/L]$ denotes the largest integer $\leq p/L$. The formula for a single step of the evolution can be written

$$\phi(p, q) = \mathbf{P}_L(p + q, p + 2q)) := ((p + q) \bmod L, (p + 2q) \bmod L) \tag{19}$$

This dynamical system is reversible in the following sense: let the sequence $(\ldots\gamma_0, \gamma_1, \gamma_2, \ldots)$ be a trajectory, meaning that it satisfies $\gamma_{n+1} = \phi(\gamma_n)$ for all $n \in \mathbf{Z}$; then the sequence $(\ldots T\gamma_2, T\gamma_1, T\gamma_0, \ldots)$, where $T(p, q) := (q, -p)$, is also a trajectory. (Proof: since the matrix

$$\mathbf{T} := \left[\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array} \right] \tag{20}$$

representing the involution T, satisfies $\mathbf{ATA} = \mathbf{T}$, it follows that if $\gamma_{n+1} = A\gamma_n$, then $T\gamma_n = AT\gamma_{n+1}$.)

This dynamicasl system has just one positive Lyapunov exponent, which is the logarithm of the larger of the two eigenvalues of the matrix $\mathbf{A}$. This eigenvalue is G^2 where G denotes the golden ratio $(1 + \sqrt{}5)/2 = 1.618\ldots$. The normalized right eigenvectors of the matrix $\mathbf{A}$ are

$$\begin{array}{rlll} \mathbf{u} := & (1 + G^2)^{-1/2}[1\ G]^T & \text{with eigenvalue} & G^2 = 1 + G, \\ \mathbf{v} := & (1 + G^2)^{-1/2}[-G\ 1]^T & \text{with eigenvalue} & G^{-2} = 2 - G \end{array} \tag{21}$$

To simplify the calculations we replace the ball $B_\epsilon(\gamma)$ used in the preceding section by a square neighbourhood $N_\epsilon(\gamma)$. The edges of $N_\epsilon(\gamma)$ are taken to be parallel to the eigenvectors and their length is 2ϵ, so that $N_\epsilon(\gamma)$ is just big enough to include the ball $B_\epsilon(\gamma)$ but small enough to be included in $B_{\sqrt{2}\epsilon}(\gamma)$. Its corners are the four points $\gamma + \epsilon(\pm\mathbf{u} \pm \mathbf{v})$. The matrix A^t converts $N_\epsilon(\gamma)$ into a rectangle with corners

$$\mathbf{A}^t\gamma + \epsilon(\pm G^{2t}\mathbf{u} \pm G^{-2t}\mathbf{v}) = (x \pm \epsilon)G^{2t}\mathbf{u} + (y \pm \epsilon)G^{-2t}\mathbf{v} \tag{22}$$

where x, y are defined by $\gamma = x\mathbf{u} + y\mathbf{v}$. This rectangle will be called 'the long rectangle'. The lengths of its sides are $2\epsilon G^{2t}$ in the $\mathbf{u}$ direction and $2\epsilon G^{-2t}$ in the $\mathbf{v}$ direction.

To apply the definition of dynamical self-correlation, eqn (3), we need the overlap area of the square $N_\epsilon(\phi^t\gamma)$ with the figure $\phi^t(N_\epsilon(\gamma))$ obtained by applying the projection $\mathbf{P}_L$ to the long rectangle. We shall take $t := t_{12}$ to be positive; the results for negative t can be obtained using the symmetry rule (5) if they should be needed. A simple case arises when $\phi^t\gamma$ is not too close to the edges of Q_L while ϵ and t are small enough for the long rectangle, whose length is $2\epsilon G^{2t}$, to lie entirely inside Q_L. Then the projection $\mathbf{P}_L$ leaves the long rectangle unaltered, and that rectangle's intersection with $N_\epsilon(\gamma)$ is simply a smaller rectangle; the width of the intersection rectangle is the same as that of the long rectangle, namely $2\epsilon G^{-2t}$, and the length of the intersection rectangle is 2ϵ, making its area $4\epsilon^2 G^{-2t}$. Putting this result into the definition (3) we find that, for this dynamical system,

$$C_\epsilon(\gamma_1, \gamma_2) = G^{-2t} \quad \text{if} \quad \epsilon G^{2t} \ll L \quad (\text{and } t > 0) \tag{23}$$

This formula can also be written in terms of the positive Lyapunov exponent $\lambda = \log(G^2)$.

$$C_\epsilon(\gamma_1, \gamma_2) = \mathrm{e}^{-\lambda t} \quad \text{if} \quad \epsilon \mathrm{e}^{\lambda t} \ll L \tag{24}$$

This formula is also true, approximately, when the neighbourhoods are taken to be balls rather than squares.

The calculation of $C_\epsilon(\gamma_1, \gamma_2)$ for the opposite case, in which $\epsilon G^{2t} \gg 1$, is more complicated, and will be given as a theorem:

Theorem 1 *For the discrete dynamical system defined by the mapping (19) the dynamical self-correlation is given by*

$$C(\gamma_1, \gamma_2) = \frac{4\epsilon^2}{L^2}\left(1 + O\left(\frac{L}{\epsilon\, G^{2t}}\right)\right) + O\left(\frac{t + \log(L/\epsilon)}{G^{2t}}\right) \tag{25}$$

The proof of this theorem depends on the following lemma

Lemma 1 *Let p_0 satisfy $0 \leq p_0 < L$, let n_1, n_2 be positive integers and consider the set of points*

$$\Sigma := \mathbf{P}_L\{p_0 + jL/G\}_{j=-n_1,-n_1+1,\ldots,n_2-1,n_2}. \tag{26}$$

where $\mathbf{P}_L(x) := x \bmod L$. Let a, b be numbers satisfying $0 \leq a < b < L$. Then the number of points from the set Σ that lie in the semi-open interval $[a, b)$, which we denote by $\sharp\{\Sigma \cap [a, b)\}$ satisfies

$$\sharp\{\Sigma \cap [a, b)\} = \frac{b-a}{L}n + O(\log n) \tag{27}$$

where $n := n_1 + n_2 + 1$.

Proof. We consider first the case where $n = F_k$ for some k, where $F_0, F_1, F_2, \ldots$ are the Fibonacci numbers, defined by the rule $F_k := F_{k-1} + F_{k-2}$ with $F_0 := 0, F_1 := 1$. In this case the error term turns out to be $O(1)$, a stronger result than the $O(\log n)$ in eqn (27).

The successive continued-fraction approximants to $1/G$ are F_{k-1}/F_k. According to the theory of continued fractions the error in using one of these approximants in place of $1/G$ has the upper bound

$$\left|\frac{1}{G} - \frac{F_{k-1}}{F_k}\right| < \frac{1}{F_k F_{k+1}} \tag{28}$$

Under this approximation each point of $\Sigma := \mathbf{P}_L\{p_0 + jL/G\}_{j=-n_1,\ldots,n_2}$ is approximated by a point from the set

$$\Sigma^{(k)} := \mathbf{P}_L\left\{p_0 + jLF_{k-1}/F_k\right\}_{j=-n_1\ldots n_2} \tag{29}$$

The set $\Sigma^{(k)}$ comprises $n = F_k$ points. Since F_{k-1} and F_k have no common factor, the F_k different numbers $-n_1F_{k-1}\ldots n_2F_{k-1}$ all give different remainders on division by F_k (for if two were to give the same remainder, their difference would at the same time be a multiple of F_k and the product of F_{k-1} by a number less than F_k, which cannot be). Therefore each of the F_k possible remainders occurs just once. Consequently, denoting the remainders by r, the set $\Sigma^{(k)}$ can be written

$$\Sigma^{(k)} := \mathbf{P}_L \left\{p_0 + rL/F_k\right\}_{r=0,\ldots,F_{k-1}} \tag{30}$$

Thus the set $\Sigma^{(k)}$ comprises $n = F_k$ points, which are equally spaced along the interval $(0, L)$, their separation being L/n.

By (28) the error in approximating the number jL/G by jLF_{k-1}/F_k is at most jL/F_kF_{k+1}, whose magnitude is bounded above by L/F_{k+1} since $|j| \leq n = F_k$.

Suppose k is large enough to make $L/F_{k+1} < \frac{1}{2}(b-a)$ and consider the open interval $[a + L/F_{k+1}, b - L/F_{k+1})$, whose length is the positive number $b - a - 2L/F_{k+1}$. Denote the number of points of $\Sigma^{(k)}$ that lie in this interval by m_k. Each of these points lies within a distance L/F_{k+1} of the corresponding point of Σ; therefore the corresponding points of Σ all lie within the larger interval $[a, b)$ and it follows that $\sharp\{\Sigma \cap [a, b)\} \geq m_k$. Moreover, since the separation of the points of $\Sigma^{(k)}$ is L/n, the length of an interval occupied by m_k of them is at most $(m_k + 1)L/n$ and so the number of them in the interval $[a + L/F_{k-1}, b - L/F_{k-1})$ satisfies $(m_k + 1)L/n \geq b - a - 2L/F_{k-1}$. Putting together these two inequalities we obtain

$$\sharp\{\Sigma \cap [a, b)\} \quad \geq \quad m_k \geq (n/L)(b - a - 2L/F_{k-1}) - 1 \tag{31}$$

In a similar way it can be shown that $\sharp\{\Sigma \cap [a, b)\}$ is bounded above by the number of points of $\Sigma^{(k)}$ in the interval $[a - L/F_k, b + L/F_k)$, which in turn is bounded above by $(b - a + 2L/F_k)n/L + 1$, so that

$$\sharp\{\Sigma \cap [a, b)\} \leq (b - a + 2L/F_k)n/L + 1 \tag{32}$$

The upper and lower bounds (32) and (31) taken together imply (27), and the proof of the lemma for the case where n is Fibonacci number is complete.

For the case where n is not a Fibonacci number, choose k to be the largest integer for which

$$n > F_k \tag{33}$$

Let k' be the largest integer for which $n - F_k \geq F_{k'}$; if $n - F_k \neq F_{k'}$ let k'' be the largest integer for which $n - F_k - F_{k'} \geq F_{k''}$ and so on. In this way we can express n as a finite sum of decreasing Fibonacci numbers:

$$n = F_k + F_{k'} + F_{k''} + \ldots \tag{34}$$

Corresponding to the decomposition (34) of the number n, we can decompose the set Σ defined in (26) into a finite number of subsets, the first comprsing F_k points, the second $F_{k'}$ points and so on :

$$\Sigma = \Sigma_k \cup \Sigma'_k \cup \Sigma''_k \cup \ldots \tag{35}$$

where

$$\begin{aligned}
\Sigma_k &:= \mathbf{P}_L\{p_0 + (i - n_1)L/G\}_{i=0,1,\ldots,F_k-1} \\
\Sigma'_k &:= \mathbf{P}_L\{p_0 + (i' - n_1 + F_k)L/G\}_{i'=0,1,\ldots,F_{k'}-1} \\
\Sigma''_k &:= \mathbf{P}_L\{p_0 + (i'' - n_1 + F_k + F_{k'})L/G\}_{i''=0,1,\ldots,F_{k''}-1} \qquad (36)
\end{aligned}$$

etc. . Because of the irrationality of G these sets are disjoint.

In the set Σ_k, we follow the procedure used in the first part of this proof, approximating $1/G$ by F_{k-1}/F_k and arriving at the result

$$\sharp\{\Sigma_k \cap [a,b)\} = (b-a)F_k/L + O(1) \qquad (37)$$

For the set Σ'_k we follow an analogous procedure, but approximating $1/G$ by $F_{k'-1}/F_{k'}$ this time. This gives the result

$$\sharp\{\Sigma'_k \cap [a,b)\} = (b-a)F_{k'}/L + O(1) \qquad (38)$$

Similarly, approximating $1/G$ by $F_{k''-1}/F_{k''}$ in Σ''_k, we obtain

$$\sharp\{\Sigma''_k \cap [a,b)\} = (b-a)F_{k''}/L + O(1) \qquad (39)$$

and so on. Adding up the equations (37), (38), (39),etc., and using the formulas (34) and (35), together with the fact that the sets $\Sigma_k, \Sigma'_k, \Sigma''_k, \ldots$ are disjoint, we obtain

$$\sharp\{\Sigma \cap [a,b)\} = (b-a)n/L + O(\log n) \qquad (40)$$

since the number of terms in the finite sum (34) is bounded above by k which is $O(\log n)$ by Binet's formula $F_k = (G^k - (-G)^{-k})/\sqrt{5}$. This completes the proof of the lemma.

Proof of theorem. We want to evaluate

$$C_\epsilon(\gamma_1, \gamma_2) := \frac{\mu(\phi_{t_2-t_1}N_\epsilon(\gamma_1) \cap N_\epsilon(\gamma_2))}{\mu(N_\epsilon)} \qquad (41)$$

as defined in eqn (3), using phase points $\gamma_1 = x\mathbf{u} + y\mathbf{v}$ and $\gamma_2 = \mathbf{P}_L(xG^{2t}\mathbf{u} + yG^{-2t}\mathbf{v})$ where $t := t_2 - t_1$, and the neighbourhoods taken to be squares with sides of length 2ϵ parallel to the eigenvectors $\mathbf{u}, \mathbf{v}$. For simplicity we assume that the distances from γ_1 and γ_2 to the edges of Q_L are greater than ϵ, so that both their neighbourhoods are inside Q_L. Then the corners of $N_\epsilon(\gamma_1)$ are the phase points $(x \pm \epsilon)\mathbf{u} + (y \pm \epsilon)\mathbf{v}$ and those of $N_\epsilon(\gamma_2)$ are $\mathbf{P}_L((xG^{2t} \pm \epsilon)\mathbf{u} + (yG^{-2t} + \epsilon)\mathbf{v})$. The region $\phi_t N_\epsilon(\gamma_1)$ is obtained by applying the projection operator $\mathbf{P}_L$ to the 'long rectangle' with corners $G^{2t}(x \pm \epsilon)\mathbf{u} + G^{-2t}(y \pm \epsilon)\mathbf{v}$. The centre of the long rectangle is the point $\gamma_2 = xG^{2t}\mathbf{u} + yG^{-2t}\mathbf{v}$ and (for positive t) its length is $2G^{2t}\epsilon$ in the $\mathbf{u}$ direction and its width is $2G^{-2t}\epsilon$ in the $\mathbf{v}$ direction.

The mid-line of the long rectangle is the line in the $\mathbf{u}$ direction joining the points $G^{2t}(x \pm \epsilon)\mathbf{u} + G^{-2t}y\mathbf{v}$. The length of this line is $2G^{2t}\epsilon$ in the $\mathbf{u}$ direction. The line, projected if necessary, meets the (horizontal) p axis at a point $(p_0, 0)$ where $p_0 = p_2 - q_2/G$. and (p_2, q_2) are the Cartesian coordinates of γ_2. Each intersection of this line with a horizontal line $q =$ integer × L will map, under the projection operator $\mathbf{P}_L$, to an intersection with the line $q = 0$, i.e. the p-axis. The number of such intersections (the n of Lemma 1) is, with accuracy $O(1)$, equal to $1/L$ times the length of the projection of the mid-line of the long rectangle onto the q axis, i.e. $(1/L)\, 2G^{2t}\epsilon\, G(1+G^2)^{-1/2} + O(1)$ (see eqn (21)). Using this expression in place of the number n in the lemma (eqn (27)), we find that the number of intersections of the image of the mid-line of the long rectangle with an arbitrary interval of length $b-a$ on the p axis is

$$\frac{(b-a)}{L}\left(\frac{2\epsilon\, G^{2t}G(1+G^2)^{-1/2}}{L} + O(1)\right) + O\left(\log\left(\frac{2\epsilon G^{2t}G(1+G^2)^{-1/2}}{L}\right)\right) \tag{42}$$

For the interval $[a,b)$ we choose the part of the p axis through which all lines in the $\mathbf{u}$ direction through $N_\epsilon(\gamma_2)$ (extended if necessary) pass. The length of this interval is $(b-a) = 2\epsilon(1+G^2)^{1/2}/G$; so the formula (42) tells us that the number of intersections of $N_\epsilon(\gamma_2)$ with the image (under $\mathbf{P}_L$) of the mid-line of the long rectangle is

$$\begin{aligned}&\frac{2\epsilon(1+G^2)^{1/2}}{LG}\left(\frac{2\epsilon\, G^{2t}G(1+G^2)^{-1/2}}{L} + O(1)\right) + O\left(\log\left(\frac{\epsilon G^{2t}}{L}\right)\right)\\ &= \frac{2\epsilon}{L}\left(\frac{2\epsilon\, G^{2t}}{L} + O(1)\right) + O\left(|t| + |\log\left(L/\epsilon\right)|\right)\end{aligned} \tag{43}$$

Each of these intersections is a line segment, nearly all having length 2ϵ. Each of these line segments is the midline of the intersection of $N_\epsilon(\gamma_2)$ with part of the image of the long rectangle; and each of these intersection rectangles, with at most two exceptions, has length 2ϵ, width $2\epsilon G^{-2t}$ and hence area $2\epsilon \times 2\epsilon G^{-2t} = 4\epsilon^2 G^{-2t}$. Multiplying by the number of such intersection rectangles, given in eqn (43), we find the total intersection area of $N_\epsilon(\gamma_2)$ with the projection of the long rectangle to be

$$\begin{aligned}&4\epsilon^2 G^{-2t}\left[\frac{2\epsilon}{L}\left(\frac{2\epsilon\, G^{2t}}{L} + O(1)\right) + O\left(t + \log\left(L/\epsilon\right)\right)\right]\\ &= \frac{16\epsilon^4}{L^2}\left(1 + O\left(\frac{L}{\epsilon\, G^{2t}}\right)\right) + O\left(\frac{\epsilon^2}{G^{2t}}\left(t + \log\left(L/\epsilon\right)\right)\right)\end{aligned} \tag{44}$$

Using this result to evaluate the numerator of eqn (41), together with the obvious formula $4\epsilon^2$ for the denominator, we conclude that

$$C(\gamma_1,\gamma_2) = \frac{4\epsilon^2}{L^2}\left(1 + O\left(\frac{L}{\epsilon\, G^{2t}}\right)\right) + O\left(\frac{t + \log(L/\epsilon)}{G^{2t}}\right) \tag{45}$$

This completes the proof of the theorem.

4 Example 2: a 'cat and kitten' map

Our second example is a dynamical system whose phase space comprises two squares on two different copies of $\mathbf{R}^2$. One of the squares is $Q := [0,L) \otimes [0,L)$;

the other is $Q' := [0, L') \otimes [0, L')$ where L and L' are positive integers. In the interesting case, L' is chosen much larger than L. For each square the opposite edges are identified so as to make it a two-dimensional torus.

The rule specifying the motion of the phase point is that, provided the phase point is outside a particular small region in $Q \cap Q'$, which will be called the 'window', it follows the Arnold dynamical rule (19) appropriate to the torus it is in; but if the phase point lands on the window it jumps to the other torus before continuing. We shall take the window to be a square neighbourhood of some point $\omega \in Q \cap Q'$, namely $N_\delta(\omega)$ where δ is a small constant. The point ω should be chosen so that $N_\delta(\omega) \cap \phi(N_\delta(\omega)) = \emptyset$, *i.e.* the phase point does not immediately jump back again at the next step. The choice $\omega = (\frac{1}{2}, \frac{1}{2})$ would be appropriate, for example.

For this dynamical system the phase point is labelled by three variables: two real variables p and q, plus an extra variable Z which takes only two values: L if the phase point is in the torus Q, and L' if it is in Q'. The analogue of the formula (19) is now

$$\begin{aligned} \phi(p,q,Z) &:= (\mathbf{P}_Z(p+q, p+2q), Z) \quad \text{if } (p,q) \notin N_\delta(\omega) \\ \text{but} &:= (\mathbf{P}_{L+L'-Z}(p+q, p+2q), L+L'-Z) \quad \text{if } (p,q) \in N_\delta(\omega) \end{aligned} \tag{46}$$

Without going into any rigorous analysis, it is plausible that, over a very long time interval, all or almost all trajectories will spend most of their time in the larger torus, but will make occasional excursions into the smaller torus. It is a reasonable conjecture that the probability per time step of hitting the window and thereby moving to the other torus is the small number δ^2/L^2 when the phase point is in Q, and the even smaller number $\delta^2/{L'}^2$ when the phase point is in Q'. Thus we can estimate the duration of each sojourn in Q to be of order L^2/δ^2, whilst the duration of each sojourn in Q' is of order L'^2/δ^2, a much larger number.

Carrying this type of reasoning a bit further we can obtain conjectural information about the dynamical self-correlations. Consider first the case where both γ_1 and γ_2 are in the larger torus Q'. Then, since most trajectories spend only a tiny fraction of their time in the smaller torus Q, the dynamical self-correlations will be very close to what they would be if the smaller torus did not exist at all; so, from (23) and (25), we may expect that (with $t > 0$ as usual)

$$\begin{aligned} \text{if} \quad \gamma_1, \gamma_2 \in Q' \quad \text{then} \quad C_\epsilon(\gamma_1, \gamma_2) &\approx G^{-2t} \quad \text{when} \quad G^{2t} < L/\epsilon \\ \text{but} &\approx 4\epsilon^2/L'^2 \quad \text{when} \quad G^{2t} \gg L/\epsilon \end{aligned} \tag{47}$$

If, on the other hand, γ_1 and γ_2 are both in the smaller torus, things are more complicated, since the fraction of the relevant trajectories that visit the other torus is significantly larger. Estimating the probability per time step of escaping from the smaller torus as δ^2/L^2, the condition for the analogue of (47) to hold is $t\delta^2/L^2 \ll 1$, so that we may expect

$$\begin{aligned} \text{if} \quad \gamma_1, \gamma_2 \in Q \quad \text{then} \quad C_\epsilon(\gamma_1, \gamma_2) &\approx G^{-2t} \quad \text{when} \quad G^{2t} < L/\epsilon \\ \text{but} &\approx 4\epsilon^2/L^2 \quad \text{when} \quad L/\epsilon \ll G^{2t},\ t \ll L^2/\delta^2 \end{aligned} \tag{48}$$

In place of the $4\epsilon^2/L^2$ in the last line, the formula $(4\epsilon^2/L^2)\mathrm{e}^{-t\delta^2/L^2}$ would probably hold over a larger time range.

The reader may find it interesting to work out what happens in the third case, for which the two ends of the trajectory segment are in different toruses.

Using (47) and (48) we can now evaluate the entropies of various trajectory segments as defined by the formula (10) with $G^{2t} \gg L/\epsilon$. The results are

$$\begin{aligned} S = k\log\frac{\mu(B_\epsilon)}{C_\epsilon(\gamma_1,\gamma_2)} \quad &\approx k\log L'^2 \quad \text{if} \quad \gamma_1,\gamma_2 \in Q' \\ \text{but} \quad &\approx k\log L^2 \quad \text{if} \quad \gamma_1,\gamma_2 \in Q \quad \text{and } t \ll L^2/\delta^2 \end{aligned} \tag{49}$$

From the first of these formulas, eqn (47), the entropy when γ_1 and γ_2 are both in the large torus is

$$k\log\frac{\mu(B_\epsilon)}{\lim_{t_{12}\to\infty} C_\epsilon(\gamma_1,\gamma_2)} = k\log L'^2 \tag{50}$$

According to eqn (47) this expression should be equal to the equilibrium entropy, and (since $L' \gg L$) it is indeed very close to the exact equilibrium entropy, which according to Boltzmann's formula (1) is $k\log(L'^2 + L^2)$.

From the second, eqn (48), we can get a non-equilibrium entropy for phase points in the smaller torus, appropriate for the time scale $t \ll L^2/\delta^2$ during which the non-equilibrium state lasts, using the formula (10):

$$S_{noneqm} = k\log\frac{\mu(B_\epsilon)}{C_\epsilon(\gamma_1,\gamma_2)}\bigg|_{\log(L/\epsilon)\ll t\ll L^2/\delta^2} = k\log L^2 \tag{51}$$

There is nothing surprising about the entropy formula (51). The expression $k\log L^2$ could have been obtained much more quickly by the *ad hoc* procedure of making a small change in the dynamics, namely closing the window completely so that the phase space splits into two mutually inaccessible parts, and then applying Boltzmann's principle to whichever part the phase point happens to be in. What the calculations illustrate, however, is that the formula (10) provides a self-contained definition of non-equilibrium entropy which requires no *ad hoc* procedures and no additional input about macroscopic descriptions. All that is necessary is to get the right time scale — the duration of a trajectory which is long enough for the initial 'transient' exponential decay to have died out ($t \gg \log L/\epsilon$) but not so long that the slow approach to equilibrium can have a significant effect ($t \ll L^2/\delta^2$).

An interesting feature of this dynamical system is that it can behave irreversibly, even though it does not satisfy the usual criteria that an irreversible system is supposed to obey. Irreversibility is generally held to be a property of systems that are large in the sense of comprising a large number of particles (and therefore having many degrees of freedom), and to reveal itself in the time evolution of macroscopic variables, such as the local density, which are defined in terms of averages over a large number of particles. But the two-torus system considered here has only two degrees of freedom, and there are no particles over which to define macroscopic variables as averages. Nevertheless, this dynamical system has the ability to behave irreversibly. If it is started at a randomly chosen point in the smaller torus Q, the chances are that, after a time of order L^2/δ^2 it will emerge from that torus and that it will then remain in the larger torus for a much longer time, of order L'^2/δ^2, before its next visit to the smaller torus. By choosing L' large enough we can make the return time L'^2/δ^2 as large as we like — larger than the age of the Universe, if desired — making the original transition from torus Q to torus Q' almost literally irreversible.

When the 'irreversible' transition from the the smaller to the larger torus takes place, the entropy increases by the amount $k \log L'^2 - k \log L^2 = 2k \log(L'/L)$, which can be arbitrary large (in comparison with k) depending on how large we choose to make the ratio L'/L. Just as in thermodynamics, the irreversible process is accompanied by a large entropy increase.

5 Conclusion

This paper gives a method for defining an entropy associated with a segment of trajectory in a chaotic dynamical system. The definition is purely dynamical ('microscopic'); it does not depend on any macroscopic or observational description. The definition depends on two parameters: ϵ, the size of the neighbourhood and t_{12}, the length of the trajectory segment. In order to get a useful result, both have to be chosen sensibly — see the conditions in eqn (48)

One way to carry this work forward would be to look for a connection between the entropy defined here and the one in Boltzmann's H theorem. It might also be possible to prove entropy increase results : for example it may be that if one trajectory segment is a subset of another, the larger segment must have the larger entropy. Another topic that could be investigated is the general connection between the dynamical self-correlation at small times and the Lyapunov exponents, illustrated by the formula (24) for the Arnold map.

6 Acknowledgement

I am grateful to John Ball for stimulating discussions which initiated this work, and to the organizers and sponsors of the International Scientific Seminar 'Entropy and Convexity for Nonlinear Partial Differential Equations' co-hosted by the Oxford Centre for Nonlinear PDE and the Royal Society of London, held on 16 and 17 June 2011 at the Kavli Royal Society International Centre in Buckinghamshire, at which these discussions took place. I am also grateful to Joel Lebowitz for discussions about irreversibility over many years.

References

[1] L Boltzmann, Weitere Studieren über das Wärmegliechgewicht unter Gasmolekülen, *Sitzungsberichte der Akademie der Wissenschaften zu Wien* II **66**

275-370 (1872). English translation in ref. [6]

[2] A Einstein, Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes, *Ann der Physik* **33** 1275 (1910)

[3] O Penrose, *Foundations of statistical mechanics* (Dover, 2005), page 170.

[4] J L Lebowitz, Microscopic origins of irreversible macroscopic behavior *Physica A* **263** 516-527 (1999) DOI 10.1016/S0378-4371(98)00514-7

[5] V I Arnold and A Avez, *Ergodic Problems of Classical Mechanics*, Benjamin 1958, section 13

[6] S G Brush, The kind of motion we call heat, *Studies in statistical mechanics*, ed. E W Montroll and J L Lebowitz, vol. VI (North-Holland, Amsterdam 1976)

Short title

entropy and irreversibility