Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chaos, Solitons and Fractals 28 (2006) 26–31
CHAOS SOLITONS & FRACTALS
www.elsevier.com/locate/chaos

Chaos and randomness: An equivalence proof of a generalized version of the Shannon entropy and the Kolmogorov–Sinai entropy for Hamiltonian dynamical systems

Roman Frigg *

Department of Philosophy, Logic and Scientific Method, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

Accepted 9 May 2005

Abstract

Chaos is often explained in terms of random behaviour; and having positive Kolmogorov–Sinai entropy (KSE) is taken to be indicative of randomness. Although seemly plausible, the association of positive KSE with random behaviour needs justification since the definition of the KSE does not make reference to any notion that is connected to randomness. A common way of justifying this use of the KSE is to draw parallels between the KSE and Shannon's information theoretic entropy. However, as it stands this no more than a heuristic point, because no rigorous connection between the KSE and Shannon's entropy has been established yet. This paper fills this gap by proving that the KSE of a Hamiltonian dynamical system is equivalent to a generalized version of Shannon's information theoretic entropy under certain plausible assumptions.

1. Introduction

On one influential suggestion, chaos is best explained in terms of randomness. It is this suggestion that this paper is concerned with. The problem with this suggestion is that to characterize a system's behaviour as random is that the concept of randomness is as much in need of analysis as the notion of chaos itself. Common physical wisdom has it that ergodic theory fits the bill. More specifically, the claim is that the ergodic hierarchy provides a set of concepts which allow for an adequate characterization of random behaviour (see for instance [7,10,14,16,19]). Among the notions introduced in this context, the Kolmogorov–Sinai entropy (KSE) is of particular importance since it is generally assumed that the move from zero to positive KSE marks the transition from regular to chaotic behaviour.

Though plausible at first glance, a closer look at the KSE reveals a conceptual gap between the KSE and random behaviour: the definition of the KSE is phrased in terms of partitions of the phase space and their time evolution and it

* Tel.: +44 20 7955 7182; fax: +44 20 7955 6845.
E-mail address: r.p.frigg@lse.ac.uk

0960-0779/$ - see front matter
doi:10.1016/j.chaos.2005.05.006

does not make reference randomness or any related concept. How then are we to understand the KSE as a indicator of random behaviour? Three suggestions have been made as to how to bridge this gap. First, connect the KSE to positive Lyapunov exponents by invoking Pesin's theorem [12]. Second, take algorithmic complexity to be a measure for randomness and relate the KSE to this notion by dint of Brudno's theorem [3]. Third, one can establish a link between the KSE and Shannon's information theoretic notion of entropy.

This brief summary reveals a lacuna in our theoretical edifice: while the association of the KSE with either positive Lyapunov exponents or algorithmic complexity is backed by powerful theorems, no theorem to the same effect is available in the case of information theory. Mathematicians in general do not discuss the relationship between the KSE and information (see for instance [1,4,15,18]). Others use the term 'information' but without elucidating how phase space topologies relate to the conceptual framework of information theory [5,6,9]. Yet others seem to be aware of the fact that there is a gap to bridge, but then do not provide more than heuristic remarks [2,8,11,13].

The aim of this paper is to bridge this gap. I first generalize Shannon's entropy and then prove that for Hamiltonian systems this generalized version of Shannon's entropy is equivalent to the KSE under plausible assumptions.

To this end, let me briefly recall some elements of dynamical systems theory. An abstract dynamical system is a triple $\mathscr{M} = (M, \mu, \Phi_t)$ where (M,μ) is a measure space equipped with a one-parameter group Φ_t of automorphisms of (M,μ), Φ_t depending measurably on t. The parameter t plays the role of time, which in what follows is assumed to evolve in discrete steps $(t_1 < t_2 < t_3 < \cdots)$. Furthermore I assume that M is normalized, $\mu(M) = 1$, and that the dynamics of the system is area preserving: $\mu(\Phi_t A) = \mu(A)$ for all measurable subsets $A \subseteq M$ and all times t.

I will use the following notational conventions: $\Phi_t(x)$ is the point in phase space onto which Φ_t maps the 'initial condition' x after time t has elapsed; and $\Phi_t(A)$ is the image of the subset $A \subseteq M$ under Φ_t. I write $\Phi_{t_i \to t_j}(A)$ to denote the image of A under the time development starting at t_i and ending at t_j.

It is often the case that the Φ_{t_i}, $i = 1,2, \ldots$, are generated by an iterative application of one single automorphism Φ. In this case we have $\Phi_{t_i} = \Phi^i$ and $\Phi_{t_i \to t_j}(A) = \Phi^{j-i}(A)$.

A partition $\alpha = \{\alpha_i | i = 1, \ldots, n\}$ of M is a division of M into finitely many measurable sets such that $\alpha_i \cap \alpha_j = \emptyset$ for all $i \neq j$ and $\mu(M - \cup_{i=1}^n \alpha_i) = 0$. The α_i are called 'atoms' or 'cells' of the partition. Furthermore notice that if $\alpha = \{\alpha_i | i = 1, \ldots, n\}$ is a partition of M then $\Phi_t \alpha := \{\Phi_t \alpha_i | i = 1, \ldots, n\}$ is a partition as well. Given two partitions $\alpha = \{\alpha_i | i = 1, \ldots, n\}$ and $\beta = \{\beta_j | j = 1, \ldots, m\}$, their least common refinement $\alpha \vee \beta$ is defined as follows: $\alpha \vee \beta = \{\alpha_i \cap \beta_j | i = 1, \ldots, n; j = 1, \ldots, m\}$.

Given this, the KSE H_Φ of an automorphism Φ is defined as

$$H_\Phi := \sup_\alpha \lim_{k \to \infty} (1/k) H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k-1}\alpha), \tag{1}$$

where the function on the right-hand side is the entropy of a partition, which is defined as follows: $H(\beta) := -\sum_{i=1}^m z[\mu(\beta_i)]$, $z(x) = x \log(x)$ if $x > 0$ and $z(x) = 0$ if $x = 0$; $\sup_\alpha$ is the supremum over all possible finite partitions α of the phase space [1].

2. Generalizing the Shannon entropy

Consider a source S, which at every discrete instant of time t_i selects one of the available messages $S_1, \ldots, S_n$ and sends it to a receiver R through a noiseless and deterministic channel. The production of a symbol by the source when time proceeds from t_i to t_{i+1} will be referred to as a 'step'. When taking the received messages down (on a paper strip, for instance), the receiver adds a time superscript to indicate when the message was received. Assuming that this process starts at time t_0 the receiver has produced the string $S_{l_1}^{t_1} S_{l_2}^{t_2} \cdots S_{l_k}^{t_k}$ at time t_k, where all the l_i range over $1, \ldots, n$ (i.e. the number of symbols available). Furthermore assume that the source is probabilistic and let $p(S_1), \ldots, p(S_n)$ be the respective probabilities that $S_1, \ldots, S_n$ occur (hence $p(S_1) + \cdots + p(S_n) = 1$). In this case, the Shannon entropy is defined as (see [17])

$$H_{\text{step}} := -\sum_{i=1}^n z[p(S_i)]. \tag{2}$$

H_{step} is a measure of the uncertainty about what symbol will crop up at the next step; the greater H_{step} the less certain we are about the outcome. I now generalize this expression in four stages.

First, we need to conditionalise on past events, as the probability that a particular symbol is chosen at time t may depend on what has been chosen beforehand. In the simplest case, a Markov process, the choice only depends on the preceding letter and not on the ones before that. In general, however, the choice can depend on the entire history of the process: the conditional probability of receiving S_i at time t_{k+1} is $p(S_i^{t_{k+1}} / S_{l_1}^{t_1} S_{l_2}^{t_2} \cdots S_{l_k}^{t_k})$. Since these probabilities may vary with k, the entropy may have a different value at every step. To make this explicit, I replace the subscript 'step' in Eq. (2) by 'k'. The expression for the entropy then is

$$H_k(S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k}) := -\sum_{i=1}^{n} z\Big[p(S_i^{t_{k+1}}/S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k})\Big] \tag{3}$$

and can be understood as a measure for the uncertainty about what symbol will show up at time t_{k+1} given that the previous history of the process (recorded on R's tape) is $S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k}$.

Second, as we usually do not know the entire history of a process it is convenient to consider the weighted average of all $H_k(S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k})$. As weight we choose the probability of the respective history, which does justice to the fact that not all histories are equally likely:

$$\overline{H}_k := \sum_{l_1,\ldots,l_k=1}^{n} p(S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k})H_k(S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k}), \tag{4}$$

where $p(S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k}) := p(S_{l_1}^{t_1})p(S_{l_2}^{t_2}/S_{l_1}^{t_1})\cdots p(S_{l_k}^{t_k}/S_{l_1}^{t_1}\cdots S_{l_{k-1}}^{t_{k-1}})$.

Third, on the basis of this we can now define the entropy $\tilde{H}_k$ of the entire process of the composition of a message of length k. Since no step is privileged over the others, this can be effected by simply taking the average of the entropy at every step of the process:

$$\tilde{H}_k := \frac{1}{k}\sum_{j=0}^{k-1}\overline{H}_j. \tag{5}$$

Fourth, we can say that the entropy of the source itself, H_S, is the average of the uncertainty at every step in the infinite limit:

$$H_S := \lim_{k\to\infty}\tilde{H}_k. \tag{6}$$

I will refer to this as 'communication-theoretic entropy' (CTE).

It is important to point out that the CTE can be used to characterize the random behaviour of a process. It is usually assumed that the probabilities are given and that the value of the CTE is calculated on the basis of these probabilities. However, this 'natural order' can be reversed. If the CTE is given, it can be used to characterize the probabilities involved even if they are not known independently. $H_S > 0$ expresses the fact that, on average, at every step there is some uncertainty about what the next symbol printed by the receiver will be. More precisely, whatever the past history of the system, we are not sure as to what the next symbol will be. And this characteristic persists forever, there exists no 'cut-off time' t_c in the process from which on the past history of the system allows us to predict with certainty what its future will be since H_S is greater than zero only if there do not cease to be $\bar{H}_k$ greater than zero. Now recall that $\bar{H}_k$ is a measure for the uncertainty about what the message printed at time $k+1$ will be. Hence, if there do not cease to be $\bar{H}_k > 0$ as time goes on, there will always be times at which we are not sure about what is going to happen next. As a consequence, we cannot predict with certainty what the future will be. (This does not imply that there are no instants of time for which this is possible; $H_S > 0$ is compatible with there being some particular $\bar{H}_k = 0$ from time to time.) In terms of probabilities this means that as the process goes on we never reach a stage where $p(S_i^{t_{k+1}}/S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k})$ equals one for some particular symbol and zero for all the others. In sum, we can characterize a system with positive entropy as one in which the past history never conveys certainty onto what will happen at the next step and more generally in the future.

3. The equivalence of KSE and CTE

The strategy now is to first carry over Eqs. (3)–(6) to the context of dynamical systems theory and then prove that the 'dynamical system version' of Eq. (6) is equivalent to the KSE as defined above.

The problem we face in doing so is that the messages we have been dealing with so far are discrete entities whereas the phase space of a dynamical system is continuous. This mismatch is removed by coarse graining the phase space. Let $\alpha = \{\alpha_1, \cdots, \alpha_n\}$ be a partition of the phase space and assume that the state of the system at t_0 is x. Then trace the trajectory $\Phi_{t_i}(x)$ of x and register at each time t_i, $i = 1,2,\ldots$ in what cell α_j, $j = 1,\ldots,n$, of the partition $\Phi_{t_i}(x)$ is. That is, write down $\alpha_j^{t_1}$ if $\Phi_{t_1}(x) \in \alpha_j$ at time t_1 and so on. If we do that up to time t_k this generates the string $\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}$, which is structurally identical to $S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k}$. Furthermore, we need something in $\mathscr{M}$ corresponding to the probability $p(S_i)$ of choosing a particular symbol S_i. By assumption, there is a normalised measure μ on M and it is a straightforward move to interpret this measure as a probability measure: interpret $\mu(\alpha_i)$ as the probability of finding the system's state in α_i.

Note, however, that although this move is quite natural, the interpretation of μ as the probability of finding the system's state in a particular cell is not compulsory. Not all measures reflect our ignorance about the system's real state; it could also simply be the spatial volume. However, this interpretation is perfectly possible, and that is all we need.

Then, the following associations are made to connect dynamical systems to communication theory:

(a) The atoms of the partition α_i correspond to the symbols (messages) S_i of the source.
(b) The measures of an atom $\mu(\alpha_i)$, interpreted as the probability of finding the system's state in cell α_i, correspond to the probability $p(S_i)$ of obtaining symbol S_i.
(c) The automorphism Φ_t corresponds to the source S, since they both do the job of generating the strings $\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}$ and $S_{l_1}^{t_1}S_{l_2}^{t_2}\cdots S_{l_k}^{t_k}$ respectively.

Given these associations, the Shannon entropy of α, commonly referred to as he 'entropy of the partition α', is $H(\alpha) := -\sum_{i=1}^{n} z[\mu(\alpha_i)]$.

In order to 'translate' Eqs. (3)–(6) into the 'language' of dynamical systems, we have to express the probabilities in terms of the measure μ. To this end realize that for any two instants of time t_i and t_j (where $t_i < t_j$) and any two measurable subsets A and B of M the following holds:

$$p(A^{t_i}B^{t_j}) = \mu\left[\Phi_{t_i\to t_j}(A)\cap B\right]. \tag{7}$$

The generalisation of this equality to any number of sets and instants of time is straightforward.

Now we spell out the conditional probabilities in terms of unconditional ones,

$$p(\alpha_i^{t_{k+1}}/\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}) = p(\alpha_i^{t_{k+1}}\&\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k})/p(\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}) \tag{8}$$

$$= p(\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}\alpha_i^{t_{k+1}})/p(\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}). \tag{9}$$

Apply Eq. (7) to this expression and then plug what we get into Eq. (3):

$$H_k(\alpha;\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}) := -\sum_{i=1}^{n} z\left[\frac{\mu\left(\alpha_i\cap\Phi_{t_k\to t_{k+1}}\alpha_{l_k}\cap\cdots\cap\Phi_{t_1\to t_{k+1}}\alpha_{l_1}\right)}{\mu\left(\alpha_{l_k}\cap\Phi_{t_{k-1}\to t_k}\alpha_{l_{k-1}}\cap\cdots\cap\Phi_{t_1\to t_k}\alpha_{l_1}\right)}\right]. \tag{10}$$

Given this, Eq. (4) carries over to dynamical systems easily,

$$\bar{H}_k(\alpha) := \sum_{l_1,\ldots,l_k=1}^{n} \mu\left(\alpha_{l_k}\cap\Phi_{t_{k-1}\to t_k}\alpha_{l_{k-1}}\cap\cdots\cap\Phi_{t_1\to t_k}\alpha_{l_1}\right) \sum_{i=1}^{n} z\left[\frac{\mu\left(\alpha_i\cap\Phi_{t_k\to t_{k+1}}\alpha_{l_k}\cap\cdots\cap\Phi_{t_1\to t_{k+1}}\alpha_{l_1}\right)}{\mu\left(\alpha_{l_k}\cap\Phi_{t_{k-1}\to t_k}\alpha_{l_{k-1}}\cap\cdots\cap\Phi_{t_1\to t_k}\alpha_{l_1}\right)}\right] \tag{11}$$

and similarly for the entropy of the process of the composition of a string of length k:

$$\tilde{H}_k(\alpha) := \frac{1}{k}\sum_{j=0}^{k-1}\bar{H}_j(\alpha). \tag{12}$$

Finally, on the basis of this we define the entropy of an automorphism with respect to partition α as follows:

$$H_{\Phi_t}(\alpha) := \lim_{k\to\infty}\tilde{H}_k(\alpha). \tag{13}$$

However, there is an important disanalogy between a source and a dynamical system. In the case of a source S, the set of possible messages $(S_1, \ldots, S_n)$ is a part of the definition of the source. This is not so with the partition α, which is no constitutive part of the dynamical system $\mathcal{M}$. Rather it has been 'imposed' on the system. This is a problem because the values we obtain for $H_{\Phi_t}(\alpha)$ essentially depend on the choice of the partition α. To get rid of this dependence, we define the *entropy of the automorphism* as the supremum of $H_{\Phi_t}(\alpha)$ over all finite measurable partitions:

$$H_{\Phi_t} = \sup_{\alpha} H_{\Phi_t}(\alpha). \tag{14}$$

Now we have to prove that this expression is equivalent to the KSE as defined in Eq. (1). To this end, first introduce an auxiliary device. Let α and β be two partitions; then the *conditional entropy of* α *with respect to* β is defined as

$$H(\alpha/\beta) := \sum_{j=1}^{m}\mu(\beta_j)\sum_{i=1}^{n} z\left[\frac{\mu(\alpha_i\cap\beta_j)}{\mu(\beta_j)}\right]. \tag{15}$$

Then realize that the standard definition of the KSE assumes that the flow is generated by the iterative application of the same automorphism Φ. Therefore, $\Phi_{t_i} = \Phi^i$ and $\Phi_{t_i \to t_j}(A) = \Phi^{j-i}(A)$. Given this, we have

Theorem 1

$$\tilde{H}_k(\alpha) = H(\alpha/\Phi\alpha \vee \Phi^2\alpha \vee \cdots \vee \Phi^k\alpha). \tag{16}$$

The prove of this theorem—along with the proof of the next theorem—will be given in the following section. Then the entropy of the process as given in Eq. (12) reads

$$\tilde{H}_k(\alpha) = \frac{1}{k}\left[H(\alpha) + H(\alpha/\Phi\alpha) + \cdots + H(\alpha/\Phi\alpha \vee \cdots \vee \Phi^{k-1}\alpha)\right]. \tag{17}$$

This can be considerably facilitated by using

Theorem 2

$$H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^k\alpha) = H(\alpha) + H(\alpha/\Phi\alpha) + \cdots + H(\alpha/\Phi\alpha \vee \cdots \vee \Phi^k\alpha). \tag{18}$$

Hence,

$$\tilde{H}_k(\alpha) = \frac{1}{k}H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k-1}\alpha). \tag{19}$$

Inserting this first into (13) and then (14) we obtain

$$H_\Phi = \sup_\alpha \lim_{k\to\infty} (1/k)H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k-1}\alpha) \tag{20}$$

and this is the definition of the entropy of an automorphism towards which we were aiming. Gathering the pieces together, we have proven the following:

Equivalence theorem

$$\begin{aligned} H_\Phi &= \sup_\alpha \lim_{k\to\infty} (1/k)H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k-1}\alpha) \\ &= \sup_\alpha \lim_{k\to\infty} \frac{-1}{k} \sum_{j=0}^{k-1} \sum_{l_1,\ldots,l_k=1}^{n} p(\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k}) \\ &\quad \sum_{i=1}^{n} z[p(\alpha_i^{t_{k+1}}/\alpha_{l_1}^{t_1}\alpha_{l_2}^{t_2}\cdots\alpha_{l_k}^{t_k})]. \end{aligned} \tag{21}$$

Since, by construction, the last term in this equation is equivalent to the CTE, the sought-after connection between the notion of entropy in dynamical systems theory and in information theory is established.

As a consequence, everything that has been said at the end of Section 2 about the unpredictable behaviour of a source can be carried over to dynamical systems one-to-one. However, a proviso with regard to the choice of a partition must be made. The exact analogue of the CTE is $H_\Phi(\alpha)$ and not H_Φ, which is defined as the supremum of $H_\Phi(\alpha)$ over all partitions α. For this reason, the characterization of randomness devised in the context of information theory strictly speaking applies to $H_\Phi(\alpha)$ rather than H_Φ. However, there is a close connection between the two: whenever $H_\Phi > 0$, there trivially is at least one partition for which $H_\Phi(\alpha) > 0$. In this case Φ_t is random in precisely the way described above with respect to this partition, and more generally with respect to all partitions for which $H_{\Phi_t}(\alpha) > 0$. Moreover, if the system is a K-system, the KSE is greater than zero for every partition which is not a trivial partition, that is a partition consisting of sets of measure 1 and zero only [4]. For this reason, statements about H_Φ and $H_\Phi(\alpha)$ naturally translate into one another.

4. Proofs of Theorems 1 and 2

In order to prove the two main theorems five lemmas are needed. The proof of Lemmas 1 and 3 can be found in [1], the other proofs are trivial.

Lemma 1. $H(\alpha \vee \beta) = H(\alpha) + H(\beta/\alpha)$.

Lemma 2. $H(\alpha) = H(\Phi_t\alpha)$.

Lemma 3. $\Phi(\alpha \vee \beta) = \Phi\alpha \vee \Phi\beta$.

Lemma 4. $H(\alpha \vee \beta) = H(\beta \vee \alpha)$.

Lemma 5. $\vee$ *is associative*: $\alpha \vee \beta \vee \gamma = (\alpha \vee \beta) \vee \gamma = \alpha \vee (\beta \vee \gamma)$.

Proof of Theorem 1. For the case of an automorphism generated by a mapping we have $\Phi_{t_i \to t_j}(A) = \Phi^{j-i}(A)$ (see above). Then, (11) becomes

$$\bar{H}_k(\alpha) = -\sum_{l_1,\ldots,l_k=1}^{n} \mu(\alpha_{l_k} \cap \cdots \cap \Phi^{k-1}\alpha_{l_1}) \sum_{i=1}^{n} z\left[\frac{\mu(\alpha_i \cap \Phi\alpha_{l_k} \cap \cdots \cap \Phi^{k}\alpha_{l_1})}{\mu(\alpha_{l_k} \cap \cdots \cap \Phi^{k-1}\alpha_{l_1})}\right]. \tag{22}$$

Using the fact that Φ is area preserving we get $\mu(\alpha_{l_k} \cap \cdots \cap \Phi^{k-1}\alpha_{l_1}) = \mu(\Phi\alpha_{l_k} \cap \cdots \cap \Phi^{k}\alpha_{l_1})$. Plugging this into Eq. (22) and taking the associativity of set intersection into account we obtain:

$$\bar{H}_k(\alpha) = -\sum_{l_1,\ldots,l_k=1}^{n} \mu(\Phi\alpha_{l_k} \cap \cdots \cap \Phi^{k}\alpha_{l_1}) \sum_{i=1}^{n} z\left[\frac{\mu(\alpha_i \cap \{\Phi\alpha_{l_k} \cap \cdots \cap \Phi^{k}\alpha_{l_1}\})}{\mu(\Phi\alpha_{l_k} \cap \cdots \cap \Phi^{k}\alpha_{l_1})}\right]. \tag{23}$$

Now realize that what the first sum effectively does is sum over all elements of a partition consisting of all intersections $\Phi\alpha_{l_k} \cap \cdots \cap \Phi^k\alpha_{l_1}$. This partition, however, is just $\Phi\alpha \vee \cdots \vee \Phi^k\alpha$. Furthermore, compare Eq. (23) with the definition of the conditional entropy in Eq. (15). We then obtain: $\bar{H}_k(\alpha) = H(\alpha/\Phi\alpha \vee \cdots \vee \Phi^k\alpha)$. □

Proof of Theorem 2. By weak induction on k:
Base case: $H(\alpha \vee \Phi\alpha) = H(\alpha) + H(\alpha/\Phi\alpha)$.

Proof: $H(\alpha \vee \Phi\alpha) = H(\Phi\alpha \vee \alpha)$, by Lemma 4, and $H(\Phi\alpha \vee \alpha) = H(\Phi\alpha) + H(\alpha/\Phi\alpha)$ by Lemma 1. Now use Lemma 2 and get $H(\alpha \vee \Phi\alpha) = H(\alpha) + H(\alpha/\Phi\alpha)$. □

Inductive step: $H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha) = H(\alpha) + \cdots + H(\alpha/\Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha)$.

Proof: Consider $H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha)$. With Lemmas 5 and 4 this is $H([\Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha] \vee \alpha)$, and now applying Lemma 1 yields $H(\Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha) + H(\alpha/[\Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha])$. Lemmas 2 and 3 together with the fact that Φ is measure preserving give: $H(\alpha \vee \cdots \vee \Phi^{k}\alpha) + H(\alpha/[\Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha])$. With the induction hypothesis this is $H(\alpha \vee \Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha) = H(\alpha) + \cdots + H(\alpha/[\Phi\alpha \vee \cdots \vee \Phi^{k+1}\alpha])$. □

References

[1] Arnold VI, Avez A. Ergodic problems of classical mechanics. New York: W.A. Benjamin; 1968.
[2] Billingsley P. Ergodic theory and information. New York: Wiley; 1965.
[3] Brudno AA. The complexity of the trajectory of a dynamical system. Russ Math Surveys 1978;33:197–8.
[4] Cornfeld IP, Fomin SV, Sinai YG. Ergodic theory. Berlin and Heidelberg: Springer; 1982.
[5] Eckmann J-P, Ruelle D. Ergodic theory of chaos and strange attractors. Review of Modern Physics 1985;57(1):617–56.
[6] Keller G. Equilibrium states in ergodic theory. Cambridge: 1998.
[7] Lichtenberg AJ, Liebermann MA. Regular and chaotic dynamics, 2nd ed. Berlin and New York: 1992.
[8] Mañé R. Ergodic theory and differentiable dynamics. Berlin and New York: 1983.
[9] Nadkarni MG. Basic Ergodic theory. Basel, 1998.
[10] Ott E. Chaos in dynamical systems. Cambridge: 1993.
[11] Parry W. Topics in ergodic theory. Cambridge: Cambridge UP; 1981.
[12] Pesin YB. Characteristic Liapunov exponents and smooth ergodic theory. Russ Math Surv 1977;32(4):55–114.
[13] Petersen K. Ergodic theory. London and New York: Cambridge UP; 1983.
[14] Reichl LE. The transition to chaos. In: Conservative classical systems: quantum manifestations. New York and Berlin: Springer; 1992.
[15] Rudolph DJ. Fundamentals of measurable dynamics. Ergodic theory on Lebesgue spaces. Oxford: Clarendon Press; 1990.
[16] Schuster HG. Deterministic chaos: an introduction. 2nd ed. Weinheim: Physik Verlag; 1988.
[17] Shannon CE, Weaver W. The mathematical theory of communication. Chicago and London: University of Illinois Press; 1949.
[18] Sinai YG, editor. Dynamical systems II. Ergodic theory with applications to dynamical systems and statistical mechanics. Berlin and New York: Springer; 1980.
[19] Tabor M. Chaos and integrability in nonlinear dynamics. New York: Wiley and Sons; 1989.