Illinois Journal of Mathematics
Volume 48, Number 1, Spring 2004, Pages 59–69
S 0019-2082

ENTROPY THEOREMS ALONG TIMES WHEN x VISITS A SET

TOMASZ DOWNAROWICZ AND BENJAMIN WEISS

Abstract. We consider an ergodic measure-preserving system in which we fix a measurable partition $\mathcal{A}$ and a set B of nontrivial measure. In a version of the Shannon-McMillan-Breiman Theorem, for almost every x, we estimate the rate of the exponential decay of the measure of the cell containing x of the partition obtained by observing the process only at the times n when $T^n x \in B$. Next, we estimate the rate of the exponential growth of the first return time of x to this cell. Then we apply these estimates to topological dynamics. We prove that a partition with zero measure boundaries can be modified to an open cover so that the S-M-B theorem still holds (up to ϵ) for this cover, and we derive the entropy function on invariant measures from the rate of the exponential growth of the first return time to the (n, ϵ)-ball around x.

1. Introduction and preliminaries

The two fundamental pointwise limit theorems in dynamics are the *ergodic theorem* and the *Shannon-McMillan-Breiman theorem*. For the former there has been much work devoted to the question of what happens when the sequence $f(T^n x)$ is replaced by a subsequence. In particular, Bourgain [B] has a striking result in which he shows that the Birkhoff ergodic theorem continues to remain valid when the n's are restricted to the visit times of some fixed set. It is rather natural to raise similar kinds of questions for the entropy limit theorems. In some sense, such questions are more delicate, and our results are correspondingly coarser in that we can only give bounds rather than precise convergence results. Indeed, examples show that in general the quantities that we study can fluctuate.

Let (X, μ, T) be an ergodic measure-preserving transformation of a probability space (we omit the indication of the sigma-field in this notation, as it remains fixed). Let $\mathcal{A}$ be a finite measurable partition of X. Let $K = \#\mathcal{A} \in \mathbb{N}$

Received January 22, 2003; received in final form May 18, 2003.
2000 *Mathematics Subject Classification.* 37A35.
The research of the first named author was supported by KBN Grant 2 P03 A 04622.

be the cardinality of $\mathcal{A}$. We denote by

$$\mathcal{A}^n := \bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$$

the partition into sets (often referred to as *cells*) of the form

$$A^n = \bigcap_{j=0}^{n-1} T^{-j}(A_j) \quad (A_j \in \mathcal{A}).$$

Every such cell will be identified with the corresponding *n-string*

$$(A_0 A_1 A_2 \dots A_{n-1})$$

over $\mathcal{A}$ viewed as a finite alphabet. By convention,

$$A^n(x) = (A(x)A(Tx)A(T^2x)\dots A(T^{n-1}x)),$$

where $A(y)$ denotes the unique cell of $\mathcal{A}$ containing y, is the unique n-string which contains x, and is called the *n-name over $\mathcal{A}$ of x*. (We often omit the part "over $\mathcal{A}$" if this is clear from the context.) We also denote by $R^n_{\mathcal{A}}(x)$ the first return time of x to $A^n(x)$:

$$R^n_{\mathcal{A}}(x) = \min\{i > 0 : T^i(x) \in A^n(x)\}.$$

The famous Shannon-McMillan-Breiman Theorem (see, e.g., [W]) states that for μ-almost every point x the measure of $A^n(x)$ decreases nearly exponentially in n with the exponent approaching the entropy of the process induced by the partition $\mathcal{A}$:

$$\lim_{n\to\infty} \frac{-1}{n} \log \mu(A^n(x)) = h_\mu(\mathcal{A}, T).$$

Another theorem, proved by Ornstein and Weiss [OW2], says that the first return times $R^n_{\mathcal{A}}(x)$ increase nearly exponentially in n with the exponent approaching the same entropy:

$$\lim_{n\to\infty} \frac{1}{n} \log R^n_{\mathcal{A}}(x) = h_\mu(\mathcal{A}, T).$$

Now we fix a set B with $\mu(B) \in (0,1)$ (usually B is large, of nearly full measure) and for each x we observe the process only at times j when $T^jx \in B$. This leads to a cell containing x defined as follows:

$$A^{*n}(x) = \bigcap_{j\in J^n(x)} T^{-j}(A(T^jx)),$$

where $J^n(x) = \{0 \le j \le n-1 : T^jx \in B\}$. Notice that the above set has the general form of an *enhanced string* $(A^*_0 A^*_1 A^*_2 \dots A^*_{n-1})$, where in addition to symbols from $\mathcal{A}$ we admit the symbol $*$ indicating that at the corresponding time we allow the point to be anywhere in the space X. For example, the string $(A_0 * A_2)$ represents the intersection $A_0 \cap T^{-2}(A_2)$. In $A^{*n}(x)$ the symbol $*$ occurs at coordinates j not belonging to $J^n(x)$. Different enhanced strings (of the same length) need not be disjoint, but this can only happen if the distributions of the symbol $*$ in both of them are different. Whenever a symbol belonging to $\mathcal{A}$ in one enhanced string is at the same position as a different symbol from $\mathcal{A}$ in the other, the corresponding intersections are obviously disjoint.

Also, we denote by

$$R^{*n}_{\mathcal{A}}(x) = \min\{i > 0 : T^i(x) \in A^{*n}(x)\},$$

the first return time of x to $A^{*n}(x)$.

Clearly, $A^{*n}(x) \supset A^n(x)$, so the measure of $A^{*n}(x)$ is not smaller, while the first return time to it is not larger, than the corresponding values for $A^n(x)$, and so behave the limit values appearing in the above two theorems:

$$\limsup_{n\to\infty} \frac{-1}{n} \log \mu(A^{*n}(x)) \le h_\mu(\mathcal{A}, T) \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n} \log R^{*n}_{\mathcal{A}}(x) \le h_\mu(\mathcal{A}, T).$$

In this work we provide the opposite limit inequalities, possibly modified by small correction terms. In fact, simple examples show that our estimates cannot be essentially improved without affecting their generality (see Example 1 below).

Notice that the sets $A^{*n}(x)$ neither form a partition nor do they appear naturally in the map induced on B (with the partition $\mathcal{A}$ restricted to B) because different points x' from the same cell $A^{*n}(x)$ usually generate different sets of times $J^n(x')$.

Later we will apply our estimates to settle the following question in topological dynamics: Can we modify a partition $\mathcal{A}$ to obtain an open cover $\mathcal{V}$ so that the measures of and the first return times to the cells $V^n(x)$ of $\mathcal{V}^n$ obey approximately the same exponential laws as in the S-M-B and O-W theorems. We will show how this can be done for a partition $\mathcal{A}$ into sets whose boundaries have measure zero. As a consequence we derive two topological ways of evaluating the entropy function on invariant measures by observing the (n, ϵ)-balls (one of which is already known from the work of Brin and Katok [BK]).

2. The main results

With the notation and setup as introduced above, we have the following theorems:

Theorem 1A (S-M-B Theorem along times when x visits a set). *If (X, μ, T) is an ergodic measure-preserving transformation, $\mathcal{A}$ is a finite partition of X, and B is a set with $\mu(B) \in (0, 1)$, then for μ-almost every x we have*

$$\liminf_{n\to\infty} \frac{-1}{n} \log \mu(A^{*n}(x)) \ge h_\mu(\mathcal{A}, T) - h_\mu(\mathcal{B}, T) - \mu(B^c) \log K,$$

where $\mathcal{B}$ is the partition of X into B and its complement B^c.

Theorem 1B (O-W Theorem along times when x visits a set). *If (X, μ, T) is an ergodic measure-preserving transformation, $\mathcal{A}$ is a finite partition of X, and B is a set with $\mu(B) \in (0,1)$, then for μ-almost every x we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log R^{*n}_{\mathcal{A}}(x) \geq h_\mu(\mathcal{A}, T) - h_\mu(\mathcal{B}, T) - \mu(B^c) \log K.$$

Example 1. To see that our estimates cannot be significantly improved, consider the independent product of two Bernoulli shifts, say

$$\mathcal{B}(\tfrac{1}{K}, \tfrac{1}{K}, \dots, \tfrac{1}{K}) \times \mathcal{B}(p, 1-p).$$

Let $\mathcal{A}$ be the partition into $2K$ sets corresponding to the pairs of symbols in both shifts, and let B be the set corresponding to the first symbol in the latter shift, so that $\mu(B) = p$. Then $h_\mu(\mathcal{A}, T) = \log K + H(p)$, where $H(p) = -p \log p - (1-p)\log(1-p)$ is the entropy $h_\mu(\mathcal{B}, T)$. Then for almost every point x it is easily calculated that $\mu(A^{*n}(x)) \approx (p/K)^{pn}$. Substituting this into our estimate of Theorem 1A, we get

$$p \log K - p \log p \geq (\log K + H(p)) - H(p) - (1-p) \log 2K.$$

This inequality differs from an equality by the term $(1-p)\log 2 - p \log p$. But for p near 1, $-p \log p$ is approximately $1-p$, so the additional term is nearly $(1-p)(1 + \log 2)$, much smaller than both of our correction terms $H(p)$ and $(1-p)\log 2K$. (The derivative of $H(p)$ is minus infinity at 1; K is assumed large in this example.) This shows that our estimate is in a sense optimal. The same example works for the estimate of Theorem 1B; we skip the details.

The proofs of both statements differ only in few details. The basic method is taken from [OW1] and [OW2]. After presenting the first proof, we will only sketch the second by indicating the necessary changes.

Before we proceed, we recall an interpretation of the S-M-B Theorem in the language of strings. The theorem says that for a typical x the measure of its m-name is approximately $e^{-m h_\mu(\mathcal{A},T)}$. This has two consequences concerning cardinalities of m-names:

(A) There is a collection of no more than $e^{m(h_\mu(\mathcal{A},T)+\delta)}$ m-names which fills up a set of nearly full measure (improving with m).

(B) Conversely, for any fixed set X' of positive measure, say $\mu(X') = \lambda$, the cardinality of m-names occurring in points $x \in X'$ exceeds, for large m, the number $\lambda e^{m(h_\mu(\mathcal{A},T)-\delta)}$, for any given $\delta > 0$.

We omit detailed proofs for these classic statements.

Proof of Theorem 1A. Clearly $A^{*n}(x) \subset A^{*n-1}(Tx)$. Thus the function

$$\Phi(x) = \liminf_{n\to\infty} \frac{-1}{n} \log \mu(A^{*n}(x))$$

is subinvariant. By preservation of the measure μ by T, Φ is μ-almost everywhere invariant, and hence, by ergodicity, constant. Denote that constant by C. In the remainder of the proof we need to appropriately estimate C from below.

Fix a $\delta > 0$ and let $N_1 \in \mathbb{N}$ be larger than $1/\delta$. There is an N_2 such that the set

$$G = \left\{ x : \frac{-1}{n} \log \mu(A^{*n}(x)) < C + \delta \text{ for some } n \in [N_1, N_2] \right\}$$

has measure strictly larger than $1 - \delta$. For each $x \in G$ we choose one value of n as above and denote it by $n(x)$. By ergodicity, for almost every x there exists m_x such that for all $m' \geq m_x$,

(a) $\operatorname{card}\{i \in [0, m'-1] : T^i x \in G\} > m'(1-\delta)$ and $\operatorname{card}\{i \in [0, m'-1] : T^i x \in B\} > m'(\mu(B) - \delta)$.

For $m \in \mathbb{N}$ denote by X_m the set of points x for which $m_x \leq m$. Clearly, X_m grows with m to a full measure set. Moreover, by the interpretation (A) of the S-M-B theorem applied to the partition $\mathcal{B}$, there exist sets Y_m with $\mu(Y_m)$ converging to 1, such that:

(b) The number of m-names over $\mathcal{B}$ of points x belonging to Y_m is less than $e^{m(h_\mu(\mathcal{B},T)+\delta)}$.

We now modify the sets Y_m so that they grow (with respect to inclusion) along a certain subsequence of indices m. Namely, passing to a subsequence we may assume that the measures of the complements Y_m^c form a convergent series. Then

$$Y'_m = \bigcap_{m' \geq m} Y_{m'}$$

provides the desired modification. From now on Y_m will denote members of this modified subsequence.

Let m_0 be such that $X' = X_{m_0} \cap Y_{m_0}$ has positive measure $\lambda > 0$. Then fix $m \geq m_0$ so large that the consequence (B) of the S-M-B Theorem holds. Clearly, $X' \subset Y_m$. Hence (b) holds with X' in place of Y_m.

Consider an element $x \in X'$. Let i_1 be the first index i such that $T^i x \in G$ and set $n_1 = n(T^{i_1}x)$ (where $n(\cdot)$ is defined as above). Let i_2 be the smallest index $i \geq i_1 + n_1$ with $T^i x \in G$ and set $n_2 = n(T^{i_2}x)$, and so on. The intervals $[i_l, i_l + n_l - 1]$ will be called *good intervals*. They contain all indices i for which $T^i x \in G$. Hence, by (a), they cover at least $m(1-\delta)$ elements of the interval $[0, m-1]$.

We will now estimate from above the cardinality of the set of possible m-names (over $\mathcal{A}$) for points $x \in X'$. In each step below we formally partition X' (or a class created in the previous step) into subclasses according to certain variable parameters and we count the number of such subclasses. The total number is then estimated by the product of the counts for each step.

(1) We first partition X' into classes according to the m-names with respect to the partition $\mathcal{B}$. By (b), there are no more than $e^{m(h_\mu(\mathcal{B},T)+\delta)}$ such classes. We further restrict our counting to one such class, so henceforth all points x yield the same set $J^m(x)$ $(=\{i \in [0,m-1] : T^i x \in B\})$.

(2) By the second part of (a), there are at most $m(\mu(B^c)+\delta)$ indices i in $[0,m-1]$ outside $J^m(x)$. We partition our class according to the configurations of symbols in the m-names over $\mathcal{A}$ occurring at the indices outside $J^m(x)$. There are at most $K^{m(\mu(B^c)+\delta)}$ different such configurations.

(3) We now compute the number of ways the good intervals can be positioned along $[0,m-1]$. Since consecutive indices i_l are separated by at least N_1 positions (each n_l is $\geq N_1$), there are at most $m\delta$ of them in $[0,m-1]$. Thus, by a standard estimate involving Stirling's formula, the number of possible ways of positioning these indices along $[0,m-1]$ is smaller than $e^{m(H(\delta)+\delta)}$ (if m is sufficiently large), where, as before,

$$H(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta).$$

With the positions of the indices i_l fixed, there are no more than $m\delta$ numbers in $[0,m-1]$ not belonging to the good intervals, which can happen again in at most $e^{m(H(\delta)+\delta)}$ different ways. Every such way is equivalent to choosing the numbers n_l, except perhaps for the last one (if $i_l+n_l > m$), for which we have no more than N_2 possibilities. Altogether, there are no more than $N_2 e^{2m(H(\delta)+\delta)}$ classes in X' determined by the positioning of the good intervals contained in or intersecting $[0,m-1]$.

(4) With the choice of the sequence of good intervals fixed, there are at most $K^{m\delta}$ configurations of symbols from $\mathcal{A}$ at the (not yet established) coordinates i outside good intervals.

(5) It remains to count the number of possible m-names of points from X' within a class with a fixed (common) set $J^m(x)$, a fixed positioning of good intervals, with a fixed pattern of symbols at all positions outside good intervals and at all positions outside $J^m(x)$. Notice that the positions which are yet to be filled coincide with the union of the sets $J^{n_l}(T^{i_l}x)$ (which does not depend on x within the considered class) over the indices l enumerating the good intervals. For each such index l, establishing the values along $J^{n_l}(T^{i_l}x)$ is equivalent to establishing the enhanced n_l-string $A^{*n_l}(T^{i_l}x)$. Because the distribution of the symbol $*$ in such a string is common for all points x in the considered class, different enhanced strings $A^{*n_l}(T^{i_l}x)$ represent disjoint sets. On the other hand, since $T^{i_l}x \in G$ and $n_l = n(T^{i_l}x)$, all such enhanced strings are associated with sets $A^{*n_l}(T^{i_l}x)$ of measure larger than $e^{-n_l(C+\delta)}$. By disjointness, there are thus at most $e^{n_l(C+\delta)}$ such choices for each index l. Since the sum of the lengths n_l of good intervals does not exceed $m+N_2$, the total number of choices in this step is not larger than $e^{(m+N_2)(C+\delta)}$.

Multiplying the estimates from steps (1)–(5), we see that the total number of m-names over $\mathcal{A}$ appearing in the elements of X' does not exceed

$$e^{m(h_\mu(\mathcal{B},T)+\delta)} \cdot K^{m(\mu(B^c)+\delta)} \cdot N_2 e^{2m(H(\delta)+\delta)} \cdot K^{m\delta} \cdot e^{(m+N_2)(C+\delta)}.$$

Combining this with the lower estimate (B), taking logarithms, dividing by m and letting m grow to infinity, we obtain

$$h_\mu(\mathcal{A},T)-\delta \le h_\mu(\mathcal{B},T)+\delta+(\mu(B^c)+\delta)\log K+2(H(\delta)+\delta)+\delta\log K+C+\delta,$$

which holds for δ arbitrarily small. Passing to the limit as $\delta \to 0$ and rearranging terms, we conclude that

$$C \ge h_\mu(\mathcal{A},T) - h_\mu(\mathcal{B},T) - \mu(B^c)\log K. \qquad \square$$

Proof of Theorem 1B. We begin by noting that, because of the inequality

$$R^{*n-1}_{\mathcal{A}}(Tx) \le R^{*n}_{\mathcal{A}}(x),$$

the function $\Psi(x) = \liminf_n \frac{1}{n}\log R^{*n}_{\mathcal{A}}(x)$ is subinvariant, and hence μ-almost everywhere equal to some constant C. From here on we proceed as in the proof of Theorem 1A with the following few modifications:

- The set G is different:

  $$G = \left\{x : \frac{1}{n}\log R^{*n}_{\mathcal{A}}(x) < C+\delta \text{ for some } n = n(x) \in [N_1, N_2]\right\}.$$

  However, it still has measure greater than $1-\delta$ for an appropriate N_2.
- When defining i_l and n_l, we also define $R_l = R^{*n_l}_{\mathcal{A}}(T^{i_l}x)$.
- We replace the estimate (5) by the following argument:

(5') With a fixed choice of the sequence of good intervals, let l' be the largest l such that $i_l + n_l + R_l < m$. Good intervals to the right of $i_{l'} + n_{l'} - 1$ can occur only to the right of $m - e^{N_2(C+\delta)} - N_2$ (because $e^{N_2(C+\delta)}$ estimates all possible values of R_l and N_2 estimates all possible values of n_l). Thus there are at most $e^{N_2(C+\delta)} + N_2$ positions contained in the good intervals with indices larger than l'. Hence there are at most $K^{e^{N_2(C+\delta)}+N_2}$ ways to fill these positions. We now choose and fix one such way.

(6) There are at most $e^{n_{l'}(C+\delta)}$ possible values of $R_{l'}$. Once the value of $R_{l'}$ is specified, the block over the good interval $[i_{l'}, i_{l'}+n_{l'}-1]$ is completely determined, because, by the definition of $R^{*n}_{\mathcal{A}}(x)$, the values at its positions i where $T^i x \in B$ (the remaining positions have been fixed in step (2)) are repeated in the already established section at a specified distance. (This works even if $R_{l'} < n_{l'}$, i.e., when the block overlaps with its repetition; in this case we reconstruct the block from right to left, using periodicity.) Again, we choose the value of $R_{l'-1}$, for which we have $e^{n_{l'-1}(C+\delta)}$ options. This determines the block over $[i_{l'-1}, i_{l'-1}+n_{l'-1}-1]$. We continue in this

manner until the first good interval is filled. The total number of choices in this step does not exceed

$$e^{\sum_{l=1}^{l'} n_l(C+\delta)} \leq e^{m(C+\delta)}.$$

Multiplying the estimates from steps (1)–(6), the total number of blocks over $[0, m-1]$ appearing in the elements of X' does not exceed

$$e^{m(h_\mu(\mathcal{B},T)+\delta)} \cdot K^{m(\mu(B^c)+\delta)} \cdot N_2 e^{2m(H(\delta)+\delta)} \cdot K^{m\delta} \cdot K^{e^{N_2(C+\delta)}+N_2} \cdot e^{m(C+\delta)}.$$

The final step of the proof requires no modifications. □

Remark. It is not very hard to see that the term $\log K$ in the estimates of Theorems 1A and 1B can be replaced by the entropy $h_{\mu_{B^c}(\mathcal{A},T_{B^c})}$ of the transformation induced on the complement of B with respect to the (suitably restricted) partition $\mathcal{A}$. We skip the details of this improvement.

3. Applications to topological dynamics

Suppose μ is a regular Borel probability measure that is invariant and ergodic for an action of a continuous map T on a compact Hausdorff space X. For a finite open cover $\mathcal{V}$ of X, the cover $\mathcal{V}^n$ is defined by exactly the same formula as for partitions. We denote by $V^n(x)$ the union of all elements of the cover $\mathcal{V}^n$ containing x, and by $R^n_{\mathcal{V}}(x)$ the first return time of x to $V^n(x)$.

Definition 1. Given a partition $\mathcal{A}$ of cardinality K, we say that an open cover $\mathcal{V}$ of X of cardinality $K+1$ is *ϵ-inscribed in $\mathcal{A}$* if it consists of K sets, each contained in a different element of $\mathcal{A}$, and an open set V_0 with $\mu(V_0) \leq \epsilon$.

Note that if $\mathcal{A}$ consists of sets whose boundaries are of measure zero, then, by regularity, there exist covers that are ϵ-inscribed in $\mathcal{A}$ for every $\epsilon > 0$. For instance, $\mathcal{V}$ may consist of the interiors of the sets $A \in \mathcal{A}$ and an open set V_0 of small measure containing the union of all boundaries of these sets. It is well known that in metric spaces such "zero measure boundary" partitions into sets of arbitrarily small diameter exist for any Borel probability measure.

Theorem 2. *Let μ be an ergodic measure in a topological dynamical system (X,T) and let $\mathcal{V}$ be a cover that is ϵ-inscribed in a finite Borel partition $\mathcal{A}$. Then for μ-almost every x we have*

$$\liminf_{n\to\infty} \frac{-1}{n} \log \mu(V^n(x)) \geq h_\mu(\mathcal{A},T) - H(\epsilon) - \epsilon \log K$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log R^n_{\mathcal{V}}(x) \geq h_\mu(\mathcal{A},T) - H(\epsilon) - \epsilon \log K.$$

Proof. Set $B = X \setminus V_0$ and notice that

$$V^n(x) \subset \bigcap_{j \in J^n(x)} T^{-j}(A(T^j x)) = A^{*n}(x).$$

This can be seen by considering in the intersection the entire space X at times j when $T^j x$ falls into V_0, i.e., when it belongs to more than one element of $\mathcal{V}$. At other times j such an element of $\mathcal{V}$ is unique and is contained in $A(T^j x)$. The assertions now follows directly from Theorems 1A and 1B. □

In our next application we assume that X is a metric space with metric d, and, in addition to the sets of the form $V^n(x)$, we consider the (n, ϵ)-balls around x defined by

$$\mathbf{B}^{(n,\epsilon)}(x) = \{y : d(T^j x, T^j y) < \epsilon \text{ for each } 0 \le j < n\}.$$

Also, we let $R^n_\epsilon(x)$ denote the first return time of x to its (n, ϵ)-ball:

$$R^n_\epsilon(x) = \min\{i > 0 : T^i(x) \in \mathbf{B}^{(n,\epsilon)}(x)\}.$$

The investigations of this paper lead us to consider the following (topological) entropy functions defined on invariant measures:

Definition 2. Fix an open cover $\mathcal{V}$ of X and an $\epsilon > 0$.

$$h^{BK}(\mu, \mathcal{V}) = \int \limsup_{n\to\infty} \frac{-1}{n} \log \mu(V^n(x)) d\mu(x),$$

$$h^{BK}(\mu, \epsilon) = \int \limsup_{n\to\infty} \frac{-1}{n} \log \mu(\mathbf{B}^{(n,\epsilon)}(x)) d\mu(x),$$

$$h^{OW}(\mu, \mathcal{V}) = \int \limsup_{n\to\infty} \frac{1}{n} \log R^n_{\mathcal{V}}(x) d\mu(x),$$

$$h^{OW}(\mu, \epsilon) = \int \limsup_{n\to\infty} \frac{1}{n} \log R^n_\epsilon(x) d\mu(x).$$

Remarks. (A) All four integrals above are easily seen to be of invariant functions (by an argument similar to that given at the beginning of the proofs of Theorems 1A and 1B). Hence, for μ ergodic, they are constant almost everywhere, so that the integration in the ergodic case can be omitted. Obviously, due to the integration, the same entropy notions are obtained for non-ergodic invariant measures by averaging with respect to the ergodic decomposition.

(B) The label BK stands for Brin-Katok, who first considered the epsilon version of this entropy (see [BK]) and proved the corresponding convergence as in Theorem 3 below. For this notion, our methods allow us to derive the same result in a different way. The label OW stands for Ornstein and Weiss, who derived the formula for entropy based on the first return time for partitions (see [OW2]). We do not know of any previous investigations of the notion of entropy based on the first return time for covers or balls.

(C) We refer the reader to [D], where these and similar notions are investigated as so-called *entropy structures.*

(D) Replacing lim sup by lim inf can lead to slightly different notions, but, as is easily verified, the convergence result of Theorem 3 below still holds.

THEOREM 3. *Let $(\mathcal{V}_k)_{k\in\mathbb{N}}$ be a sequence of finite open covers with maximal diameters of elements decreasing to zero. The following convergence results hold for every invariant measure μ:*

$$\lim_k h^{BK}(\mu,\mathcal{V}_k) = \lim_k h^{OW}(\mu,\mathcal{V}_k) = \lim_{\epsilon\to 0} h^{BK}(\mu,\epsilon) = \lim_{\epsilon\to 0} h^{OW}(\mu,\epsilon) = h_\mu(T).$$

Proof. It suffices to prove the convergences for the case when μ is ergodic. In this case we can find a sequence of partitions $\mathcal{A}_{\mu,k}$ into sets of maximal diameter ϵ_k decreasing to 0, and with boundaries of μ-measure zero. Note that then $h_\mu(\mathcal{A}_{\mu,k},T) \underset{k}{\to} h_\mu(T)$. For each k we find a cover $\mathcal{V}_{\mu,k}$ which is $\epsilon_k/\log\#\mathcal{A}_{\mu,k}$-inscribed in $\mathcal{A}_{\mu,k}$. By a direct application of Theorem 2 we have

$$\liminf_{k\to\infty} h^{BK}(\mu,\mathcal{V}_{\mu,k}) \geq h_\mu(T), \quad \text{and} \quad \liminf_{k\to\infty} h^{OW}(\mu,\mathcal{V}_{\mu,k}) \geq h_\mu(T).$$

Further, for each k_0, ϵ_k becomes eventually smaller than the Lebesgue number of $\mathcal{V}_{\mu,k_0}$; in particular, the covers $\mathcal{V}_k$ are eventually inscribed in $\mathcal{V}_{\mu,k_0}$. So, for each n, $\mathbf{B}^{(n,\epsilon_k)}(x) \subset V^n_{\mu,k_0}(x)$ and $V^n_k(x) \subset V^n_{\mu,k_0}(x)$. This easily implies that all four lower limits in question are not smaller than the entropy. Conversely, the partitions $\mathcal{A}_{\mu,k}$ satisfy $A^n_{\mu,k}(x) \subset \mathbf{B}^{(n,\epsilon_k)}(x)$ and, for each k_0, we eventually have $A^n_{\mu,k}(x) \subset V^n_{k_0}(x)$, which implies that the four upper limits do not exceed the entropy. □

The following natural question arises: How far is the convergence of Theorem 3 from being uniform on invariant measures? The answer to this question is provided in [D], where it is shown that uniform convergence of any of these four sequences (as well as of several other entropy functions with topological flavor) is equivalent to the system (X,T) being asymptotically h-expansive (which is a relatively strong topological restriction; see [M] for the definition).

References

[B] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, with an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein, Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45. MR **90k:**28030

[BK] M. Brin and A. Katok, *On local entropy*, Geometric dynamics (Rio de Janeiro, 1981), Lecture Notes in Math., vol. 1007, Springer, Berlin, 1983, pp. 30–38. MR **85c:**58063

[D] T. Downarowicz, *Entropy structure*, preprint.

[M] M. Misiurewicz, *Topological conditional entropy*, Studia Math. **55** (1976), 175–200. MR 54#3672

[OW1] D. S. Ornstein and B. Weiss, *The Shannon-Mcmillan-Breiman theorem for a class of amenable groups*, Israel J. Math. **44** (1983), 53–60. MR **85f:**28018

[OW2] ______, *Entropy and data compression schemes*, IEEE Trans. Inform. Theory **39** (1993), 78–83. MR **93m:**94012

[W] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982. MR **84e:**28017

Tomasz Downarowicz, Institute of Mathematics, Wroclaw University of Technology, Wybrzee Wyspiańskiego 27, 50-370 Wrocław, Poland
E-mail address: downar@im.pwr.wroc.pl

Benjamin Weiss, The Hebrew University of Jerusalem, Jerusalem, 91904, Israel
E-mail address: weiss@math.huji.ac.il