Entropy in Dynamical Systems

Lai-Sang Young[1]

In this article, the word *entropy* is used exclusively to refer to the entropy of a dynamical system, i.e. a map or a flow. It measures the *rate of increase in dynamical complexity* as the system evolves with *time*. This is not to be confused with other notions of entropy connected with spatial complexity.

I will attempt to give a brief survey of the role of entropy in dynamical systems and especially in smooth ergodic theory. The topics are chosen to give a flavor of this invariant; they are also somewhat biased toward my own interests. This article is aimed at nonexperts. After a review of some basic definitions and ideas, I will focus on one main topic, namely the relation of entropy to Lyapunov exponents and dimension, which I will discuss in some depth. Several other interpretations of entropy, including its relations to volume growth, periodic orbits and horseshoes, large deviations and rates of escape are treated briefly.

1 Topological and metric entropies: a review

We begin with topological entropy because it is simpler, even though chronologically metric entropy was defined first. Topological entropy was first introduced in 1965 by Adler, Konheim and McAndrew [AKM]. The following definition is due to Bowen and Dinaburg. A good reference for this section is [W], which contains many of the early references.

Definition 1.1 *Let $f : X \to X$ be a continuous map of a compact metric space X. For $\varepsilon > 0$ and $n \in \mathbb{Z}^+$, we say $E \subset X$ is an (n, ε)-separated set if for every $x, y \in E$, there exists i, $0 \leq i < n$, such that $d(f^i x, f^i y) > \varepsilon$. Then the topological entropy of f, denoted by $h_{\rm top}(f)$, is defined to be*

$$h_{\rm top}(f) \;=\; \lim_{\varepsilon \to 0} \left\{ \limsup_{n \to \infty} \frac{1}{n} \log N(n, \varepsilon) \right\}$$

where $N(n, \varepsilon)$ is the maximum cardinality of all (n, ε)-separated sets.

[1]Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012, email lsy@cims.nyu.edu. This research is partially supported by a grant from the NSF

Roughly speaking, we would like to equate "higher entropy" with "more orbits", but since the number of orbits is usually infinite, we need to fix a finite "resolution", i.e. a scale below which we are unable to tell points apart. Suppose we do not distinguish between points that are $< \varepsilon$ apart. Then $N(n, \varepsilon)$ represents the number of distinguishable orbits of length n, and if this number grows like $\sim e^{nh}$, then h is the topological entropy. Another way of counting the number of distinguishable n-orbits is to use (n, ε)-*spanning* sets, i.e. sets E with the property that for every $x \in X$, there exists $y \in E$ such that $d(f^i x, f^i y) < \varepsilon$ for all $i < n$; $N(n, \varepsilon)$ is then taken to be the minimum cardinality of these sets. The original definition in [AKM] uses open covers but it conveys essentially the same idea.

Facts *If follows immediately from Definition 1.1 that*

(i) $h_{\rm top}(f) \in [0, \infty]$*, and*

(ii) if f *is a differentiable map of a compact* d*-dimensional manifold, then* $h_{\rm top}(f) \leq d \log \|Df\|$.

Metric or *measure-theoretic* entropy for a transformation was introduced by Kolomogorov and Sinai in 1959; the ideas go back to Shannon's information theory. We begin with some notation. Let $(X, \mathcal{B}, \mu)$ be a probability space, i.e. X is a set, $\mathcal{B}$ a σ-algebra of subsets of X, and μ a probability measure on $(X, \mathcal{B})$. Let $f : X \to X$ be a measure-preserving transformation, i.e. for all $A \in \mathcal{B}$, we have $f^{-1}A \in \mathcal{B}$ and $\mu(A) = \mu(f^{-1}A)$. Let $\alpha = \{A_1, \cdots, A_k\}$ be a finite partition of X. Then $f^{-i}\alpha = \{f^{-i}A_1, \cdots f^{-i}A_k\}$, and we say j is the α-*address* of $x \in X$ if $x \in A_j$. For two partitions α and β, $\alpha \vee \beta := \{A \cap B : A \in \alpha, B \in \beta\}$, so that elements of $\bigvee_{i=0}^{n-1} f^{-i}\alpha$ are sets of the form $\{x : x \in A_{i_0},\ fx \in A_{i_1},\ \cdots,\ f^{n-1}x \in A_{i_{n-1}}\}$ for some $(i_0, i_1, \cdots, i_{n-1})$, which we call the α-addresses of the n-orbit starting at x.

Definition 1.2 *The metric entropy of* f*, written* $h_\mu(f)$*, is defined as follows:*

$$H(\alpha) := H(\mu(A_1), \cdots, \mu(A_k)) \quad \text{where} \quad H(p_1, \cdots, p_k) = -\sum p_i \log p_i \ ;$$

$$h_\mu(f, \alpha) \ := \ \lim_{n\to\infty} \frac{1}{n} H(\bigvee_0^{n-1} f^{-i}\alpha) \ = \ \lim_{n\to\infty} H(f^{-n}\alpha | \bigvee_0^{n-1} f^{-i}\alpha) \ ;$$

$$h_\mu(f) \ := \ \sup_\alpha h_\mu(f, \alpha) \ .$$

The objects in the definition above have the following interpretations: $H(\alpha)$ measures the amount of uncertainty, on average, as one attempts to predict the α-address of a randomly chosen point. The two limits in the second line can be shown to be equal. The quantity in the middle is the average uncertainty *per iteration* in guessing the α-addresses of a typical n-orbit, while the one on the right is the uncertainly in guessing the α-address of $f^n x$ conditioned on the knowledge of the α-addresses of $x, fx, \cdots, f^{n-1}x$. The following facts make it easier to compute the quantity in the third line: $h_\mu(f) = h_\mu(f, \alpha)$ if α is a generator; it can also be realized as $\lim h_\mu(f, \alpha_n)$ where α_n is an increasingly refined sequence of partitions such that $\bigvee \alpha_n$ partitions X into points.

The following theorem gives a very nice interpretation of metric entropy:

Theorem 1 (The Shannon-McMillan-Breiman Theorem) *Let f and α be as above, and write $h = h_\mu(f, \alpha)$. Assume for simplicity that (f, μ) is ergodic. Then given $\varepsilon > 0$, there exists N such that the following holds for all $n \geq N$:*

(i) $\exists Y_n \subset X$ *with* $\mu(Y_n) > 1 - \varepsilon$ *such that* Y_n *is the union of* $\sim e^{(h\pm\varepsilon)n}$ *atoms of* $\bigvee_0^{n-1} f^{-i}\alpha$ *each having* μ*-measure* $\sim e^{-(h\pm\varepsilon)n}$*;*

(ii) $-\frac{1}{n} \log \mu((\bigvee_0^{n-1} f^{-i}\alpha)(x)) \to h$ *a.e. and in* L^1 *where* $\alpha(x)$ *is the element of* α *containing* x*.*

SUMMARY Both h_μ and h_{top} measure the exponential rates of growth of n-orbits:
– h_μ counts the number of **typical** n-orbits, while
– h_{top} counts **all** distinguishable n-orbits.

We illustrate these ideas with the following coin-tossing example. Consider $\sigma : \Pi_0^\infty\{H, T\} \to \Pi_0^\infty\{H, T\}$, where each element of $\Pi_0^\infty\{H, T\}$ represents the outcome of an infinite series of trials and σ is the shift operator. Since the total number of possible outcomes in n trials is 2^n, $h_{\text{top}}(\sigma) = \log 2$. If $P(H) = p,\ P(T) = 1 - p$, then $h_\mu(\sigma) = -p \log p - (1 - p) \log(1 - p)$. In particular, if the coin is biased, then the number of *typical* outcomes in n trials is $\sim e^{nh}$ for some $h < \log 2$.

From the discussion above, it is clear that $h_\mu \leq h_{\text{top}}$. We, in fact, have the following *variational principle*:

Theorem 2 *Let f be a continous map of a compact metric space. Then*

$$h_{\text{top}}(f) \ = \ \sup_\mu \ h_\mu(f)$$

where the supremum is taken over all f-invariant Borel probability measures μ.

The idea of expressing entropy in terms of local information has already appeared in Theorem 1 (ii). Here is another version of this idea, more convenient for certain purposes. In the context of continuous maps of compact metric spaces, for $x \in X$, $n \in \mathbb{Z}^+$ and $\varepsilon > 0$, define

$$B(x, n, \varepsilon) \ := \ \{y \in X : d(f^i x, f^i y) < \varepsilon,\ 0 \leq i < n\} \ .$$

Theorem 3 [BK] *Assume (f, μ) is ergodic. Then for μ-a.e. x,*

$$h_\mu(f) \ = \ \lim_{\varepsilon \to 0} \left\{ \limsup_{n\to\infty} -\frac{1}{n} \log \mu B(x, n, \varepsilon) \right\} \ .$$

Thus metric entropy also has the interpretation of being the rate of loss of information on nearby orbits. This leads naturally to its relation to Lyapunov exponents, which is the topic of the next section.

2 Entropy, Lyapunov Exponents and Dimension

In this section, we focus on a set of ideas in which entropy plays a central role.

Setting for this section: We consider (f, μ) where $f : M \to M$ is a C^2 diffeomorphism of a compact Riemannian manifold M and μ is an f-invariant Borel probability measure. For simpllicity, we assume (f, μ) is ergodic. (This assumption is not needed but without it the results below are a little more cumbersome to state.) Let $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ denote the distinct Lyapunov exponents of (f, μ), and let E_i be the linear subspaces corresponding to λ_i, so that the dimension of E_i is the multiplicity of λ_i.

We have before us two ways of measuring dynamical complexity: metric entropy, which measures the growth in randomness or number of "typical" orbits, and Lyapunov exponents, which measure the rates at which nearby orbits diverge. The first is a purely probabilistic concept, while the second is primarily geometric. We now compare these two invariants. Write $a^+ = \max(a, 0)$.

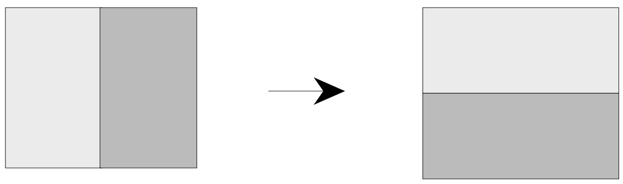
Theorem 4 (Pesin's Formula) [P] *If μ is equivalent to the Riemannian measure on M, then*

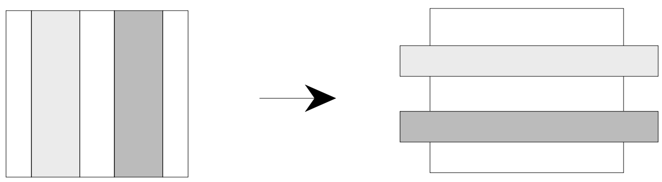
$$h_\mu(f) \ = \ \sum \lambda_i^+ \dim E_i \ .$$

Theorem 5 (Ruelle's Inequality) [R2] *Without any assumptions on μ, we have*

$$h_\mu(f) \ \leq \ \sum \lambda_i^+ \dim E_i \ .$$

To illustrate what is going on, consider the following two examples: Assume for simplicity that both maps are affine on each of the shaded rectangles and their images are as shown. The first map is called "baker's transformation". Lebesgue measure is preserved, and $h_\mu(f) = \log 2 = \lambda_1$, the positive Lyapunov exponent. The second map





is easily extended to Smale's horseshoe; we are interested only in points that remain in the shaded vertical strips in all (forward and backward) times. As with the first map, we let μ be the Bernoulli measure that assigns mass $\frac{1}{2}$ to each of the two vertical strips. Here $h_\mu(f) = \log 2 < \lambda_1$.

Theorems 4 and 5 suggest that entropy is created by the exponential divergence of nearby orbits. In a conservative system, i.e. in the setting of Theorem 4, all the expansion goes back into the system to make entropy, leading to the equality in Pesin's entropy formula. A strict inequality occurs when some of the expansion is "wasted", and that can happen only if there is "leakage" or dissipation from the system.

The reader may have noticed that the results above involve only positive Lyapunov exponents. Indeed there is the following complete characterization:

Theorem 6 ([LS], [L], [LY1]) *For (f, μ) with $\lambda_1 > 0$, Pesin's entropy formula holds if and only if μ is an SRB measure.*

Definition 2.1 *An f-invariant Borel probability measure μ is called a* **Sinai-Ruelle-Bowen measure** *or* **SRB measure** *if f has a positive Lyapunov exponent μ-a.e. and μ has absolutely continuous conditional measures on unstable manifolds.*

The last sentence is an abbreviated way of saying the following: If f has a positive Lyapunov exponent μ-a.e., then local unstable manifolds are defined μ-a.e. Each one of these submanifolds inherits from M a Riemannian structure, which in turn induces on it a Riemannian measure. We call μ an SRB measure if the conditional measures of μ on local unstable manifolds are absolutely continuous with respect to these Riemannian measures.

SRB measures are important for the following reasons: (i) They are more general than invariant measures that are equivalent to Lebesgue (they are allowed to be singular in stable directions); (ii) they can live on attractors (which cannot support invariant measures equivalent to Lebesgue); and (iii) singular as they may be, they reflect the properties of positive Lebesgue measure sets in the sense below:

Fact [PS] If μ is an ergodic SRB measure with no zero Lyapunov exponents, then there is a positive Lebesgue measure set V with the property that for every continuous function $\varphi : M \to \mathbb{R}$,

$$\frac{1}{n}\sum_{i=0}^{n-1} \varphi(f^i x) \ \to \ \int \varphi d\mu$$

for Lebesgue-a.e. $x \in V$.

Theorems 4, 5 and 6 and Definition 2.1 are generalizations of ideas first worked out for Anosov and Axiom A systems by Sinai, Ruelle and Bowen. See [S], [R1] and [B] and the references therein.

Next we investigate the discrepancy between entropy and the sum of positive Lyapunov exponents and show that it can be expressed precisely in terms of fractal dimension. Let ν be a probability measure on a metric space, and let $B(x,r)$ denote the ball of radius r centered at x.

Definition 2.2 *We say that the dimension of ν, written $\dim(\nu)$, is well defined and equal to α if for ν-a.e. x,*

$$\lim_{\varepsilon\to 0} \frac{\log \nu B(x,\varepsilon)}{\log \varepsilon} = \alpha \ .$$

The existence of $\dim(\nu)$ means that locally ν has a scaling property. This is clearly not true for all measures.

Theorem 7 [LY2] *Corresponding to every $\lambda_i \neq 0$, there is a number δ_i with $0 \leq \delta_i \leq \dim E_i$ such that*

(a) $h_\mu(f) \ = \ \sum \lambda_i^+ \delta_i \ = \ -\sum \lambda_i^- \delta_i$;

(b) $\dim(\mu|W^u) = \sum_{\lambda_i>0} \delta_i$, $\dim(\mu|W^s) = \sum_{\lambda_i<0} \delta_i$.

The numbers δ_i have the interpretation of being the *partial dimensions* of μ in the directions of the subspaces E_i. Here $\mu|W^u$ refers to the conditional measures of μ on unstable manifolds. The fact that these conditional measures have well defined dimensions is part of the assertion of the theorem.

Observe that the entropy formula in Theorem 7 can be thought of as a refinement of Theorems 4 and 5: When μ is equivalent to the Riemannian volume on M, $\delta_i = \dim E_i$ and the formula in Theorem 7(a) is Pesin's formula. In general, we have $\delta_i \leq \dim E_i$. Plugged into (a) above, this gives Ruelle's Inequality. We may think of $\dim(\mu|W^u)$ as a measure of the *dissipativeness* of (f,μ) in forward time.

I would like to include a very brief sketch of a proof of Theorem 7, using it as an excuse to introduce the notion of *entropy along invariant foliations.*

Sketch of Proof *I. The "conformal" case.* By "conformal", we refer here to situations where all the Lyapunov exponents are $= \lambda$ for some $\lambda > 0$. This means in particular that f cannot be invertible, but let us not be bothered by that. In fact, let us pretend that locally, f is a dilation by e^λ. Let $B(x,n,\varepsilon)$ be as defined toward the end of Section 1. Then

$$B(x,n,\varepsilon) \ \sim \ B(x, \varepsilon e^{-\lambda n}) \ .$$

By Theorem 3,

$$\mu B(x,n,\varepsilon) \ \sim \ e^{-nh}$$

where $h = h_\mu(f)$. Comparing the expressions above and setting $r = e^{-\lambda n}$, we obtain

$$\mu B(x,r) \ \sim \ r^{\frac{h}{\lambda}} \ ,$$

which proves that $\dim(\mu)$ exists and $= \frac{h}{\lambda}$.

II. The general picture. Let $\lambda_1 > \cdots > \lambda_u$ be our positive Lyapunov exponents. It is a fact from nonuniform hyperbolic theory that for each $i \leq u$, there exists at μ-a.e. x an immersed submanifold $W^i(x)$ passing through x and tangent to $E_1(x) + \cdots + E_i(x)$. These manifolds are invariant, i.e. $fW^i(x) = W^i(fx)$, and the leaves of W^i are contained in those of W^{i+1}. Our strategy here is to work our way up this hierarchy of W^is, dealing with one exponent at a time.

For each i, we introduce a notion of entropy along W^i, written h_i. Intuitively, this number is a measure of the randomness of f along the leaves of W^i; it ignores what happens in transverse directions. Technically, it is defined using (infinite) partitions whose elements are contained in the leaves of W^i. We prove that for each i there exists δ_i such that

(i) $h_1 = \delta_1 \lambda_1$;
(ii) $h_i - h_{i-1} = \delta_i \lambda_i$ for $i = 2, \ldots, u$, and
(iii) $h_u = h_\mu(f)$.

The proof of (i) is similar to that in the "conformal" case since it involves only one exponent. To give an idea of why (ii) is true, consider the action of f on the leaves of W^i, and pretend somehow that a quotient dynamical system can be defined by collapsing the leaves of W^{i-1} inside W^i. This "quotient" dynamical system has exactly one Lyapunov exponent, namely λ_i. It behaves as though it leaves invariant a measure with dimension δ_i and has entropy $h_i - h_{i-1}$. A fair amount of technical work is needed to make this precise, but once properly done, it is again the single exponent principle at work. Summing the equations in (ii) over i, we obtain $h_u = \sum\limits_{i=1}^{u} \sigma_i \lambda_i$. Step (iii) says that zero and negative exponents do not contribute to entropy. This completes the outline of the proof in [LY2]. $\square$

Random dynamical systems

We close this section with a brief discussion of dynamical systems subjected to random noise. Let ν be a probability measure of $\mathrm{Diff}(M)$, the space of diffeomorphisms of a compact manifold M. Consider one- or two-sided sequences of diffeomorphisms

$$\cdots, \; f_{-1}, \; f_0, \; f_1, \; f_2, \; \cdots$$

chosen independently with law ν. This setup applies to stochastic differential equations

$$d\xi_t \;=\; X_0 dt \;+\; \sum_i X_i \circ dB_t^i$$

where the X_i are time-independent vector fields and the f_i are time-one-maps of the stochastic flow.

The random dynamical system above defines a Markov process on M with $P(E|x)$ $= \nu\{f, f(x) \in E\}$. Let μ be the marginal of a stationary measure of this process, i.e.

$\mu = \int f_*\mu \; d\nu(f)$. As individual realizations of this process, our random maps also have a system of invariant measures $\{\mu_{\bar f}\}$ defined for $\nu^{\mathbb{Z}}$-a.e. $\bar f = (f_i)_{i=-\infty}^{\infty}$. These measures are invariant in the sense that $(f_0)_*\mu_{\bar f} = \mu_{\sigma \bar f}$ where σ is the shift operator, and they are related to μ by $\mu = \int \mu_{\bar f} \; d\nu^{\mathbb{Z}}$.

Extending slightly ideas for a single diffeomorphism, it is easy to see that Lyapunov exponents and entropy are well defined for almost every sequence $\bar f$ and are nonrandom. We continue to use the notation $\lambda_1 > \cdots > \lambda_r$ and h.

Theorem 8 [LY3] *Assume that μ has a density wrt Lebesgue measure. If $\lambda_1 > 0$, then*

$$h \;=\; \sum \lambda_i^+ \dim E_i \; .$$

We remark that μ has a density if the transition probabilities $P(\cdot|x)$ do. In light of Theorem 6, when $\lambda_1 > 0$ the $\mu_{\bar f}$ may be regarded as *random SRB measures.*

3 Other interpretations of entropy

3.1 Entropy and volume growth

Let M be a compact m-dimensional C^∞ Riemannian manifold, and let $f : M \to M$ be a C^1 mapping. In this subsection, $h(f)$ is the topological entropy of f. As we will see, there is a strong relation between $h(f)$ and the rates of growth of areas or volumes of f^n-images of embedded submanifolds. To make these ideas precise, we begin with the following definitions:

For $\ell, k \geq 1$, let $\Sigma(k,\ell)$ be the set of C^k mappings $\sigma : Q^\ell \to M$ where Q^ℓ is the ℓ-dimensional unit cube. Let $\omega(\sigma)$ be the ℓ-dimensional volume of the image of σ in M counted with multiplicity, i.e. if σ is not one-to-one, and the image of one part coincides with that from another part, then we will count the set as many times as it is covered. For $n = 1, 2, \cdots$, $k \leq \infty$ and $\ell \leq m$, let

$$V_{\ell,k}(f) \;=\; \sup_{\sigma \in \Sigma(k,\ell)} \; \limsup_{n\to\infty} \frac{1}{n} \log \omega(f^n \circ \sigma) \; ,$$

$$V(f) \;=\; \max_{\ell} \; V_{\ell,\infty}(f) \; ,$$

and

$$R(f) \;=\; \lim_{n\to\infty} \frac{1}{n} \log \; \max_{x\in M} \|Df^n(x)\| \; .$$

Theorem 9 (i) [N] *If f is $C^{1+\varepsilon}$, $\varepsilon > 0$, then $h(f) \;\leq\; V(f)$.*
(ii) [Yom] *For $f \in C^k$, $k = 1, \cdots, \infty$,*

$$V_{\ell,k}(f) \;\leq\; h(f) + \frac{2\ell}{k} R(f) \; .$$

In particular, for $f \in C^\infty$, (i) and (ii) together imply $h(f) = \; V(f)$.

For (i), Newhouse showed that $V(f)$ as a volume growth rate is in fact attained by a large family of disks of a certain dimension. The factor $\frac{2\ell}{k}R(f)$ in (ii) is a correction term for pathologies that may (and do) occur in low differentiability.

Ideas related to (ii) are also used to resolve a version of the *Entropy Conjecture*. Let $S_\ell(f), \ell = 0, 1, \cdots, m$, denote the logarithm of the spectral radius of $f_* : H_\ell(M, \mathbb{R}) \to H_\ell(M, \mathbb{R})$ where $H_\ell(M, \mathbb{R})$ is the ℓth homology group of M, and let $S(f) = \max_\ell S_\ell(f)$.

Theorem 10 [Yom] *For $f \in C^\infty$, $S(f) \leq h(f)$.*

Intuitively, $S(f)$ measures the complexity of f on a global, topological level; it tells us which handles wrap around which handles and how many times. These crossings create "generalized horseshoes" which contribute to entropy. This explains why $S(f)$ is a lower bound for $h(f)$. On the other hand, entropy can also be created locally, say, inside a disk, without involving any action on homology. This is why $h(f)$ can be strictly larger.

3.2 Horseshoes and growth of periodic points

Recall that a "horseshoe" is a uniformly hyperbolic invariant set on which the map is topologically conjugate to $\sigma : \Sigma_2 \to \Sigma_2$, the full shift on two symbols. The next theorem says that in the absence of neutral directions, entropy can be thought of as carried by sets of this type.

Theorem 11 [K] *Let $f : M \to M$ be a C^2 diffeomorphism of a compact manifold, and let μ be an invariant Borel probability measure with no zero Lyapunov exponents. If $h_\mu(f) > 0$, then given $\varepsilon > 0$, there exist $N \in \mathbb{Z}^+$ and $\Lambda \subset M$ such that*

(i) $f^N(\Lambda) = \Lambda$ and $f^N|\Lambda$ is uniformly hyperbolic,

(ii) $f^N|\Lambda$ is topologically conjugate to $\sigma : \Sigma_s \to \Sigma_s$ for some s,

(iii) $\frac{1}{N} \log h_{\text{top}}(f^N|\Lambda) > \frac{1}{N} \log h_\mu(f) - \varepsilon$.

In particular, if dim(M) $= 2$ and $h_{\text{top}}(f) > 0$, then (i) and (ii) hold and the inequality in (iii) is valid with $h_\mu(f)$ replaced by $h_{\text{top}}(f)$.

The assertion for $\dim(M) = 2$ follows from the first part of the theorem because in dimension two, any μ with $h_\mu(f) > 0$ has one positive and one negative Lyapunov exponent.

Since shift spaces contain many periodic orbits, the following is an immediate consequence of Theorem 11:

Theorem 12 [K] *Let f be a C^2 surface diffeomorphism. Then*

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n \;\geq\; h_{\text{top}}(f)$$

where P_n is the number of fixed points of f^n.

For Anosov and Axiom A diffeomorphisms in any dimension, we in fact have

$$\lim_{n\to\infty} \frac{1}{n} \log P_n \ = \ h_{\text{top}}(f) \ ,$$

(see [B]) but this is a somewhat special situation. Recent results of Kaloshin show that in general P_n can grow much faster than entropy.

For more information on the relations between topological entropy and various growth properties of a map or flow, see [KH].

3.3 Large deviations and rates of escape

We will state some weak results that hold quite generally and some stronger results that hold in specialized situations. This subsection is taken largely from [You].

Let $f : X \to X$ be a continuous map of a compact metric space, and let m be a Borel probability measure on X. We think of m as a reference measure, and assume that there is an invariant probability measure μ such that for all continuous functions $\varphi : X \to \mathbb{R}$, the following holds for m-a.e. x:

$$\frac{1}{n} S_n\varphi(x) \ := \ \frac{1}{n}\sum_{i=0}^{n-1} \varphi(f^i x) \ \to \ \int \varphi d\mu \ := \ \bar{\varphi} \qquad \text{as} \ \ n \to \infty \ .$$

For $\delta > 0$, we define

$$E_{n,\delta} \ = \ \{x \in X : |\frac{1}{n} S_n\varphi(x) - \bar{\varphi}| > \delta\}$$

and ask how fast $mE_{n,\delta}$ decreases to zero as $n \to \infty$.

Following standard large deviation theory, we introduce a dynamical version of relative entropy: Let ν be a probability measure on X. We let $h_m(f;\nu)$ be the essential supremum with respect to ν of the function

$$h_m(f, x) \ = \ \lim_{\varepsilon\to 0} \ \limsup_{n\to\infty} \ -\frac{1}{n} \log mB(x, n, \varepsilon)$$

where $B(x, n, \varepsilon)$ is as defined in Section 1. (See also Theorem 3.) Let $\mathcal{M}$ denote the set of all f-invariant Borel probability measures on X, and let $\mathcal{M}_e$ be the set of ergodic invariant measures. The following is straightforward:

Proposition 3.1 *For every $\delta > 0$,*

$$\limsup_{n\to\infty} \ \frac{1}{n} \log mE_{n,\delta} \ \geq \ \sup\left\{ h_\nu(f) - h_m(f;\nu) : \nu \in \mathcal{M}_e, \ |\int \varphi d\nu - \bar{\varphi}| > \delta \right\} \ .$$

Without further conditions on the dynamics, a reasonable upper bound cannot be expected. More can be said, however, for systems whose dynamics have better statistical properties. A version of the following result was first proved by Orey and Pelikan. For $\nu \in \mathcal{M}_e$, let λ_ν be the sum of the positive Lyapunov exponents of (f, ν) counted with multiplicity.

Theorem 13 *Let $f|\Lambda$ be an Axiom A attractor, and let $U \supset \Lambda$ be its basin of attraction. Let m be the (normalized) Riemannian measure on U, and let μ be the SRB measure on the attractor (see Section 2). Then $\frac{1}{n}S_n\varphi$ satisfies a large deviation principle with rate function*

$$k(s) \;=\; -\sup\left\{h_\nu(f) - \lambda_\nu : \nu \in \mathcal{M}_e,\; \int \varphi d\mu = s\right\} \;.$$

A similar set of ideas applies to *rates of escape* problems. Consider a differentiable map $f : M \to M$ of a manifold with a compact invariant set Λ (which is not necessarily attracting). Let U be a compact neighborhood of Λ and assume that Λ is the maximal invariant set in $\bar{U}$, the closure of U. Define

$$E_{n,U} \;=\; \{x \in \bar{U} : f^i x \in \bar{U},\; 0 \le i < n\} \;.$$

Here, $\mathcal{M}$ is the set of invariant measures in $\bar{U}$. For $\nu \in \mathcal{M}$, let λ_ν be as before, averaging over ergodic components if ν is not ergodic. Let m be Lebesgue measure as before. The uniformly hyperbolic case of the next theorem is first proved in [B].

Theorem 14 *In general,*

$$\liminf_{n\to\infty} \frac{1}{n} \log m E_{n,U} \;\ge\; \sup\{h_\nu(f) - \lambda_\nu : \nu \in \mathcal{M}\} \;.$$

If Λ is uniformly hyperbolic or partially uniformly hyperbolic, then the limit above exists and is equal to the right side.

The reader may recall that $h_\nu(f) \le \lambda_\nu$ is Ruelle's Inequality (Theorem 5). The quantity on the right side of Theorem 14 is called *topological pressure*. There is a variational principle similar to that in Theorem 2 but with a potential (see [R1] or [B] for the theory of equilibrium states of uniformly hyperbolic systems).

We finish with the following interpretations: The numbers $\lambda_\nu,\ \nu \in \mathcal{M}$, describe the forces that push a point away from Λ. For example, if Λ consists of a single fixed point of saddle type, then the sum of the logarithms of its eigenvalues of modulus greater than one is precisely the rate of escape from a neighborhood of Λ. The entropies of the system, $h_\nu(f)$, represent, in some ways, the forces that keep a point in U, so that $h_\nu(f) - \lambda_\nu$ gives the net escape rate. One cannot, however, expect equality in general in Theorem 14, the reason being that not all parts of an invariant set are "seen" by invariant measures. For a simple example, consider the time-one-map of the "figure 8" flow where $\Lambda \subset \mathbb{R}^2$ consists of a saddle fixed point p together with its separatrices both of which are homoclinic orbits. If $|\det Df(p)| < 1$, then Λ is an attractor, from whose neighborhood there is no escape. For its unique invariant measure $\mu = \delta_p$, however, $0 = h_\nu < \lambda_\nu$.

References

[AKM] R. Adler, A. Konheim and M. McAndrew, *Topological entropy*, Trans. Amer. Math. Soc., **114** (1965), 309-319.

[B] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes in Math. **470** (1975).

[BK] M. Brin and A. Katok, *On local entropy*, Geometric Dynamics, Springer Lecture Notes, **1007** (1983), 30-38.

[K] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, Publ. Math. IHES, **51** (1980), 137-174.

[KH] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press 1995.

[L] F. Ledrappier, *Proprietes ergodiques des mesures de Sinai*, Publ. Math. IHES **59** (1984), 163-188.

[LS] F. Ledrappier and J.-M. Strelcyn, *A proof of the estimation from below in Pesin entropy formula*, Ergod. Th. & Dynam. Sys., **2** (1982), 203-219.

[LY1] F. Ledrappier and L.-S. Young, *The metric entropy of diffeomorphisms, Part I*, Ann. Math. **122** (1985), 509-539.

[LY2] F. Ledrappier and L.-S. Young, *The metric entropy of diffeomorphisms, Part II*, Ann. Math. **122** (1985), 540-574.

[LY3] F. Ledrappier and L.-S. Young, *Entropy formula for random transformations*, Prob. Th. Rel. Fields **80** (1988), 217-240.

[N] S. Newhouse, *Entropy and Volume*, Ergod. Th. & Dynam. Sys., **8** (1988), 283-299.

[P] Ya. Pesin, *Characteristic Lyapunov exponents and smooth ergodic theory*, Russ. Math. Surveys **32** (1977), 55-114.

[PS] C. Pugh and M. Shub, *Ergodic attractors*, Trans. AMS **312** (1989), 1-54.

[R1] D. Ruelle, *Thermodynamic Formalism*, Addison Wesley, Reading, MA, 1978.

[R2] D. Ruelle, *An inequality of the entropy of differentiable maps*, Bol. Sc. Bra. Mat. **98** (1978), 83-87

[S] Ya. G. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surveys **27** No. 4 (1972), 21-69.

[W] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Math., Springer-Verlag (1982).

[Yom] Y. Yomdin, *Volume growth and entropy*, Israel J. of Math., **57** (1987), 285-300.

[You] L.-S. Young, *Some large deviation results for dynamical systems*, Trans. Amer. Math. Soc., Vol 318, No. 2 (1990), 525-543.