Ergodic Theory of Chaotic Dynamical Systems

Lai-Sang Young [1] [2]

This is the text of the author's plenary lecture at the International Congress of Mathematical Physics in 1997

This article is about the ergodic theory of differentiable dynamical systems in finite dimensions. Limiting our discussions to discrete time, we are concerned with iterations of maps from $\mathbb{R}^n$ or finite dimensional manifolds to themselves. We consider only maps that generate chaotic dynamics, and our focus is on their statistical properties.

Complex geometric behavior in dynamical systems was observed by Poincaré, but no attempt was made to study it systematically until the 1960's when Smale proposed the idea of a hyperbolic invariant set or "horseshoe" [37]:

DEFINITION 1 A diffeomorphism $f : M \to M$ is said to be *uniformly hyperbolic* on a compact invariant set Λ if at each $x \in \Lambda$ the tangent space splits into Df-invariant subspaces E^u_x and E^s_x with the following properties: $\exists C \geq 1$ and $\lambda > 1$ independent of x such that $\forall n \geq 0$, $|Df^n_x v| \geq C^{-1}\lambda^n|v|$ for $v \in E^u_x$ and $|Df^n_x v| \leq C\lambda^{-n}|v|$ for $v \in E^s_x$.

Hyperbolic sets were used as geometric models of complex behavior because hyperbolicity implies a sensitive dependence on initial conditions: the orbits of most pairs of nearby points diverge exponentially fast in both forward and backward times. A system is said to satisfy *Axiom A* if all of its essential parts are uniformly hyperbolic (see [37]).

In the 10 years or so since Smale's invention of Axiom A, an ergodic theory for these systems emerged, due largely to Sinai [35] who first developed the theory for Anosov systems and to Ruelle [32] who worked in the more general Axiom A setting using Markov partitions constructed by Bowen [7]. Sinai and Ruelle brought various ideas from statistical mechanics into dynamical systems. The connection goes like this: given a partition of the phase space, one can assign to each point a bi-infinite symbol sequence describing its itinerary; this sequence can then be thought of as a configuration in a 1-dimensional statistical mechanical system. This dictionary works very well for Axiom A systems, a satisfactory ergodic theory of which was developed by the mid 1970's.

Since then hyperbolic theory has expanded considerably in scope, and some of the concepts have been clarified. It is these post-Axiom A developments that I wish to write about in this article. Section 1 contains some highlights of a nonuniform hyperbolic theory which studies maps that are hyperbolic almost everywhere with respect to some invariant measure. Alongside this general theory, detailed analyses of several specific classes of (non-Axiom A) maps have also been carried out. I will present a few examples in Section 2, including billiards and the Hénon attractors. In Section 3, I would like to report on some recent work of my own on the topic of correlation decay and related problems in systems with some hyperbolic behavior.

[1]Department of Mathematics, University of California, Los Angeles, CA 90095-1555 Email: lsy@math.ucla.edu

[2]The author is partially supported by the NSF

In a short note such as this one, it is impossible to give a systematic account of an area, only snapshots that will hopefully convey its flavor. While snapshots record the scenery, they also reflect the choices of the individual pointing the camera. This is very much the case here.

1 Nonuniform Hyperbolic Theory

While some natural objects such as geodesic flows on compact manifolds of negative curvature are uniformly hyperbolic [1], people have, over the years, come to realize that uniformly hyperbolic sets are too restrictive as models of chaos. One way to weaken this notion is to require only nonzero Lyapunov exponents. Recall that λ is called a Lyapunov exponent at x if $|Df_x^n v| \sim e^{\lambda n}$ for some tangent vector v at x. Oseledec's theorem [28] tells us that Lyapunov exponents are well defined almost everywhere with respect to invariant measures. The local picture along typical orbits, including the almost everywhere existence of stable and unstable manifolds, was worked out by Pesin [29]. (An exposition of some of this material is given in [40].)

1.1 Physically relevant invariant measures

Taking the view that only properties that hold on positive Lebesgue measure sets are observable, a first attempt to characterize physically relevant invariant measures is to require that they have densities. Thus for a Hamiltonian system, Liouville measure is regarded as the relevant measure (even though the system may admit many other invariant measures). This criterion runs into difficulties with dissipative systems. Consider, for example, a volume decreasing map f with an *attractor* Λ, by which we refer to a compact invariant set with the property that all points in a neighborhood U of Λ (called its *basin*) are attracted to Λ, *i.e.* for all $x \in U$, $f^n x \to \Lambda$ as $n \to \infty$. Since the dynamics on $U - \Lambda$ is transient, all invariant measures are supported on Λ, but Λ must have Lebesgue measure 0 if f is volume decreasing.

This leads to the notion of *Sinai-Ruelle-Bowen* or *SRB measures*. The idea of an SRB measure is that the invariant measure itself need not have a density as long as its properties are reflected on a positive Lebesgue measure set. Thus in the attractor situation above, we would regard an invariant measure as physically relevant as long as it governs the behavior of a positive Lebesgue measure set of points in the basin. This may not seem feasible: how can a measure tell us about points far away from its support?

The following definition is motivated by the work of Sinai [35] and Ruelle [33] on Axiom A attractors (by which we include Anosov systems):

DEFINITION 2 Let f be a diffeomorphism and μ an f-invariant Borel probability measure with some positive Lyapunov exponents μ-a.e. We call μ an *SRB measure* if the conditional measures of μ on unstable manifolds are compatible with the volume elements induced on these submanifolds.

A few remarks are in order. First, why are SRB measures physically relevant? We say that a point x is generic with respect to a measure μ if for all continuous observables φ,

$\frac{1}{n}\sum_0^{n-1}\varphi\circ f^i(x)\to\int\varphi d\mu$. Note that if x is μ-generic, then so is every point y on its stable manifold since the distance between f^nx and f^ny tends to 0 as $n\to\infty$. Let μ be an ergodic SRB measure with no zero Lyapunov exponents. Then with respect to the induced Riemannian measure on some unstable manifold, almost every point is μ-generic, and the set of stable manifolds through these μ-generic points form a positive Lebesgue measure set in phase space. (We have used implicitly the absolute continuity property of the stable foliation; see [31]).

Our second remark concerns the relation between SRB measures and invariant measures with densities. It is a fact that if f is a diffeomorphism, μ has a density, and μ-a.e. f has no zero Lyapunov exponents, then μ has absolutely continuous conditional measures on unstable manifolds [30]. Thus in a sense SRB measures *generalize* the notion of smooth invariant measures for conservative systems.

Our third remark concerns the *existence* or *prevalence* of SRB measures. The fact that we have articulated a definition does not mean that there really are measures with these properties in real life. The situation is as follows. Axiom A attractors always admit SRB measures; they are constructed in [35], [33] and [8]. Outside of the Axiom A category, numerical evidence is generally in favor of existence in the sense that for arbitrarily chosen initial conditions in the basin of an attractor, trajectory plots tend to produce pictures that are very much alike. This suggests that many orbits share the same set of statistics, something implied by the presence of an SRB measure. Analytically, there are few results beyond Axiom A. The Hénon attractors, which are perhaps the simplest, genuinely nonuniformly hyperbolic attractors, are shown to admit SRB measures only a few years ago (see Section 2.3); attractors without SRB measures have also been observed recently [17]. Away from specific examples these existence questions are very difficult and are not likely to be resolved in the near future.

1.2 Structure of attractors

THEOREM 1 ([21], [30]). *Let μ be an SRB measure, and assume that (f,μ) has no zero Lyapunov exponents. Then*
(a) *μ has at most a countable number of ergodic components $\{\mu_i\}$;*
(b) *for each i, there is a decomposition of the support of μ_i into disjoint measurable sets $X_1^i,\cdots,X_n^i$ cyclically permuted by f with the property that for each j, $(f^n,\mu_i|X_j^i)$ is mixing and isomorphic to a Bernoulli shift.*

Recall that (f,μ) is called *mixing* if $\mu(f^{-n}A\cap B)\to\mu(A)\mu(B)$ as $n\to\infty$ for all measurable sets A and B. Mixing properties will be discussed further in Section 3.

It is easy to get some feeling for what the ergodic components look like. First, points on a stable manifold W^s (resp. unstable manifold W^u) belong in the same ergodic component because their trajectory averages for continuous observables in forward (resp. backward) time converge to the same number. Recalling the picture in the paragraph on why SRB measures are physically relevant, we see immediately that generic points of ergodic components of SRB measures occupy positive Lebesgue measure sets (hence there is at most a countable number of them). This is a version of an argument due to Hopf. More generally, if $\gamma^{(1)},\cdots,\gamma^{(k)}$ are positive μ-measure sets of local W^u- and

W^s-leaves, W^u for odd i and W^s for even i, and if for each i every leaf in $\gamma^{(i)}$ intersects every leaf in $\gamma^{(i+1)}$, then the union of the $\gamma^{(i)}$'s belong in the same ergodic component. Intuitively, ergodic components can be thought of as the union of the largest sets reached by these W^u/W^s-chains and their images.

When studying ergodicity questions one sometimes distinguishes between *local* and *global* issues. Local ergodicity is about whether or not ergodic components are open modulo sets of Lebesgue measure 0. In systems that are not Axiom A, local W^u- and W^s-manifolds vary measurably; they differ in sizes and may twist and turn as the base point varies. It is not clear if arbitrarily nearby points are connected by W^u/W^s-chains. Global ergodicity, on the other hand, has to do with the transitivity of "large" regions, such as whether there are "walls" separating the phase space into non-interacting domains.

1.3 Entropy, Lyapunov exponents and dimension

We assume throughout this subsection that f is a C^2 diffeomorphism of a compact Riemannian manifold and μ is an f-invariant Borel probability measure. We consider here two different ways of measuring dynamical complexity: Lyapunov exponents, which measure *geometrically* how fast orbits diverge, and the metric entropy of (f,μ), which measures randomness in the sense of *information.* (The notion of entropy for a transformation was introduced by Kolmogorov and Sinai around 1960; it measures the amount of uncertainty one faces when attempting to predict future behaviors of orbits based on knowledge of their pasts.) The distinct Lyapunov exponents of (f,μ) are denoted by $\lambda_1 > \lambda_2 > \cdots > \lambda_r$, E_i are the linear subspaces corresponding to λ_i, and entropy is written $h_\mu(f)$. Our first theorem gives the relation between these two sets of invariants.

THEOREM 2 ([34], [30]; [22], [21], [23]). Writing $a^+ = \max\,(a,0)$,

$$h_\mu(f) \;\le\; \int \sum \lambda_i^+ \dim E_i \; d\mu \;;$$

equality holds if and only if μ is SRB.

The inequality above is known as *Ruelle's Inequality*; when equality holds, it is called *Pesin's Formula.* One way to understand these results is as follows: Randomness is created by the separation of nearby orbits, which in turn is caused by expansions in a map. In a conservative system, or when the measure is SRB, all the expansion goes back into the system to make entropy, hence we have Pesin's formula. A strict inequality signifies some "wasted" expansion or "leakage" from the system, which can happen only if the invariant measure has "holes" on unstable manifolds.

This leads naturally to the idea of *dimension.* Let ν be a Borel probability measure on a metric space X, and let $B(x,r)$ denote the ball of radius r about x. We say $dim(\nu)$ is well defined and is equal to α if for ν-a.e. x, $\nu B(x,r) \sim r^\alpha$ as $r \to 0$. The relation between $dim(\cdot)$ and Hausdorff dimension is that if $dim(\nu) = \alpha$, then $\text{Inf}\{HD(Y) : Y \subset X, \nu(Y) = 1\} = \alpha$. A practical use of dimension is that it tells us how many variables are needed to faithfully describe an object.

Returning to our dynamical system (f,μ), let $\mu|W^u$ denote the conditional measures of μ on unstable manifolds.

THEOREM 3 (Part II of [23]) *Assume for simplicity that* (f, μ) *is ergodic. Then corresponding to each* λ_i*, there is a number* σ_i*,* $0 \leq \sigma_i \leq 1$*, such that*

(a) $dim(\mu|W^u)$ *exists and is* $= \sum_{\lambda_i>0} \sigma_i dimE_i$*;*

(b) $h_\mu(f) = \sum_i \lambda_i^+ \sigma_i dimE_i$*.*

The numbers "$\sigma_i \mathrm{dim} E_i$" are essentially the dimensions of μ in the direction of E_i. In the case where μ is SRB, $\sigma_i = 1$ for all i, so that (a) above becomes $dim(\mu|W^u) = \sum_{\lambda_i>0} \mathrm{dim}E_i = dimW^u$ and (b) becomes Pesin's formula. It has been proved very recently that in the absence of zero exponents, $dim(\mu) = dim(\mu|W^u) + dim(\mu|W^s)$; in particular, the limit in the definition of $dim(\mu)$ always exists [2].

As a model of a randomly perturbed dynamical system consider the composition $f_0 \circ f_1 \circ f_2 \circ \cdots$ where the f_i's are an *iid* sequence with repect to a probability measure on the space of C^2 diffeomorphisms of a manifold. (This setup is compatible with that of stochastic differential equations; see e.g. [20].) Let μ be an invariant measure for this process, and let $\{\mu_\omega\}$ denote the distintegration of μ on bi-infinite sample paths $\omega = \{f_i\}_{i=-\infty}^{\infty}$. Dynamical invariants such as Lyapunov exponents, entropy and dimension continue to make sense in this setting; moreover, they are nonrandom.

THEOREM 4 ([24], [25]). *Assume that the* f_i*'s are sufficiently random. Then:*
(a) *if* $\lambda_1 > 0$*, then a.s. the* μ_ω*'s have the SRB property;*
(b) *if* $\lambda_i \neq 0$ $\forall i$*, then there is an* i^* *s.t.* $\sigma_i = 1$ *for* $i < i^*$ *and* $\sigma_i = 0$ *for* $i > i^*$*.*

Thus mass has a tendency to align itself with the more expanding directions when a system is stochastically purturbed. This is an example of the simplified dynamical picture created by the averaging effects of random noise.

1.4 Approximations by horseshoes

We have discussed so far two ways of describing hyperbolic behavior in a dynamical system: Axiom A, and (f, μ) where μ is an invariant probability measure with $\lambda_i \neq 0$ $\forall i$. We wish now to clarify the relation between these models. First, given (f, μ) as above, it is always possible to approximate its dynamics by uniformly hyperbolic sets. The following theorem gives two ways of doing that:

THEOREM 5 (a) [30] *There exist closed sets* $\Lambda_1 \subset \Lambda_2 \subset \cdots$ *with* $\mu(\cup\Lambda_i) = 1$ *such that orbits starting from each* Λ_i *are uniformly hyperbolic, although the strength of hyperbolicity may decrease as i increases.* (The Λ_i's here are, in general, *not* invariant.)
(b) [19] f *leaves invariant uniformly hyperbolic sets* Λ *with* $h_{top}(f|\Lambda) > h_\mu(f) - \varepsilon$ *for every* $\varepsilon > 0$. (Here $h_{top}(\cdot)$ is topological entropy, and $\mu(\Lambda) = 0$ in general.)

Thus, for example, if an SRB measure with no zero exponents exists and gives positive mass to every open set of an attractor, then it follows that uniformly hyperbolic sets or horseshoes are arbitrarily dense on the attractor. The converse, however, is not true (see *e.g.* [17]). The existence of an invariant measure with nonzero exponents implies not only the presence of horseshoes but also certain coherent structures among them.

2 Some Specific Examples

The last 20-30 years also saw the emergence of detailed analyses of several specific classes of (non-Axiom A) examples including logistic maps, billiards etc. These examples were originally studied as unrelated objects using techniques adapted to the individual characateristics of each class. It is becoming clear, however, that many of them can be understood under one unifying framework, which I will describe in 2.1.

2.1 Systems with localized sources of nonhyperbolicity

Consider a map f with the following properties:

(a) There is an identifiable, localized (*i.e.* small), compact set, not necessarily invariant, that is responsible for all the nonhyperbolic or nonuniformly hyperbolic behavior in the system; that is to say, away from this "bad set" the system is uniformly hyperbolic, but most orbits get near this set at some point in time, and their hyperbolicity is spoiled to varying degrees with each visit to the vicinity of this set.

(b) Both the mechanism with which hyperbolicity is spoiled and the mechanism with which it is eventually recovered are known to us.

What we have described above is a natural generalization and considerable broadening of the Axiom A condition (which requires that the "bad set" be empty). The existence of an identifiable, localized "bad set", however, is a rather special property among all maps with some hyperbolic behavior; this therefore is a strictly smaller subclass. In the next few pages I will present three of the most important nonuniformly hyperbolic examples that have been studied, with a view toward these "bad set–recovery" ideas.

2.2 The logistic maps

Consider $f_a : [-1, 1] \to [-1, 1]$ defined by $f_a(x) = 1 - ax^2$, where $a \in [0, 2]$ is a parameter. Since we are in 1-dimension, hyperbolicity here means expansion, and the "bad set" consists of a single point, namely 0, the critical point of f. It is clear how expansion is lost: when an orbit comes to a distance δ of 0, it experiences a contraction of $\sim \delta$; the orbit recovers from this derivative loss if it subsequently stays away from the critical point for a sufficiently long time. The question is: will expansion or contraction prevail for typical orbits?

The answer to this very innocent question turns out to be less than simple. A partial answer is contained in the following two theorems:

THEOREM 6 [18]. *There exists a positive Lebesgue measure set of parameters a with the property that f_a has an invariant density with a positive Lyapunov exponent.*

THEOREM 7 [15]. *There exists an open and dense set of parameters a with the property that f_a has a periodic cycle which attracts the orbit of Lebesgue-a.e. point.*

Thus on a positive measure set of parameters, expansion wins, and on an open and dense set, contraction wins. This intermingling of parameters with antipodal behaviors underscores the complexity of the dynamical picture. It has been announced very recently that these two types of dynamical behaviors account for a full measure set of parameters [27].

To control the dynamics, it is natural to go to the source of nonhyperbolicity, *i.e.* to impose conditions on the critical orbit. The following ideas go back to [13]; we follow the formulation in [3], [4] where conditions (i) and (ii) are used to produce chaotic behavior for a positive measure set of parameters: for some $\alpha > 0$ and $\lambda > 1$,

(i) $|f_a^n(0)| \geq e^{-\alpha n}$ for all $n \geq 1$;

(ii) $|(f_a^n)'(f0)| \geq \lambda^n$ for all $n \geq 1$.

These conditions ensure a full and immediate recovery after each visit to the bad set: an orbit that comes near the critical point will follow the critical orbit for some time; condition (ii) says that the critical orbit is expanding, while condition (i) allows us to compare the derivatives of the two orbits. For maps satisfying (i) and (ii), the recovery time for a visit to $\sim \delta$ of 0 is easily estimated to be $\sim \log \frac{1}{\delta}$.

2.3 Hénon-type attractors

The Hénon maps are a 2-parameter family of maps of $\mathbb{R}^2$ given by $f_{a,b}(x,y) = (1 - ax^2 + y, bx)$. For $b = 0$ and $a < 2$, $f_{a,b}$ maps all of $\mathbb{R}^2$ onto the x-axis, and on the x-axis it is equal to f_a in the last subsection, so it is not hard to see that for b sufficiently small $f_{a,b}$ has an attractor which looks like the graph of the corresponding 1-dimensional map "thickened up". With this picture in mind, one sees that away from a small neighborhood of the y-axis, roughly horizontal vectors are expanded and stay roughly horizontal, whereas near the y-axis, Df rotates horizontal vectors to all possible directions.

A naive probabilistic model is as follows: suppose we flip a (very) biased coin, for which a head shows up 99% of the time. When a head occurs, we take the matrix $\begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, and when a tail occurs we pick a random rotation. These matrices are then multiplied together, and we wish to know if this random matrix product has a positive Lyapunov exponent. ("Head" here corresponds to when the orbit is away from the y-axis, and the matrix product corresponds to Df^n.)

It is an easy exercise to show that the random matrix product above has a positive Lyapunov exponent. This model, however, does not accurately reflect the true situation, and the answer to whether or not Hénon maps have positive exponents turns out to be a much harder one. To a casual observer, the matrices $Df_z,\ Df_{fz},\ Df_{f^2z}, \cdots$, where f is the Hénon map and $z \in \mathbb{R}^2$, may resemble the random sequence in the last paragraph, but there is absolutely nothing independent, nothing probabilistic at all, when going from one iterate to the next: the location of f^iz determines completely Df_{f^iz} and $f^{i+1}z$. This is the nature of *deterministic chaos*: the data generated may look random, but all is determined once an initial condition is chosen.

In [4] Benedicks and Carleson developed a machinery for analyzing the dynamics of $f_{a,b}$ for a positive measure set of parameters Δ near $a = 2$ and $b = 0$. The next theorem builds on their analysis:

THEOREM 8 [5]. *For $(a, b) \in \Delta$, $f = f_{a,b}$ admits an SRB measure μ. Moreover, μ is unique, and (f^n, μ) is ergodic for all n, which is equivalent to (f, μ) being mixing and Bernoulli.*

The analysis in [4] is very involved. Through an inductive construction (which can be carried through only for the selected parameters) the authors of [4] identified a fractal set near the y-axis as the source of all nonhyperbolicity. Points in this set are tangencies of stable and unstable manifolds. When an orbit passes near this "bad set", hyperbolicity is spoiled by the near-reversal of its stable and unstable directions, meaning that the direction most expanded in previous iterates will now undergo a series of contractions. As with the logistic maps, conditions imposed on the orbits of the "bad set" ensure full recovery, and these conditions are arranged through parameter selection.

2.4 Billiards

By a billiard, we refer as usual to the uniform motion of a point mass in $\Omega \subset \mathbb{R}^2$ or $\mathbb{T}^2$ where $\partial\Omega$ is the union of a finite number of smooth curves. Points in phase space are represented by (z, v) where $z \in \Omega$ and v is a unit vector in the direction of the flow. We assume that collisions with $\partial\Omega$ are totally elastic, so that the angle of incidence is equal to the angle of reflection. Let $M = \partial\Omega \times [-\frac{\pi}{2}, \frac{\pi}{2}]$ be the usual cross-section to the billiard flow, and let $f : M \to M$ be the section map. Then with (r, φ) denoting the coordinates in M, f preserves the measure $\mu = \cos\varphi dr d\varphi$.

A great deal has been written about billiards and their higher dimensional generalizations including the problem of hard balls (see e.g. [10], [36], [39], and [38] and the references theirin). We mention two well known mechanisms for producing nonzero Lyapunov exponents in 2-dimensional billiards:

(a) Concave boundaries: intuitively, nearly parallel rays approaching these boundary pieces become divergent upon reflection.

(b) Convex boundaries, such as those in the stadium [9], can also produce hyperbolicity if certain conditions are met, for even though nearly parallel rays first become convergent upon reflection, they diverge after focussing, and expansion in phase space results if, before the next collision, they have diverged more than they have converged. See [39] for precise formulations.

For billiards all of whose boundary pieces are concave, nonhyperbolicity is caused by trajectories meeting $\partial\Omega$ tangentially or going into "corners" (or points of nondifferentiability of $\partial\Omega$). Otherwise f is essentially uniformly hyperbolic. Its "bad set", therefore, is made up of curves of discontinuity. When boundary pieces of type (b) are present, one of the obstructions to hyperbolicity is when a billiard trajectory makes consecutive reflections off the convex pieces at nearly tangential angles. For the map f this translates into an orbit spending a long time near a part of the "bad set" consisting of a line of fixed points at which the derivative of f is parabolic.

3 Correlation decay and related problems

This section is a report on some recent and ongoing research of the author.

3.1 The problems

Our aim is to understand the statistical properties of dynamical systems with a lot of expansion or hyperbolicity on large parts but not necessarily all of their phase spaces. In particular we would like to address the following questions:

(1) Under what conditions can one conclude the existence of an SRB measure?

Assume that an SRB measure μ exists and is mixing (see Theorem 1).

(2) What is the speed of mixing? More importantly, what kinds of mechanisms produce the various speeds?

(3) When does the Central Limit Theorem hold?

By the speed of mixing, or the speed of correlation decay, we refer to the speed with which the quantity

$$C_n(\varphi, \psi) \;:=\; \left| \int (\varphi \circ f^n) \psi d\mu \;-\; \int \varphi d\mu \int \psi d\mu \right|$$

decreases to zero. Here φ and ψ are observables or test functions defined on phase space. We consider only Hölder continuous test functions, for in a totally deterministic setting such as ours mixing can be arbitrarily slow if we allow all measurable test functions.

In the next subsection we will propose some answers to these questions, but first, a sample of previously known results: For Axiom A maps, the questions discussed here were resolved in the 70's [32]. Correlation decay questions for Axiom A *flows* remain not well understood; for recent progress see [12], [14]. (Our discussion does not apply to flows.) There have also been a few results proving exponential mixing for examples that are not Axiom A (*e.g.* [16], [26]). Known techniques for proving exponential mixing include spectral gaps for the Perron-Frobenius or transfer operator and the invariant cones method proposed in [26] a few years ago. To my knowledge systematic methods for studying slower decay rates have – up until now – not been developed.

3.2 Renewal times, growth of unstable manifolds, and the speed of mixing

Motivated by countable Markov chains theory and examples from dynamics, we propose to formulate answers to Questions (1)–(3) above in terms of *recurrence* or *renewal times.* The precise setup is discussed in [41]. The idea is as follows: Pick an arbitrary set with reasonable properties, think of it as a reference set and regard a part of the dynamical system as having "renewed" itself when it makes a "full" return to this set, *i.e.*, when it crosses over the reference set completely, at least in the unstable direction. We propose to give conditions for (1)–(3) in terms of the asymptotics of $m\{R > n\}$ where R is the return time function and m is a reference measure (such as Lebesgue measure) on the reference set.

Let us describe a little more precisely this construction of dynamical renewal. Consider first an expanding map f. Let Λ be a small open disk. We run f until $f^n\Lambda \supset \Lambda$. Let $\Lambda_1 \subset \Lambda$ be such that f^n maps Λ_1 diffeomorphically onto Λ. We think of Λ_1 as having made a "full return" at time n and stop considering it. The rest of Λ is iterated until something else makes a "full return", and so on. If f has reasonable recurrence properties, this procedure will result in a decomposition of Λ into a disjoint union of sets $\cup_i \Lambda_i$ with the property that for each i, there is a positive integer R_i such that $f^{R_i}\Lambda_i = \Lambda$. The return time function R is defined to be $R|\Lambda_i = R_i$. For invertible hyperbolic maps, the picture is more complicated. To avoid messy estimates I would choose Λ with a product structure (*i.e.* Λ is the intersection of transversal families of W^u and W^s-disks) even though these sets are not open in general. Here m is Lebesgue measure on W^u, and we require that $m(\Lambda \cap W^u) > 0$.

There are also a few technical requirements, the most important of which is a regularity condition for $Df^{R_i}|\Lambda_i$ which puts a uniform bound on the nonlinearities of $f^{R_i}|\Lambda_i$. This is a natural condition for C^2 maps that are sufficiently expanding; it is responsible for ensuring some resemblance to independence for the dynamics between successive returns to the reference set. Please see [41] for the precise formulations.

We now explain how this construction is used to study the questions posed at the beginning of this section. Again we omit details, sketching only the main ideas.

First we relate the statistical properties of f to the asymptotics of $m\{R > n\}$. We call these "abstract" results because they do not depend on the characteristics of the individual dynamical system other than the tail of the return time function R.

ABSTRACT THEOREM ([41], [42]). *Let f, Λ, m and R be as above. Then:*
(a) *If $\int R dm < \infty$, then f admits an SRB measure μ.*
(b) *If, additionally,* $\gcd\{R_i\} = 1$, *then (f, μ) is mixing.*
(c) *If $m\{R > n\} < C\theta^n$ for some $\theta < 1$, then $\exists\ \tilde{\theta} < 1$ s.t. $\forall \varphi, \psi,\ C_n(\varphi, \psi) < C\tilde{\theta}^n$.*
(d) *If $m\{R > n\} = \mathcal{O}(n^{-\alpha})$ for some $\alpha > 1$, then $C_n(\varphi, \psi) = \mathcal{O}(n^{-\alpha+1})$.*
(e) *If R is as in* (d) *and $\alpha > 2$, then the CLT holds for all φ.*

Next, we argue that conceptually $m\{R > n\}$ is essentially the speed with which arbitrarily small pieces of unstable manifolds grow to a specified size. (This is *not* the same as Lyapunov exponents, which measure pointwise growth rates.) First we describe the picture. If f has good hyperbolic properties, then we can cover most of phase space with a finite number of sets $\Gamma_1, \cdots, \Gamma_k$ with product structures (they look like $W^u \times W^s$ trelises). If f is mixing, then in finite time, $f^n\Gamma_i$ crosses over Γ_j in the unstable direction for every i, j. These structures give the dynamics the flavor of a finite Markov chain, but one should not carry the analogy too far, for $\cup\Gamma_i$ is not all of phase space, nor is it an invariant set. The rest of phase space is made up of small bits of stable and unstable manifolds that twist and turn as described in Section 1.2. Returning to the problem of estimating $m\{R > n\}$, suppose that Γ_1 is our reference set. Since f is ergodic, it is inevitable that some parts of Γ_1 will get into the messy regions of phase space before they return. It is necessary, therefore, to know how long it takes structures of *arbitrarily small scales* to "straightout out" and grow to the scale of the Γ_i's. This is also sufficient, for once a W^u-leaf reaches a size comparable to the Γ_i's, it will soon cross over one of them, and once it crosses over one Γ_j, it will cross over Γ_1 in a finite number of steps via the Markov-like action on $\cup\Gamma_i$ described earlier on.

Finally we observe that while in general it is impossible to know the detailed structures of a map to arbitrarily small scales, the type of estimates to which the problem has now been reduced is feasible if we know the *rules* of the game. In particular, if there is a recognizable "bad set" with known mechanisms as described in Section 2.1, then the messy parts are created by interactions with the "bad set", which also determines how they evolve. The speed in question is therefore related to the speed with which the influence of the "bad set" is overcome.

3.3 Applications

In Section 3.2 we proposed a generic scheme for obtaining statistical information for dynamical systems with some hyperbolic behavior. We now implement this scheme for some well known examples. Most of the results discussed below, including those for the Hénon maps and billiards, are new. See [41] and [42] for additional references.

EXAMPLE 1 *Expanding maps in 1-d with a neutral fixed point* [42]. Here the "bad set" consists exactly of the neutral fixed point, which we call 0. If $f'(0) = 1$ and $f''(0) \approx |x|^{\gamma-1}$ for some $\gamma > 0$, then taking Λ to be a suitable interval, it is an easy exercise to see that $m\{R > n\} = \mathcal{O}(n^{-\alpha})$ where $\alpha = \frac{1}{\gamma}$. Once this is computed, the abstract theorem in 3.2 gives immediately the existence of an invariant probability density with correlation decay rate $\mathcal{O}(n^{-\frac{1}{\gamma}+1})$ for $\gamma < 1$, and the CLT for $\gamma < \frac{1}{2}$.

EXAMPLE 2 *Logistic and Hénon-type maps.* For the parameter values studied in [4], the time that it takes an orbit to regain its hyperbolicity after coming to a distance of δ from the "bad set" is $\sim \log \frac{1}{\delta}$ (see 2.2; the same estimate holds for the Hénon maps). Thus after each visit to the "bad set", it is as though there is unobstructed, uniform growth until the derivatives are fully recovered. This translates into the estimate $m\{R > n\} < C\theta^n$ for some $\theta < 1$, from which we conclude exponential decay of correlations and CLT. (In the case of the Hénon maps the constructions require more technical work than we have indicated; they are carried out in [6].)

EXAMPLE 3 *Billiards.* (a) First we consider billiards on $\mathbb{T}^2$ with convex scatterers and finite horizon. Earlier results [10] have shown that their correlation decay rates are bounded above by $\sim e^{-\sqrt{n}}$. I had the feeling this may not be the true decay rate, so I ran these much studied examples through the analysis in 3.2. Here is what I have found [41]: As observed in 2.4, the only obstruction to uniform growth along W^u-curves are a finite number of discontinuity curves transversal to W^u. To get an idea of what might happen, let γ be a short W^u-curve and imagine a scenario in which each component of $f^n\gamma$ is expanded by $\frac{3}{2}$ and cut into 2 roughly equal pieces with each iteration – it would be very hard for these components to grow to unit length! Unlike the logistic and Hénon maps, for which parameters are chosen to guarantee full and immediate recovery after each visit to the "bad set", the components of $f^n\gamma$ are not guaranteed to grow long before they get cut again. We rely instead on the geometry of billiards and a *statistical* argument: It is observed in [11] that no more than Kn branches of the discontinuity set of f^n can meet in one point, K depending only on the billiard table. Thus in n iterates the image of a sufficiently short W^u-curve has at most $Kn+1$ components while its total length grows by a factor of λ^n for some $\lambda > 1$. On average, therefore, exponential growth prevails. This translates, after some work, into the estimate $m\{R > n\} < C\theta^n$, from which we conclude that the speed of correlation decay is actually $\sim e^{-\alpha n}$. (b) We mention as a last example the stadium, in which we have to overcome parabolic behavior caused by trajectories (nearly) tangential to the circular pieces (see 2.4) and those perpendicular to the two straight sides. The correlation decay rate for this billiard map is not known; my preliminary investigations along the lines of 3.2 seem to suggest that it is $\sim \frac{1}{n}$.

References

[1] Anosov, D., Geodesic flows on closed Riemann manifolds with negative curvature, AMS Translation (1969) 1-235.

[2] Barreira, L., Pesin, Y. and Schmeling, J., Dimension of hyperbolic measures – a proof of the Eckmann-Ruelle Conjecture, preprint.

[3] Benedicks, M. and Carleson, L., On iterations of $1 - ax^2$ on $(-1, 1)$, Ann. Math. **122** (1985) 1-25.

[4] Benedicks, M. and Carleson, L., The dynamics of the Henon map, Ann. Math. **133** (1991) 73-169.

[5] Benedicks, M. and Young, L.-S., SBR measures for certain Henon maps, Inventiones Math. **112** (1993) 541-576.

[6] Benedicks, M. and Young, L.-S., Markov extension and decay or correlations for certain Hénon maps, 1996 preprint.

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

[8] Bowen, R. and Ruelle, D., The ergodic theory of Axiom A flows, Invent. Math. **29** (1975) 181-202.

[9] Bunimovich, L. A., On the ergodic properties of nowhere dispersing billiards, Commun. Math. Phys. **65** (1979) 295-312.

[10] Bunimovich, L. A., Sinai, Ya. G., and N. I. Chernov, Statistical properties of 2-dimensional hyperbolic billiards, Russ. Math. Surv., **46** (1991) 47-106.

[11] Bunimovich, L. A., Sinai, Ya. G., and Chernov, N. I., Markov partitions for two-dimensional billiards, Russ. Math. Surv., **45** (1990) 105-152.

[12] Chernov, N. I., Markov approximations and decay of correlations for Anosov flows, to appear in Ann. Math.

[13] Collet, P. and Eckmann, J.-P., Positive Lyapunov exponents and absolute continuity, Ergod. Th. & Dynam. Sys., **3** (1983) 13-46.

[14] Dolgopiat, D., On the decay of correlations in Anosov flows, to appear in Ann. Math.

[15] Graczyk, G. and Swiatek, G., Generic hyperbolicity in the logistic family, Ann. Math., **146** (1997) 1-52.

[16] Hofbauer, F. and Keller, G., Ergodic properties of invariant measures for piecewise monotonic transformations, Math. Z., **180** (1982), 119-140.

[17] Hu, H. and Young., L.-S., Nonexistence of SBR measures for some systems that are "almost Ansov", Ergod. Th. & Dynam. Sys., **15** (1995) 67-76.

[18] Jakobson, M., Absolutely continuous invariant measures for one-parameter families of one-dimensional maps, Commun. Math. Phys., **81** (1981), 39-88.

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

[20] Kunita, H., *Stochastic flows and stochastic differential equations*, Cambridge Univ. Press (1990).

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

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

[23] Ledrappier, F. and Young, L.-S., The metric entropy of diffeomorphisms, Ann. Math. **122** (1985) 509-574.

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

[25] Ledrappier, F. and Young, L.-S., Dimension formula for random transformations, Commun. Math. Phys. **117** (1988) 529-548.

[26] Liverani, C., Decay of correlations, Ann. Math. **142** (1995) 239-301.

[27] Lyubich, M., Almost every real quadratic map is either regular or stochastic, preprint.

[28] Oseledec, V. I., A multiplicative ergodic theorem: Liapunov characteristic numbers of dynamical systems, Trans. Moscow Math. Soc. **19** (1968) 197-231.

[29] Pesin, Ya. B., Families if invariant manifolds corresponding to non-zero characteristic exponents, Math. of the USSR, Izvestjia **10** (1978) 1261-1305.

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

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

[32] Ruelle, D., The thermodynamics formalism for expanding maps, Commun. Math. Phys., Vol. 125 (1989) 239-262.

[33] Ruelle, D., A measure associated with *Axiom A* attractors, Amer. J. Math. **98** (1976) 619-654.

[34] Ruelle, D., An inequality of the entropy of differentiable maps, Bol. Sc. Bra. Mat. **9** (1978) 83-87.

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

[36] Sinai, Ya. G., Dynamical systems with elastic reflections: ergodic properties of dispersing billiards, Russ. Math. Surveys **25**, No. 2 (1970) 137-189.

[37] Smale, S., Differentiable dynamical systems, Bull. AMS **73** (1967) 747-817.

[38] Szász, D., Boltzmann's Ergodic Hypothesis, a conjecture for centuries?, Studia Sci. Math. Hung. **31** (1996) 299-322.

[39] Wojtkowski, M., Principles for the design of billiards with nonvanishing Lyapunov exponents, Commun. Math. Phys. **105** (1986) 391-414.

[40] Young, L.-S., Ergodic theory of differentiable dynamical systems, *Real and complex dynamical systems*, Ed. Branner and Hjorth, Kluwer Acad. Publ. (1995) 293-336.

[41] Young, L.-S., Statistical properties of dynamical systems with some hyperbolicity, to appear in Ann. Math.

[42] Young, L.-S., Recurrence times and rates of mixing, 1997 preprint.