BENEATH THE NOISE, CHAOS

by

Steven P. Lalley
Purdue University

Technical Report #97-14

Department of Statistics
Purdue University
West Lafayette, IN USA

August 1997

BENEATH THE NOISE, CHAOS

STEVEN P. LALLEY
PURDUE UNIVERSITY

ABSTRACT. The problem of extracting a signal x_n from a noise-corrupted time series $y_n = x_n + e_n$ is considered. The signal x_n is assumed to be generated by a discrete-time, deterministic, chaotic dynamical system F – in particular, $x_n = F^n(x_0)$, where the initial point x_0 is assumed to lie in a compact hyperbolic $F-$invariant set. It is shown that (1) if the noise sequence e_n is gaussian then it is impossible to consistently recover the signal x_n, but (2) if the noise sequence consists of i.i.d. random vectors uniformly bounded by a constant $\delta > 0$, then it is possible to recover the signal x_n *provided* $\delta < 5\Delta$, where Δ is a *separation threshold* for F. A filtering algorithm for the latter situation is presented.

1. INTRODUCTION

Physical and numerical experiments carried out over the past 30+ years suggest that the phenomenon of *deterministic chaos* is ubiquitous in physical systems. Experience has shown that inference of the mathematical objects (the differential equations, equilibrium measures, Lyapunov exponents, etc.) governing the dynamics of such systems from time series data is a delicate problem even when this data is uncorrupted by noise. See [3] for an extensive review and bibliography. Inference from *noisy* data is therefore bound to be doubly difficult. Although various *ad hoc* "noise reduction" algorithms have been proposed (some seemingly quite effective when tested on computer-generated data from low-dimensional chaotic systems, e.g., [9] and [5]), their theoretical properties are largely unknown.

The purpose of this paper is to address the following basic question: Is it possible to consistently recover a "signal" $\{x_n\}_{n\in\mathbb{Z}}$ generated by an Axiom A system from a time series of the form

$$y_n = x_n + e_n \tag{1}$$

where e_n is observational noise? A positive answer would essentially reduce the problem of inference from *noisy* time series data to that of inference from *non*-noisy data. The following scenario for the signal will be considered here:

$$x_n = F^1(x_{n-1}) = F^n(x_0), \tag{2}$$

where F is a C^2 diffeomorphism and x_0 is a point lying in a hyperbolic invariant set or in the basin of attraction of a hyperbolic attractor (see section 2 for definitions and examples). Our main result is that the possibility of consistent signal extraction depends on the nature and amplitude of the noise. If the noise e_n is uniformly bounded and the bound is below a certain threshold Δ then consistent signal extraction is possible; but if the noise is *unbounded*, in particular gaussian, then consistent signal extraction is impossible (even when the L^2-norm of e_n is far below the threshold Δ).

Date. July 15, 1997.
Supported by NSF Grant DMS-.

In a companion paper [7] we shall consider a different but related scenario for the signal x_n, which is technically (and perhaps also conceptually) more difficult but probably of greater practical importance. In this scenario, the underlying dynamical system is a topologically mixing Axiom A *flow* F^t, but observations on the orbit $x_t = F^t(x_0)$ are made only at integer times n. It will be shown that the dichotomy between bounded and unbounded noise persists, and an algorithm for noise removal (more complicated than that given in this paper) will be presented.

We must emphasize at the outset that the results of this and the companion paper, and in particular the type of asymptotics considered, may not be relevant or appropriate for all signal extraction problems connected with noisy data from chaotic dynamical systems. In various circumstances more or less will be known *a priori* about the dynamical system than we assume here. In many circumstances, inference about the dynamics F^t and/or the basic set Λ will be of greater importance than extraction of the signal x_n itself. Finally, when dealing with flows F^t rather than diffeomorphisms, the experimenter may sometimes be able to control the frequency of observation.

2. Background: Attractors, Hyperbolicity, and Axiom A

2.1. Invariant Sets and Attractors. The model for a smooth discrete-time dynamical system is a C^2 diffeomorphism [1] F of a phase space M, which, for simplicity, we take to be an open subset of $\mathbb{R}^d$. The *orbits* of the system are the (two-sided) sequences $\{x_n\}_{n\in\mathbb{Z}}$ such that $x_{n+1} = F(x_n)\ \forall\, n$. A compact subset Λ of the phase space will be called $F-$*invariant* if $F^{-1}(\Lambda) = \Lambda$, so that the restriction $F|\Lambda$ of F to Λ is a homeomorphism of Λ. Especially important among the invariant sets are *attractors*, which arise when the phase space contains a relatively compact open set Ω such that closure$(F\Omega) \subset \Omega$. If there exists such a set Ω, the set $\Lambda = \cap_{n>0} F^n\Omega$ is a nonempty $F-$invariant compact set, called an *attractor* for the diffeomorphism, and Ω is contained in its *basin of attraction.* All orbits $x_n = F^n(x_0)$ beginning at points $x_0 \in \Omega$ converge to Λ.

2.2. Example: Smale's Solenoid Mapping. The following example, Smale's *solenoid* mapping, shows that attractors may have a complex structure. The set Ω is a solid torus in $\mathbb{R}^3$ centered at the origin that may be parametrized by a real coordinate $\theta \in [0, 2\pi)$ and a complex coordinate $z \in \{|z| < 1\}$. (Picture the torus as a solid of revolution obtained by rotating the solid disc $\{|z| < 1\}$ once around the origin.) Fix $\alpha \in (0, \frac{1}{2})$, and define

$$F_\alpha(\theta, z) = (2\theta, \alpha z + e^{2i\theta}/2) \tag{3}$$

where 2θ is reduced mod 2π if $\theta \geq \pi$. In geometric terms, the mapping F_α is obtained as follows: (1) Cut the torus and unroll it to get a solid cylinder. (2) Stretch the cylinder lengthwise by a factor of two, then compress the resulting cylinder in the directions orthogonal to its length by a factor of α. (3) Wrap the resulting long, thin cylinder twice around the origin and place it so that it is entirely inside the original solid torus, and reattach the two ends. See Figure 1.

[1] A C^2 *diffeomorphism* is a bijective mapping F such that both F and F^{-1} are twice continuously differentiable.

Figure 1

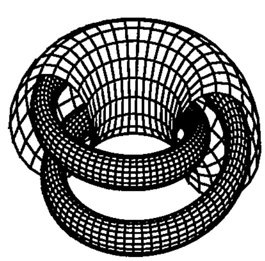
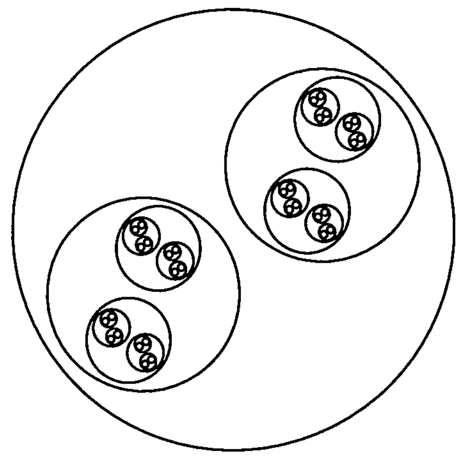


Figure 2



For each α the diffeomorphism F_α has an attractor $\Lambda \subset \Omega$ whose intersection with any "slice" $\Omega_\beta = \{(e^{i\theta}, z) : \theta = \beta\}$ is a Cantor set – see Figure 2. For each $\xi \in \Lambda \cap \Omega_\beta$ there is a smooth curve γ_ξ through ξ transverse to Ω_β that is contained in Λ. The homeomorphism $F_\alpha|\Lambda$ multiplies distances locally along each γ_ξ by 2, and multiplies distances in $\Omega_\beta \cap \Lambda$ by α. See Devaney (1986), section. 2.5, for helpful diagrams and further details.

2.3. Hyperbolicity and Orbit Separation. A compact invariant set Λ is called *hyperbolic* if at every point $\xi \in \Lambda$ the space of tangent vectors splits as a direct sum $E^u \oplus E^s$ of subspaces in such a way that for all $n \geq 1$,

$$\|DF^n v\| \geq c_u \lambda^n \|v\| \qquad \forall v \in E^u \tag{4}$$

$$\|DF^n v\| \leq c_s \lambda^n \|v\| \qquad \forall v \in E^s, \tag{5}$$

with suitable constants $0 < c_s, c_u < \infty$. The solenoid attractor is hyperbolic: for $\xi \in \Lambda$, E^s is the two-dimensional space of vectors pointing into the slice Ω_β containing ξ, and E^u is the one-dimensional space of vectors pointing in the direction of the curve γ_ξ. Hyperbolicity (together with compactness of the invariant set Λ and smoothness of F) implies that orbits of nearby points diverge rapidly. In particular, there exist constants $1 > \Delta > 0$ (which we shall call a *separation threshold*) and $C > 0$ such that if $0 < |x - x'| < \Delta$ for two points $x, x' \in \Lambda$ then

$$|F^n(x) - F^n(x')| > \Delta \qquad \text{for some } |n| \leq -C \log |x - x'|. \tag{6}$$

2.4. Axiom A Attractors. A compact hyperbolic invariant set Λ is called an *Axiom A basic set* if (i) periodic points are dense in Λ, and (ii) there exists $x \in \Lambda$ such that for every $m \geq 0$ the forward orbit $\{F^n(x)\}_{n \geq m}$ is dense in Λ. (NOTE: See [1] for the standard definition.) We shall only deal with Axiom A basic sets that are *topologically mixing*: This means that for any two (relatively) open sets U, V there exists an integer n_* such that for any $n \geq n_*$,

$$F^n(U) \cap V \neq \emptyset. \tag{7}$$

(By Smale's spectral decomposition theorem [[1], section 3.5], there is really no loss of generality in assuming that the basic set is topologically mixing.) It is not difficult to verify that the solenoid is a topologically mixing Axiom A attractor. Theoretical results (e.g., the

Structural Stability Theorem – see [10], Corollary) suggest that Axiom A basic sets occur commonly[2] in dynamical systems.

The ergodic theory of Axiom A basic sets and attractors is well understood – see [1] for a thorough exposition. Of special importance in the study of Axiom A *attractors* is the existence of a (unique) strongly mixing F–invariant probability measure μ_*, the so-called "*SRB* measure", that is supported by Λ and has the following property: for every continuous function $\varphi : \Omega \to \mathbb{R}$ and for a.e. $x \in \Omega$ (relative to Lebesgue measure on Ω),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} \varphi(F^k(x)) = \int \varphi \, d\mu_*. \tag{8}$$

It is the *SRB* measure that one would expect to "see" in time series data. For our purposes, the essential fact about the *SRB* measure is that it is a *Gibbs state* in the sense of [1], chapter 1. See section 7 below for the important facts about Gibbs states and more on the ergodic theory of Axiom A diffeomorphisms.

2.5. Lyapunov Exponents. The *Lyapunov exponents* measure the exponential rates at which orbits separate. For the solenoid mapping F_α there are two exponents, $\log 2$ and $\log \alpha$. In general, there are $l \le d$ Lyapunov exponents $\lambda_1 > \lambda_2 > \cdots > \lambda_l$. For μ_*–a.e. point $x \in \Lambda$ there are vector subspaces $E_1 \supset E_2 \supset \dots E_{l+1}$ of the space E of tangent vectors at x such that $E = E_1$, $E_{l+1} = \{0\}$, and for every $v \in E_j - E_{j+1}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \|DF^n v\| = \lambda_j. \tag{9}$$

This implies that the rate of separation of orbits will in general depend on the direction of the difference $x' - x$ in the initial points, a fact that may have important ramifications for the smoothing algorithm defined in section 3.2 below. See Eckmann and Ruelle (1986) for a detailed discussion of Lyapunov exponents.

More background on Axiom A diffeomorphisms, Gibbs states, and *SRB* measures, of a more technical nature, is given in the Appendix below. This additional material is needed for the proofs, but not the statements, of the results stated in the following section.

3. Signal Extraction

3.1. Bounded Noise. Consider now the problem of reconstructing an orbit $\{x_n\}$ from a noise-corrupted time series $y_n = x_n + e_n$. The sequence x_n is generated by (2), and we assume that the initial point x_0 is either an element of a (compact) hyperbolic invariant set or in the basin of attraction of a hyperbolic attractor. We first consider the problem of noise removal under the hypothesis that the noise is uniformly bounded:

Hypothesis 1. *Conditional on the sequence $\{x_n\}$ (equivalently, conditional on x_0) the random vectors e_n are independent, uniformly bounded by a constant $\delta > 0$, and have expectations*

$$E(e_n \mid x_0) = 0. \tag{10}$$

[2] whatever this means

3.1.1. *Smoothing Algorithm D.* This algorithm is designed for time series produced by a diffeomorphism (hence the D), with noise satisfying Hypothesis 1, and assumes that a suitable bound $\delta > 0$ for the noise is known *a priori.* The algorithm takes as input a finite sequence $\{y_n\}_{0\leq n\leq m}$ and produces as output a sequence $\{\hat{x}_n\}_{0\leq n\leq m}$ of the same length that will approximate the unobservable signal $\{x_n\}_{0\leq n\leq m}$. Let κ_m be an increasing sequence of integers such that

$$\lim_{m\to\infty} \kappa_m = \infty \quad \text{and} \quad \lim_{m\to\infty} \frac{\kappa_m}{\log m} = 0; \tag{11}$$

e.g., $\kappa_m = \log m / \log\log m$. For each integer $\kappa_m < n < m - \kappa_m$, define A_n to be the set of indices $\nu \in \{0, 1, \ldots, m\}$ such that

$$\max_{|j|\leq\kappa_m} |y_{\nu+j} - y_{n+j}| < 3\delta, \tag{12}$$

with the convention that $|y_j - y_i| = \infty$ if either of i or j is not in the range $[0, m]$. For $n \leq \kappa_m$ or $n \geq m - \kappa_m$, define A_n to be the singleton $\{n\}$. In rough terms, A_n consists of the indices of those points in the time series whose orbits "shadow" the orbit of x_n for κ_m time units. In Lemma 1 below we will show that $\nu \in A_n$ implies that $|x_\nu - x_n|$ is small. Thus, even though the values x_j are unobservable, neighboring points may still be picked out by virtue of having similar orbits. Now define

$$\hat{x}_n = \frac{1}{|A_n|} \sum_{\nu\in A_n} y_\nu. \tag{13}$$

Theorem 1. *Assume that x_0 is either an element of a compact hyperbolic invariant set Λ or an element of the basin of attraction of a compact hyperbolic attractor Λ, and assume that the noise sequence e_n satisfies Hypothesis1. Let Δ be a separation threshold for the invariant set. If $5\delta < \Delta$ then for every $\varepsilon > 0$,*

$$\lim_{m\to\infty} P\{m^{-1} \sum_{n=0}^{m} \mathbf{1}\{|\hat{x}_n - x_n| \geq \varepsilon\} \geq \varepsilon\} = 0. \tag{14}$$

Theorem 1 is valid for *every* orbit $\{x_n\}_{n\geq 0}$ contained in Λ, but the conclusion is only a weak convergence statement. For "generic" orbits of an Axiom A basic set, the conclusion can be strengthened to an a.s. convergence statement.

Theorem 2. *Assume that x_0 is chosen at random from a Gibbs state μ_* supported by an Axiom A basic set Λ, and assume that the noise sequence e_n satisfies Hypothesis 1. Let Δ be a separation threshold for the attractor. If $5\delta < \Delta$ then with probability one,*

$$\lim_{m\to\infty} \max_{\kappa_m < n < m - \kappa_m} |\hat{x}_n - x_n| = 0. \tag{15}$$

The most important case (probably) is when Λ is an Axiom A attractor and μ_* is the *SRB* measure. In practice, when dealing with an attractor, the initial point might be chosen at random from an absolutely continuous distribution on the basin of attraction Ω, and an initial segment of the orbit would then be discarded. Theorem 2 remains valid under this hypothesis.

Since the almost sure convergence statement (15) holds for points x_0 chosen at random from *any* F−invariant Gibbs state, and since Gibbs states are dense in the space of ergodic F−invariant probability measures on Λ, one might at first suspect that Theorems 1-2 might be strengthened to the stronger statement that (15) holds for *every* x_0 in Λ. This is false: it can be shown that *every* Axiom A basic set contains orbits for which (15) fails.

Theorems 1 and 2 will be proved in sections 5 and 6 below, respectively. In both cases, only the proofs for orbits x_n contained in Λ will be given, as the proofs for orbits initiated in the basin of attraction are nearly identical. The proof of Theorems 1 is relatively elementary, but that of Theorems 2 requires deeper results from the ergodic theory of Axiom A basic sets, which are collected in the Appendix.

3.1.2. *Implementation.* One might expect to use filters of the type described above on time series of length $m = 10^6$ or more, and so from a practical standpoint efficient *implementation* may be as important as *statistical* efficiency. Although implementation of Smoothing Algorithm D in the form described above may require on the order of $O(m^2)$ comparisons, there are simple modifications that can be implemented by $O(m \log m)$–step algorithms. In perhaps the simplest such modification, the indices $n \in [1, m]$ are sorted into bins B_v indexed by integer vectors v, with $n \in B_v$ if and only if v is the integer vector closest to $2x_n/3\delta$. The indices $n \in [1, m]$ are then re-sorted into bins B^*_w indexed by *arrays* w of integer vectors of length $2\kappa_m$, with $n \in B^*_w$ if and only if for each $|j| \leq \kappa_m$ the index $n + j \in B_{v_j}$, where v_j is the jth entry of w. The nth entry $\hat{x}_n$ of the smoothed sequence is then gotten by averaging the vectors y_ν over the indices ν in the bin B^*_w containing n.

3.1.3. *Consequences for Axiom A Attractors.* (A) By the Ergodic Theorem, it is almost surely the case that the empirical distribution of the points $x_1, x_2, \ldots, x_m$ converges weakly to the Gibbs state μ_*. Therefore, by Theorem 2, the empirical distribution of the points $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m$ converges weakly to μ_*.

(B) Since the *positive* Lyapunov exponents can be recovered from *non*-noisy data x_n (see [3] and [4]), it follows from Theorem 2 (and some auxiliary results) that the positive Lyapunov exponents can be estimated from noisy data y_n. Moreover, since the entropy is just the sum of the positive exponents, it too can be consistently estimated. Finally, since the *correlation dimension* of the *SRB* measure can be estimated from non-noisy data x_n, it can also be estimated from noisy data.

(C) Since F is continuous and the support of μ_* is dense in Λ, the set of ordered pairs $(\hat{x}_n, \hat{x}_{n+1})$, where $\kappa_m < n < m - \kappa_m$, converges in Hausdorff metric to the graph of $F|\Lambda$. Thus, one can in effect reconstruct the basic set Λ and the mapping $F|\Lambda$ from noisy data. However, it may *not* always be possible to recover all of the partial derivatives of F, as the support of the *SRB* measure may not fill up the "stable" directions E^s at $\xi \in \Lambda$. This was noted in [3].

3.1.4. *Second Stage Smoothing.* There is, obviously, a bias-variance tradeoff in the choice of the sequence κ_m used in the smoothing algorithm of section 3.2. Decreasing the rate of growth of κ_m increases the number of points in A_n, and therefore decreases the variance of the average (13); however, the values of x_ν included in the average will then tend to be further from x_n, therefore increasing the bias. But there is an even larger impediment to the accuracy of the algorithm that derives from the fact that the rate of orbital separation depends on the *direction* of the difference between initial points. In particular, the dynamical distance between orbit segments $\{F^\nu(x)\}_\nu$ and $\{F^\nu(x')\}_\nu$ will tend to be smaller when $x' - x$ points approximately in a "Lyapunov direction" corresponding to a smaller Lyapunov exponent. Thus, for most n it will be the case that the points $\{x_\nu\}_{\nu \in A_n}$ will lie in a (very) long, thin ellipsoid, and that many ν for which $|x_\nu - x_n|$ is relatively small will be excluded from A_n.

This peculiarity might, in principle, be exploited to obtain more accurate estimates of the points x_n. Fix $\beta \in (0, 1)$, and for each $1 \leq n \leq m$ let B_n be the set consisting of those m^β integers $\nu \in [1, m]$ for which $|\hat{x}_\nu - \hat{x}_n|$ is smallest. Define

$$\tilde{x}_n = m^{-\beta} \sum_{\nu \in B_n} y_\nu. \tag{16}$$

We conjecture that, with a suitable choice of β, use of this second stage filter might considerably improve the accuracy of estimation of x_n.

3.2. Gaussian Noise. Hypothesis 1 is quite a bit more stringent than one would like. However, if the errors are unbounded, even gaussian, then it is *impossible* to consistently reconstruct the signal x_n, or even a part of it, from a long stretch of the time series y_n. In fact, it is impossible to infer even a single value x_0 of the signal from the entire two-sided time series $y_n = x_n + e_n$.

Hypothesis 2. *Conditional on the sequence $\{x_n\}$ (equivalently, conditional on x_0) the random vectors e_n are independent and Gaussian with mean vector 0 and nonsingular covariance matrix Σ.*

Theorem 3. *Assume that x_0 is chosen at random from a Gibbs state μ_* supported by an Axiom A basic set Λ. If the errors e_n satisfy Hypothesis 2 then there is* no *measurable function $\xi_* = \xi_*(\{y_n\}_{n\in\mathbb{Z}})$ such that*

$$x_0 = \xi_* \qquad \text{with probability 1.} \tag{17}$$

The proof , which will be given in section 4 below, will show that orbit reconstruction is impossible even if the macroscopic features of the dynamics (the diffeomorphism F, the attractor Λ, and the *SRB* measure μ_*) are known *a priori.* Furthermore, it should be clear from the proof that the result extends to a large class of error distributions. We shall refrain, however, from trying to state and prove an extremely general form of the result.

Although it is not possible to consistently recover the signal $\{x_n\}$ from the time series y_n when the noise e_n is gaussian, it is nevertheless possible to consistently estimate important features of the dynamics provided the covariance matrix Σ is known. In particular, Birkhoff's ergodic theorem implies that for every polynomial $g(x)$ in d variables,

$$\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} g(y_i) = \int_{\mathbb{R}^d}\int_{\Lambda} g(\xi+\zeta)\, d\mu_*(\xi)\varphi_{0,\Sigma}(\zeta)\, d\zeta \tag{18}$$

where $\varphi_{0,\Sigma}$ is the gaussian density with parameters $0, \Sigma$. This implies that the moments of μ_* can be consistently estimated; since μ_* has compact support, it is determined by its moments, and so μ_* can be consistently estimated. Similarly, the joint distribution of $(X, F(X))$, where $X \sim \mu_*$, may also be consistently estimated. Since the support of this latter distribution is the graph of $F|\Lambda$, this too may be consistently estimated.

Unfortunately, proving the existence of consistent estimators is not the same as the construction of good or useful estimators. The substantial problem of inference about the dynamics of F from time series data $y_n = x_n + e_n$ when the noise e_n is gaussian will be left to another paper.

4. Proof of Theorem 3

Proof. The proof that there is no such ξ_* uses the existence of *homoclinic pairs* – see section 7.4 in the appendix below. By Proposition 2 of the appendix, on some probability space are defined random vectors x_0 and x'_0, each with marginal distribution μ_*, such that (a) with positive probability, $x'_0 \neq x_0$; and (b) with probability one x_0 and x'_0 constitute a homoclinic pair, i.e., for some $\alpha > 0$,

$$\lim_{|n|\to\infty} (1+\alpha)^{|n|}|x_n - x'_n| = 0, \tag{19}$$

where $x_n = F^n(x_0)$ and $x'_n = F^n(x'_0)$. We may assume that the probability space also accomodates a sequence e_n of gaussian random vectors that are jointly independent of x_0 and x'_0. Define $y_n = x_n + e_n$ and $y'_n = x'_n + e_n$; then conditional on the values of x_0 and

x'_0 the sequences $\mathbf{y} = \{y_n\}_{n\in\mathbb{Z}}$ and $\mathbf{y}' = \{y'_n\}_{n\in\mathbb{Z}}$ have gaussian distributions with the same autocovariance, and mean vector sequences $\{x_n\}_{n\in\mathbb{Z}}$, $\{x'_n\}_{n\in\mathbb{Z}}$ satisfying (19). Since (19) implies that $\sum_{n\in\mathbb{Z}} |x_n - x'_n|^2 < \infty$, a theorem of Kakutani (see, for instance, [6], section II.2, Theorem 2.1 and Example 3) implies that the conditional distributions of the sequences $\mathbf{y}$ and $\mathbf{y}'$, given x_0 and x'_0, are mutually absolutely continuous. Consequently, for any Borel measurable function $\xi_* : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}^d$, the conditional distributions of the random vectors $\xi_*(\mathbf{y})$ and $\xi_*(\mathbf{y}')$, given x_0 and x'_0, are also mutually absolutely continuous. If there were a function $\xi_* = \xi_*(\mathbf{y})$ such that $x_0 = \xi_*(\mathbf{y})$ almost surely, then it would also be the case that $x'_0 = \xi_*(\mathbf{y}')$ almost surely, and so the mutual absolute continuity of the conditional distributions would then imply that $x'_0 = x_0$ almost surely, a contradiction.

□

5. Proof of Theorem 1

In essence, the proof of Theorem 1 consists of showing (1) that the sets A_n are large (so that averaging over A_n will get rid of the errors); (2) that the sets A_n contain only indices ν such that $|x_n - x_\nu|$ is small; and (3) that although the sets A_n and the error random vectors e_ν are not *a priori* independent (since the sets A_n are defined using the values y_ν), the dependence may be circumvented in the averaging. It is only for task (2) that hyperbolicity of the invariant set Λ is needed.

Lemma 1. *There exists a constant $C > 0$ such that if $\nu \in A_n$ then*

$$|x_n - x_\nu| \leq \exp\{-\kappa_m/C + 2/C\}. \tag{20}$$

Proof. This is a consequence of the orbit separation property, which in turn follows from the hyperbolicity of Λ. By hypothesis, $5\delta < \Delta$, where Δ is a separation threshold for the attractor (see equation (6)), and by Hypothesis 1, $|e_n| < \delta$. Consequently, if $\nu \in A_n$ (ie.e, if inequality(12) holds), then

$$\max_{|j|\leq\kappa_m} |x_{n+j} - x_{\nu+j}| < 5\delta < \Delta.$$

But this cannot hold unless (20) is true, by the orbit separation property (6) (the constant C being the same as the constant C in (6)). Thus, $\nu \in A_n$ implies (20). □

Lemma 2. *For every $\varepsilon > 0$,*

$$\lim_{m\to\infty} \frac{1}{m} \sum_{n=0}^{m} \mathbf{1}\{|A_n| \leq m^{1-\varepsilon}\} = 0. \tag{21}$$

Proof. This follows from the hypothesis (11) that $\kappa_m = o(\log m)$ as $m \to \infty$, by a routine counting argument. Since Λ is compact, it has a finite subset B that is $\delta/2-$dense. Since $\kappa_m = o(\log m)$, the cardinality N_m of the set $B^{2\kappa_m+1}$ of length-$(2\kappa_m + 1)$ sequences with entries in B satisfies

$$N_m = o(m^\varepsilon) \qquad \text{as } m \to \infty \tag{22}$$

for every $\varepsilon > 0$. If B is $\delta/2-$dense in Λ, then for every $x \in \Lambda$, there is at least one element $\xi = (\xi_0, \xi_1 \dots, \xi_{2\kappa_m})$ of $B^{2\kappa_m+1}$ that $\delta/2-$shadows the orbit segment $\{F^n(x)\}_{-\kappa_m\leq n\leq\kappa_m}$, i.e., such that

$$|F^n(x) - \xi_{n-\kappa_m}| < \delta/2 \qquad \forall\, |n| \leq \kappa_m. \tag{23}$$

For each $\xi \in B^{2\kappa_m+1}$ define $B_m(\xi)$ to be the set of all indices $\nu \in \{0, 1, 2, \dots, m\}$ such that (23) holds with $x = x_\nu$. Every index ν is contained in at least one of the sets $B_m(\xi)$. If two indices n, ν both lie in the same set $B_m(\xi)$, then by (23) and the triangle inequality, $|x_{n+j} - x_{\nu+j}| < \delta$ and hence $|y_{n+j} - y_{\nu+j}| < 3\delta$ for all $|j| \leq \kappa_m$; thus, $\nu \in A_n$. Therefore, to prove (21) it suffices to show that for large m most of the indices ν lie in sets $B_m(\xi)$ with at least $m^{1-\varepsilon}$ elements. But by (22),

$$\sum_{\xi} |B_m(\xi)| \mathbf{1}\{|B_m(\xi)| \leq \sqrt{m/N_m}\} \leq \sqrt{mN_m} = o(m^{\frac{1}{2}+\varepsilon}) \tag{24}$$

Consequently, all but at most $o(m^{\frac{1}{2}+\varepsilon})$ of the indices $\nu \in \{0, 1, 2, \ldots, m\}$ lie in sets $B_m(\xi)$ with at least $m^{1-\varepsilon}$ elements. □

Proof of Theorem 1. The estimate $\hat{x}_n$ is obtained by averaging the vectors y_ν over the indices $\nu \in A_n$ (equation (13)). Since $y_\nu = x_\nu + e_\nu$, we have

$$\hat{x}_n = x_n + \frac{1}{|A_n|} \sum_{\nu \in A_n} e_\nu + \frac{1}{|A_n|} \sum_{\nu \in A_n} (x_\nu - x_n). \tag{25}$$

Lemma 1 implies that the latter average converges to zero uniformly for $\kappa_m < n < m - \kappa_m$ as $m \to \infty$. Thus, it suffices to show that for most of the indices n the average of the errors e_ν for $\nu \in A_n$ is small, with probability approaching 1 as $m \to \infty$. If the random vectors e_ν were independent of the index sets A_n then in view of Lemma 2 the result would follow immediately from the Chebyshev inequality. However, the random vectors e_ν are *not* independent of the index sets A_n; thus, some delicacy is required.

For each index n, define A_n^* to be the set of all indices ν such that $\nu \in A_n$ and $|n-\nu| \leq \kappa_m$; note that $|A_n^*|$ is no larger than $2\kappa_m + 1 = o(\log m)$, so on the event that $|A_n| > m^{3/4}$ the indices $\nu \in A_n^*$ have a negligible effect on the average $\sum_{\nu \in A_n} e_\nu / |A_n|$. For each index n and each integer $i \in [1, 2\kappa_m + 1]$, define A_n^i to be the set of all indices $\nu \notin A_n^*$ such that $\nu \in A_n$ and $\nu \equiv i \bmod 2\kappa_m + 1$. Obviously, the sets $A_n^*, A_n^1, A_n^2, \ldots, A_n^{\kappa_m}$ are pairwise disjoint, and

$$A_n = A_n^* \cup \left(\bigcup_{i=1}^{2\kappa_m+1} A_n^i \right). \tag{26}$$

For each i the set A_n^i is independent of the collection of random vectors $\{e_\nu\}$ indexed by integers $\nu \equiv i \bmod 2\kappa_m + 1$. To see this, consider an integer $\nu \equiv i \bmod 2\kappa_m + 1$. The event $\nu \in A_n^i$ is completely determined by the values of y_{n+j} and $y_{\nu+j}$ for $|j| \leq \kappa_m$; furthermore, no other event $\nu' \in A_n^i$, where $\nu' \neq \nu$, is influenced by the values of $y_{\nu+j}$ for $|j| \leq \kappa_m$ (this is the point of partitioning the indices ν into blocks of size $2\kappa_m + 1$). Moreover, the event $\nu \in A_n^i$ is not affected by the value of e_ν, because if $|y_{\nu+j} - y_{n+j}| < 3\delta$ for all $1 \leq |j| \leq \kappa_m$ then by the same argument as in the proof of Lemma 1, $|x_n - x_\nu| < \delta/2$ (provided m is large) and so $|y_n - y_\nu| < 3\delta$ *regardless* of the values of e_n and e_ν. Thus, the composition of the set A_n^i can be determined without reference to the values of the random vectors $\{e_\nu\}$ indexed by integers $\nu \equiv i \bmod 2\kappa_m + 1$.

For each index n, the sets A_n^i may be partitioned as $\mathcal{I} \cup \mathcal{J}$, where $\mathcal{I}$ consists of the special index $*$ and those indices i for which $|A_n^i| < \sqrt{m}$, and $\mathcal{J}$ consists of the remaining indices. For each $i \in \mathcal{J}$, Chebyshev's inequality implies that for any $\varepsilon > 0$,

$$P\left(|\sum_{\nu \in A_n^i} e_\nu| / |A_n^i| > \varepsilon \,\middle|\, A_n^i \right) \leq \delta^2 / |A_n^i| \varepsilon^2 \leq \delta^2 / (\sqrt{m}\varepsilon^2), \tag{27}$$

since the random vectors e_ν indexed by $\nu \in A_n^i$ are independent of A_n^i, by the preceding paragraph. Since there are no more than $2\kappa_m + 2$ elements of $\mathcal{I}$, and $|A_n^i| < \sqrt{m}$ for each $i \in \mathcal{I}$,

$$|\sum_{i \in \mathcal{I}} \sum_{\nu \in A_n^i} e_\nu| \leq (2\kappa + 2)\sqrt{m}\delta.$$

Consequently, if $|A_n| \geq m^{3/4}$ and m is sufficiently large that $(2\kappa_m + 2)/m^{1/4} < \varepsilon/\delta$ then the event $|\sum_{\nu \in A_n} e_\nu|/|A_n| > 2\varepsilon$ is contained in the union over $i \in \mathcal{J}$ of the events $|\sum_{\nu \in A_n^i} e_\nu| > \varepsilon |A_n^i|$. It therefore follows from inequality (27) that

$$P\left(|\sum_{\nu \in A_n} e_\nu|/|A_n| > 2\varepsilon \,\middle|\, |A_n| \geq m^{3/4} \right) \leq (2\kappa_m + 1)\delta^2/(\sqrt{m}\varepsilon^2).$$

Together with Lemma 2, this implies that

$$\sum_{n=0}^{m} P\{|\sum_{\nu \in A_n} e_\nu|/|A_n| > 2\varepsilon\} = o(m), \tag{28}$$

which, in view of Lemma 1, proves (14). □

6. Proof of Theorem 2

The proof of Theorem 2 differs from that of Theorem 1 in two respects: (A) Lemma 2 must be replaced by the stronger statement that the cardinality of A_n is large for *every* index n between κ_m and $m - \kappa_m$; and (B) Chebyshev's inequality must be replaced by an exponential large deviations probability inequality. The latter change is relatively minor; the former, however, requires hard results from the ergodic theory of Gibbs states on Axiom A basic sets. See the appendix below for a resume of the most important definitions and facts, and [1] for a detailed exposition of the theory.

Assume that Λ is an Axiom A basic set for F, that μ_* is a Gibbs state for F supported by Λ (see section 7.3 below for the definition and basic properties), and that the initial point x_0 of the orbit x_n is distributed in Λ according to μ_*.

Lemma 3. *For every $\varepsilon > 0$, all sufficiently large m, and all integers $n \in (\kappa_m, m - \kappa_m)$,*

$$P(|A_n| \leq m^{1-4\varepsilon}) \leq \exp\{-m^\varepsilon\} \tag{29}$$

Proof. The basic set Λ admits a *Markov partition* $\mathcal{M}$ of diameter less than δ (see section 7.2 below). Let $z_0, z_0' \in \Lambda$ be points with orbits $z_j = F^j(z_0)$ and $z_j' = F^j(z_0')$ and itineraries $\{i_j\}$ and $\{i_j'\}$ (relative to the Markov partition $\mathcal{M}$), respectively. If $i_j = i_j'$ for all $|j| \leq \kappa_m$ then $|z_j - z_j'| < \delta$ for all $|j| \leq \kappa_m$, since the diameters of the sets G_i of $\mathcal{M}$ are less than δ. Consequently, if x_n and x_ν are two points on the orbit of $x = x_0$ with itineraries $\{i_j\}, \{i_j'\}$ that coincide for $|j| \leq \kappa_m$, then $|y_{n+j} - y_{\nu+j}| < 3\delta$ for all $|j| \leq \kappa_m$, and so $\nu \in A_n$. Thus, to prove the inequality (29) it suffices to prove that for every finite itinerary $\mathbf{i} = \{i_j\}_{|j| \leq \kappa_m}$ of length $2\kappa_m + 1$, the probability that fewer than $m^{1-\varepsilon}$ of the points $\{x_n\}_{1 \leq n \leq m}$ share the itinerary $\mathbf{i}$ is smaller than $\exp\{-\sqrt{m}\}$.

Let I be the (doubly infinite) itinerary of a random point of Λ with distribution μ_*. Because μ_* is a Gibbs state, there exists a constant $\beta > 0$ and an integer L, both independent of m, such that the following is true (see inequalities (38) and (39) of the appendix below): For any infinite itinerary $\mathbf{i}$ and any finite itinerary $\mathbf{i}^*$ of length $2\kappa_m + 1$,

$$P(I_{L+n} = i_n^* \;\forall\, 1 \leq n \leq 2\kappa_m + 1 \mid I_n = i_n \;\forall n \leq 0) \geq \beta^{2\kappa_m + 1}. \tag{30}$$

Thus, if the random itinerary I is broken up into segments of length $L + 2\kappa_m + 1$, each segment will provide an opportunity for the letters $\mathbf{i}^*$ to occur with success probability at least $\beta^{2\kappa_m+1}$. Hence, if $N(\mathbf{i}^*)$ is the number of times that the finite string $\mathbf{i}^*$ occurs in the first m entries of I, then $N(\mathbf{i}^*)$ stochastically dominates the sum of $k = [[m/(L + 2\kappa_m + 1)]]$ i.i.d. Bernoulli random variables with success parameter $\beta^{2\kappa_m+1}$. Since $\kappa_m = o(\log m)$, for sufficiently large m this success probability is, for any $\varepsilon > 0$, eventually larger than $m^{-\varepsilon}$, and furthermore $k \geq m^{1-\varepsilon}$. It follows that the expectation of the sum is larger than $m^{1-2\varepsilon}$.

Consequently, by a very crude probability inequality for sums of independent Bernoulli random variables,

$$P\{N(\mathrm{i}^*) \le m^{1-4\varepsilon}\} \le \exp\{-m^{\varepsilon}\}. \tag{31}$$

□

Lemma 4. *With probability one,*

$$\lim_{m\to\infty} \max_{\kappa_m < n < m-\kappa_m} \frac{1}{|A_n|} \sum_{\nu \in A_n} e_\nu = 0. \tag{32}$$

Proof. The proof will use the following standard large deviations probability estimate for sums of independent random variables: If $\xi_1, \xi_2, \dots$ are independent random variables (or vectors) uniformly bounded by a constant $\delta < \infty$ and if $E\xi_j = 0$ for every j, then for every $\eta > 0$ there exists $\gamma = \gamma(\eta, \delta) > 0$ such that for all sufficiently large n,

$$P\{\frac{1}{n}|\sum_{j=1}^{n} \xi_j| \ge \eta\} \le \exp\{-n\gamma\}. \tag{33}$$

As in the proof of Theorem 1, the set A_n may be decomposed as the disjoint union of the sets A_n^* and A_n^i — see equation (26). Recall that for each i the set A_n^i is independent of the collection of random vectors $\{e_\nu\}$ indexed by integers $\nu \equiv i \bmod 2\kappa_m + 1$. Recall also that the indices $*, i$ may be partitioned as $\mathcal{I} \cup \mathcal{J}$, where $\mathcal{I}$ consists of the special index $*$ and those indices i for which $|A_n^i| < \sqrt{m}$, and $\mathcal{J}$ consists of the remaining indices. For each $i \in \mathcal{J}$, the probability inequality (33) implies that for any $\varepsilon > 0$ and all sufficiently large m,

$$P\left(\frac{1}{|A_n^i|}\Big|\sum_{\nu\in A_n^i} e_\nu\Big| > \varepsilon \;\Big|\; A_n^i\right) \le \exp\{-\gamma|A_n^i|\} \le \exp\{-\gamma\sqrt{m}\} \tag{34}$$

for a constant $\gamma > 0$ depending on ε and δ but not on m. Now for sufficiently large m,

$$\{\sum_{\nu\in A_n} e_\nu| > 2\varepsilon|A_n|\} \subset \{|A_n| \le m^{3/4}\} \cup \left(\bigcup_{i\in\mathcal{J}}\{|\sum_{\nu\in A_n^i} e_\nu| > \varepsilon|A_n^i|\}\right). \tag{35}$$

Consequently, by Lemma 3 and inequality (34), for all large m and $\kappa_m < n < m - \kappa_m$,

$$P\left(\frac{1}{|A_n|}\Big|\sum_{\nu\in A_n^i} e_\nu\Big| > 2\varepsilon\right) \le (2\kappa_m + 1)\exp\{-\gamma\sqrt{m}\} + \exp\{-m^{1/16}\}. \tag{36}$$

Since the series $\sum_m me^{-am^\alpha}$ is summable for any values of $a > 0$ and $\alpha > 0$, the result (32) follows from the Borel-Cantelli Lemma. □

7. Appendix: Markov Partitions for Axiom A Basic Sets

7.1. Example: Smale's Solenoid. In this example there is a simple Markov partition, and the resulting "symbolic dynamics" is relatively transparent. Partition the attractor Λ (or its basin of attraction Ω) into two sets

$$G_0 = \{(\theta, z) : 0 \le \theta \le \pi\}$$
$$G_1 = \{(\theta, z) : \pi \le \theta \le 2\pi\}$$

(This isn't really a partition in the usual sense of the word, since the sets have nonempty intersection, nor would Markov understand why his name is attached, but it is called a *Markov partition* anyway.) For any point $x \in \Lambda$, define an *itinerary* of x to be a doubly infinite sequence $\mathbf{i} = \{i_n\}_{n\in\mathbb{Z}}$ of 0s and 1s such that $F^n(x) \in G_{i_n}$ for each integer n. Observe that if $\mathbf{i}$ is an itinerary of $x = (\theta, z)$ then $i_0 i_1 i_2 \dots$ is a binary expansion of $\theta/2\pi$; moreover, if $x \in \Lambda_\beta$ for some particular cross-sectional slice Λ_β then the value of i_{-1} indicates which of the two "first generation" circles (see Figure 2) contains x, and $i_{-n}i_{-n+1}\dots i_{-1}$ determines which of the 2^n "nth generation" circles contains x . With this in mind, it is not difficult to see that (a) every infinite sequence of 0s and 1s is an itinerary of a unique $x \in \Lambda$; and (b) for μ_*–a.e. x there is only one itinerary. The projection from sequence space to Λ (semi-)conjugates the forward shift operator on sequence space to the solenoid mapping F_α. (In fact, Smale invented the solenoid mapping for just this reason.) See [2], chapter 2, for further details concerning this example.

7.2. Markov Partitions and Symbolic Dynamics. Every Axiom A basic set admits *Markov partitions* of arbitrarily small diameter, but in general neither the partitions nor their construction are simply described. See [1], chapter 3, or [10], chapter 10 for the precise definition and construction. A Markov partition $\mathcal{M}$ consists of finitely many closed sets $G_1, G_2, \dots, G_r$ whose union contains Λ, and such that for μ_*–a.e. z_0, every point $z_n = F^n(z_0)$ in the orbit of z_0 lies in only one of the sets G_i. The *diameter* of the partition is the maximum of the diameters of its constituent sets. For any point $z_0 \in \Lambda$, define an *itinerary* of z_0 to be a two-sided sequence $\dots i_{-1}i_0i_1\dots$ such that for each n, $z_n \in G_{i_n}$; note that for μ_*–a.e. z_0, there is only one itinerary. If the diameter of $\mathcal{M}$ is sufficiently small then no two distinct points $x, x' \in \Lambda$ may share the same itinerary, since this would entail a violation of the orbit separation property mentioned in section 2.3 above.

Let Σ be the space of all doubly infinite itineraries, and let σ be the forward shift operator on Σ. Since distinct points of Λ may not share the same itinerary, there is a projection $\pi : \Sigma \to \Lambda$ that maps each itinerary $\mathbf{i}$ to the unique point $x \in \Lambda$ with itinerary $\mathbf{i}$. It is not difficult to see that π is continuous (and even Hölder continuous with respect to the appropriate metric on Σ – see [1] or [10]). Clearly, $F\circ\pi = \pi\circ\sigma$, and so σ is a homeomorphism of Σ, since F is a homeomorphism of Λ. Not every sequence $\mathbf{i}$ need be an element of Σ; however, the Markov property of the partition $\mathcal{M}$ implies that the space Σ of all doubly infinite itineraries, together with the forward shift operator σ, is a topologically mixing *shift of finite type* (see [1], Lemma 1.3 and Proposition 3.19). A shift (Σ, σ) is of *finite type* if there exists a finite set $\mathcal{F}$ of finite words from the alphabet $\mathcal{A} = \{1, 2, \dots, r\}$ such that for any doubly infinite sequence $\mathbf{i}$ with entries in $\mathcal{A}$, $\mathbf{i}$ is an element of Σ if and only if $\mathbf{i}$ contains none of the words in $\mathcal{F}$. The shift (Σ, σ) is *topologically mixing* if there exists an integer $M < \infty$ such that for every pair $\omega, \omega' \in \Sigma$ there exists a finite word w of length M such that the concatenation

$$\dots\omega_{-2}\omega_{-1}\omega_0 w_1 w_2 \dots w_M \omega'_1\omega'_2\dots$$

is an element of Σ.

7.3. Gibbs States. A Gibbs state μ_* on Λ is defined to be an invariant probability measure whose pullback to a shift-invariant probability measure $\bar{\mu}_*$ on the sequence space Σ has the *Gibbs property* described in [1], chapter 1 (see [1], chapter 4 for the proof). In particular, $\bar{\mu}_*$ must satisfy a system of inequalities

$$C_1 \leq \frac{\bar{\mu}_*\{\mathbf{w} \in \Sigma : w_j = i_j \; \forall\, 0 \leq j \leq n\}}{\exp\{-\lambda n + \sum_{j=0}^{n} \varphi(\sigma^j \mathbf{i})\}} \leq C_2, \tag{37}$$

valid for all itineraries $\mathbf{i}$ and all integers $n \geq 0$, for constants $0 < C_1 < C_2 < \infty$ independent of n and of the itinerary $\mathbf{i}$. Here φ is a real-valued, Hölder continuous function on the space of all doubly infinite sequences $\mathbf{i}$, σ is the forward shift operator, and $\lambda \in \mathbb{R}$ is a constant called the *pressure*. See [1], section 1.4, for details. Note that (37) implies that there exists a constant $\beta > 0$ such that for any finite itinerary $i_1 i_2 \dots i_n$,

$$\bar{\mu}_*\{\mathbf{w} \in \Sigma : w_j = i_j \; \forall\, 1 \leq j \leq n\} \geq \beta^n \tag{38}$$

The *SRB* measure μ_* for an Axiom A attractor is a Gibbs state – see [1], chapter 4 for a proof. For Smale's solenoid mapping F_α the measure $\bar{\mu}_*$ is the product Bernoulli-1/2 measure, i.e., the measure that makes the coordinate random variables i.i.d. Bernoulli-1/2. In general, Gibbs states enjoy very strong mixing properties, among which the following, concerning the conditional distribution of the future given the past, is perhaps the most useful.

Proposition 1. *There exist constants* $\rho_k > 0$ *satisfying* $\lim_{k\to\infty} \rho_k = 1$ *and such that for every infinite itinerary* $\mathbf{i} = \dots i_{-1} i_0 i_1 \dots \in \Sigma$ *and every finite itinerary* $\mathbf{i}^* = i_1^* i_2^* \dots i_n^*$ *(of any positive length),*

$$\begin{aligned}(39)\qquad \bar{\mu}_*(w_{j+M+k} = i_j^* \;\forall\, 1 \le j \le n \mid w_j = i_j \;\forall\, j \le 0) \\ \ge \rho_k \bar{\mu}_*\{w_j = i_j^* \;\forall\, 1 \le j \le n\}.\end{aligned}$$

See [8] for a proof.

Equations (38)-(39) have the following consequence: there is a constant $\beta > 0$ such that for any finite itinerary $\mathbf{i}^* = i_1^* i_2^* \dots i_n^*$ (of any positive length), the conditional probability, given the past, that the next $M + n$ steps of the itinerary will end in $i_1^* i_2^* \dots i_n^*$ is at least β^n.

7.4. Homoclinic Pairs. One of the important features of Axiom A (and, more generally, hyperbolic) systems is the existence of *homoclinic* pairs. Two *distinct* points x and x' are said to be a homoclinic pair if for some $\varepsilon > 0$,

$$(40)\qquad \lim_{|n|\to\infty} (1+\varepsilon)^{|n|} |F^n(x) - F^n(x')| = 0;$$

in words, x, x' are distinct but their orbits approach each other exponentially fast both forwards and backwards in time. In Axiom A systems, homoclinic pairs are dense: in particular, for any points $\xi, \xi' \in \Lambda$ and any $\delta > 0$ there exists a homoclinic pair of points such that $|x - \xi| < \delta$ and $|x' - \xi'| < \delta$.

This may be proved using the existence of Markov partitions of small diameter. Let $\mathbf{i}$ and $\mathbf{i}'$ be itineraries of ξ and ξ', respectively. By the separation of orbits property, there exists an integer k such that if the itinerary $\mathbf{i}''$ of a point $x \in \Lambda$ satisfies $i_j'' = i_j$ for all $|j| \le k$, then $|x - \xi| < \delta$, and similarly, if $i_j'' = i_j'$ for all $|j| \le k$, then $|x - \xi'| < \delta$. But topological mixing (see section 7.2 above) guarantees that itineraries may be spliced together to obtain itineraries $\mathbf{i}^*$ and $\mathbf{i}^{**}$ so that (a) $i_j^* = i_j$ for all $|j| \le k$; (b) $i_j^{**} = i_j'$ for all $|j| \le k$; and (c) $i_j^* = i_j^{**}$ for all $|j| > M + k$. If x and x' have itineraries $\mathbf{i}^*$ and $\mathbf{i}^{**}$, respectively, then $|x - \xi| < \delta$ and $|x' - \xi'| < \delta$, by (a) and (b), and x, x' are a homoclinic pair, by (c) and the orbit separation property.

The foregoing argument may be adapted to prove the following proposition, which is the key to Theorem 3 above.

Proposition 2. *On some probability there exist random vectors* X', X'' *valued in* Λ *such that*

(a) *each of* X' *and* X'' *has marginal distribution* μ_*;
(b) *with probability 1,* X' *and* X'' *are a homoclinic pair; and*
(c) *with positive probability,* $X' \ne X''$.

Proof. The probability space should be large enough to accomodate a random vector X with distribution μ_* and several independent uniform-(0,1) random variables. Let $\mathbf{I} = \dots I_{-1} I_0 I_1 \dots$ be the itinerary of X. Construct new itineraries $\mathbf{I}', \mathbf{I}''$ as follows: For some large integer k, set $I_j' = I_j'' = I_j$ for all $|j| > k$; and choose the random vectors $(I_{-k}', \dots, I_k')$ and $(I_{-k}'', \dots, I_k'')$ independently from the conditional distribution of $(I_{-k}, \dots, I_k)$ given

$\{I_j\}_{|j|>k}$. (This is possible if the underlying probability space supports uniform random variables independent of $\mathbf{I}$.) By construction, each of $\mathbf{I}'$ and $\mathbf{I}''$ will be an itinerary. Define X' and X'' to be the unique points with itineraries $\mathbf{I}'$ and $\mathbf{I}''$, respectively. Clearly, each of X' and X'' has the same marginal distribution as X. Moreover, since the itineraries of X' and X'' coincide except in finitely many entries, X' and X'' must be a homoclinic pair. Finally, Proposition 1 implies that if k is large then the joint distribution of (X', X'') approximates the product measure $\mu_* \times \mu_*$. Since under $\mu_* \times \mu_*$ there is positive probability that the coordinates are not equal, the same is true for the joint distribution of (X', X''). $\square$

References

[1] R. Bowen (1975) Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. *Springer Lecture Notes in Math.* 470.

[2] R. Devaney (1986) *An Introduction to Chaotic Dynamical Systems.* Benjamin & Cummings, Menlo Park CA.

[3] J.-P. Eckmann & D. Ruelle (1985) Ergodic theory of chaos and strange attractors. *Rev. Modern Phys.* **57** 617

[4] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, & S. Ciliberto (1986) Lyapunov exponents from times series. *Phys. Rev. A* **34** 4971-4979

[5] E.Kostelich & J. Yorke (1990) Noise reduction – finding the simplest dynamical system consistent with the data. *Physica D* **41** 183-196

[6] H. H. Kuo *Gaussian Measures in Banach Spaces.* Springer Lecture Notes in Math. **463**

[7] S. Lalley (1997) More noise, more chaos. Manuscript.

[8] F. Ledrappier (1976) Sur la condition de Bernoulli faible et ses applications. *Théorie Ergodique.* Springer Lecture Notes in Math. **532** 152-159

[9] T. Sauer (1992) A noise reduction method for signals from nonlinear systems. *Physica D* **58** 193-201

[10] M. Shub (1987) *Global Stability of Dynamical Systems.* Springer-Verlag, New York.

Purdue University, Department of Statistics, West Lafayette IN 47907