Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Physica D 183 (2003) 260–272
PHYSICA D
www.elsevier.com/locate/physd

Filtering entropy☆

Wael Bahsoun*, Paweł Góra, Abraham Boyarsky, Mehran Ebrahimi

Department of Mathematics and Statistics, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec, Canada H4B 1R6

Received 10 October 2002; received in revised form 12 May 2003; accepted 29 May 2003
Communicated by C.K.R.T. Jones

Abstract

We consider a discrete time deterministic chaotic dynamical system: $x_{n+1} = T(x_n)$, where T is a nonlinear map of the unit interval into itself. The effect of noise contamination is modeled by $x_{n+1} = T(x_n) + \xi_n$, where ξ_n is an independent random variable with small noise level. We present a new theoretical result for estimating the metric entropy of T from the observed data of the noisy system. We introduce a skew product representation for the noisy system and then show that this skew product satisfies a formula similar to Adler's natural entropy formula.

MSC: Primary 37A35; 37H20

Keywords: Entropy; Noise; Skew product; Absolutely continuous invariant measure

1. Introduction

Metric entropy is an important measure of the chaos in a dynamical system. When dealing with a system modeled by a discrete time, nonlinear difference equation:

$$x_{n+1} = T(x_n), \tag{1.1}$$

the method described in [1] and implemented in [10] provides an algorithm for computing metric entropy. In [5] it is proved that if the dynamical system (1.1) possesses an ergodic absolutely continuous (with respect to Lebesgue measure) invariant measure μ, having positive metric entropy, then the metric entropy of (1.1) is equal to the Lyapunov exponent:

$$l = \int_0^1 \log|T'(x)|\,\mathrm{d}\mu.$$

☆ The research was supported by NSERC grant. WB is a recipient of Concordia University Graduate scholarship. ME is a recipient of Institut des Sciences Mathématiques (ISM) scholarship.
* Corresponding author.
E-mail addresses: wab@alcor.concordia.ca (W. Bahsoun), pgora@vax2.concordia.ca (P. Góra), boyar@vax2.concordia.ca (A. Boyarsky), mehra_eb@alcor.concordia.ca (M. Ebrahimi).

0167-2789/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0167-2789(03)00207-0

However, when the system is contaminated by noise

$$x_{n+1} = T(x_n) + \xi_n, \tag{1.2}$$

its entropy, in general, is infinite, while that of the underlying deterministic source system is finite. To extract the entropy of the dynamical system from the noisy data, computational techniques have been developed [1,10], but these are often difficult to implement and their accuracy cannot easily be verified. In [8], there is a conjecture that the metric entropy of the deterministic dynamical system is the difference between the entropy of the noisy system and the entropy of the noise itself. The basic assumption is that the large (often infinite) entropies of the noisy system and the noise itself will cancel to reveal the entropy of the chaotic system. However, there are no proofs in [8] and it is not clear under what conditions, if any, the method applies. Furthermore, the proposed method assumes a special structure for the noise. The main objective of this note is to propose a model in which this conjecture is true in a fairly general setting.

In Section 2, we state the notation and formulate the problem with the aid of a skew product representation. Section 3 contains the formula for the entropy of the skew product which in this special setting is similar to Adler's natural formula. In Section 4, using the entropy formula of Section 3, we show that the entropy of the chaotic map T can be filtered from the data of the noisy system. In Section 5, we present the method for computing the dynamical entropy from observed data. In Section 6, we present a computational example that verifies our method.

2. Notation and entropy of skew products

Definition 2.1. Let $(X, \mathfrak{A}, \sigma, \pi)$ be a dynamical system and let $(Z, \mathfrak{B}, \phi_x, \mu_x)_{x\in X}$ be a family of dynamical systems such that the function ϕ_x is $X\times Z$ measurable. A skew product of σ and $\{\phi_x\}_x$ is a transformation $F : X\times Z \to X\times Z$ defined by

$$F(x, z) = (\sigma(x), \phi_x(z)),$$

$x \in X, z \in Z$.

In [2], Adler remarked that the natural conjecture for the formula of the entropy of a skew product is

$$h(F, \nu) = h(\sigma, \pi) + \int_X h(\phi_x, \mu_x)\, \mathrm{d}\pi(x), \tag{2.1}$$

where $h(F, \nu)$, $h(\sigma, \pi)$ and $h(\phi_x, \mu_x)$ are the entropies of F, σ, ϕ_x, respectively.

We now define the skew product we will use in the sequel. In Section 3, we will prove that this skew product satisfies a formula similar to (2.1).

Definition 2.2. Let $([0, 1], \mathfrak{A}, \tau, \lambda)$ be an ergodic dynamical system, where λ is Lebesgue measure. Let $([0, 1], \mathfrak{B}, T, \mu)$ be an ergodic dynamical system. We assume that both τ and T are piecewise C^1. We consider the following perturbation of T:

$$T_x(z) = T(z) + g(x) \pmod 1,$$

where g is piecewise C^1 and $|g'| \le M$, $M < \infty$. Consider the family of dynamical systems, $([0, 1], \mathfrak{B}, T_x,)_{x\in[0,1]}$. We assume that $|\tau'| > \sup_x \sup_z |T_x'(z)|$. The skew product of τ and T_x is the transformation:

$$S : [0, 1] \times [0, 1] \to [0, 1] \times [0, 1]$$

defined by

$$S(x, z) = (\tau(x), T_x(z)).$$

We assume that S preserves the two-dimensional measure $\nu = \lambda \times \mu_R$ and that the system $([0, 1] \times [0, 1], \mathfrak{A} \times \mathfrak{B}, S, \nu)$ is ergodic.

Then, μ_R satisfies the equation $\mu_R(A) = \int_0^1 \mu_R(T_x^{-1}(A))\, d\lambda(x)$, for any measurable set $A \subset [0, 1]$. In particular, this holds if all T_x preserve the same measure μ_R. For the second iterate S^2, we have

$$S^2(x, z) = (\tau^2(x), T_{\tau(x)} \circ T_x(z)) = (\tau^2(x), T(T(z) + g(x)) + g(\tau(x)). \tag{2.2}$$

In general,

$$S^n(x, z) = (\tau^n(x), R^n(z)), \tag{2.3}$$

where

$$R^n(z) = T_{\tau^{n-1}(x)} \circ \cdots \circ T_{\tau(x)} \circ T_x(z).$$

The skew product S is our model for a random map

$$R = T_x(z) \quad \text{with probability } \lambda,$$

i.e., x is chosen according to Lebesgue measure. Below, we recall the notion of Lyapunov exponents and their relation to entropy.

Definition 2.3. Let $f : M \to M$ be an endomorphism on a manifold M of dimension m. Let $|\cdot|$ be the norm on tangent vectors induced by a Riemannian metric on M. For each $x \in M$ and $v \in T_x M$ let

$$l(x, v) = \lim_{k\to\infty} \frac{1}{k} \log(|Df_x^k v|), \tag{2.4}$$

whenever the limit exits.

Remark 2.4. The multiplicative ergodic theorem of Oseledec [9] says that, for almost all $x \in M$:

(i) the limit in (2.4) exists for all tangent vectors $v \in T_x M$, and
(ii) there are at most m distinct values of $l(x, v)$ for one point x.

Let $s(x)$ be the number of distinct values of $l(x, v)$ at x for $v \in T_x M$, with tangent vectors $v^j \in T_x M$ for $1 \le j \le s(x)$ giving distinct values:

$$l_j(x) = l(x, v^j)$$

with

$$l_1(x) < l_2(x) < \cdots < l_{s(x)}(x).$$

These distinct values are called the Lyapunov exponents at x.

Theorem 2.5 (Robinson [9]). *If the measure μ_R is absolutely continuous with respect to Lebesgue measure, then the measure theoretic entropy of S is then given by*

$$h(S,\nu)=\int_{[0,1]\times[0,1]}\chi^u(x,z)\,\mathrm{d}\nu, \tag{2.5}$$

where

$$\chi^u(x,z)=\sum_{j=r(x,z)+1}^{s(x,z)}k_j(x,z)l_j(x,z),$$

i.e., $\chi^u(x,z)$ is the sum of the positive Lyapunov exponents $l_j(x,z)$ with multiplicity $k_j(x,z)$.

Remark 2.6. Formula (2.5) is known as Pesin's formula (see [9]).

3. Entropy of skew product

The goal of this section is to show that

$$h(S,\nu)=h(\tau,\lambda)+\int_0^1\int_0^1\log|T_x'(z)|\,\mathrm{d}\mu_R(z)\,\mathrm{d}\lambda(x).$$

This formula have been proved before by other methods, see [3] or [6].

In a special case, when all T_x preserve the same measure μ_R, this reduces to

$$h(S,\nu)=h(\tau,\lambda)+\int_0^1 h(T_x,\mu_R)\,\mathrm{d}\lambda(x).$$

First, we observe that the derivative matrix of S is given by

$$A=\begin{pmatrix}\tau'(x) & 0\\ g'(x) & T'(z)\end{pmatrix}.$$

Lemma 3.1. *Let l_1 and l_2 be the Lyapunov exponents of S. Then, $l_1=h(\tau,\lambda)$ and $l_2=\int_0^1\int_0^1\log|T_x'(z)|\,\mathrm{d}\mu_R(z)$ $\mathrm{d}\lambda(x)$.*

Proof. Let

$$l_1=\lim_{k\to\infty}\frac{1}{k}\log|A(S^{k-1}(x,z))\cdots A(S(x,z))A(x,z)v_1|,$$

where $v_1=\begin{pmatrix}1\\0\end{pmatrix}$. We have

$$A(x,z)v_1=\begin{pmatrix}\tau'(x)\\ g'(x)\end{pmatrix},$$

$$A(S(x,z))A(x,z)v_1=\begin{pmatrix}\tau'(\tau(x))\tau'(x)\\ g'(\tau(x))\tau'(x)+g'(x)T'(T_x(z))\end{pmatrix}$$

and

$$A(S^2(x,z))A(S(x,z))A(x,z)v_1 = \begin{pmatrix} \tau'(\tau^2(x)\tau'(\tau(x))\tau'(x) \\ g'(\tau^2(x))\tau'(\tau(x))\tau'(x) + g'(x)T'(T_x(z))T'(T_{\tau(x)}(z)) + g'(\tau(x))\tau'(x)T'(T_{\tau(x)}(z)). \end{pmatrix}.$$

First, we estimate l_1 from below. Notice that

$$A(S^{k-1}(x,z))\cdots A(S(x,z))A(x,z)v_1 = \begin{pmatrix} \tau'(\tau^{k-1}(x))\cdots\tau'(\tau(x))\tau'(x) \\ \sharp \end{pmatrix},$$

where $\sharp$ is a number which we will estimate in the second part of the proof. It follows that

$$\|A(S^{k-1}(x,z))\cdots A(S(x,z))A(x,z)v_1\| \geq |\tau'(\tau^{k-1}(x))\cdots\tau'(\tau(x))\tau'(x)|.$$

Since log is an increasing function, we have

$$\log\|A(S^{k-1}(x,z))\cdots A(S(x,z))A(x,z)v_1\| \geq \log|\tau'(\tau^{k-1}(x))\cdots\tau'(\tau(x))\tau'(x)|.$$

Therefore,

$$l_1 \geq \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1}\log|\tau'(\tau^i(x))| = \int_0^1 \log|\tau'(x)|\,\mathrm{d}\lambda = h(\tau,\lambda). \tag{3.1}$$

Now, we estimate l_1 from above. For $k=2$:

$$|\sharp| = |g'(\tau(x))\tau'(x) + g'(x)T'(T_x(z))| \leq M|\tau'(x)| + M|T'(T_x(z))| \leq 2M|\tau'(x)|.$$

It can be seen that, for any k:

$$|\sharp| \leq Mk|\tau'(\tau^{k-1}(x))\cdots\tau'(\tau(x))\tau'(x)|.$$

Then, we have

$$\|A(S^{k-1}(x,z))\cdots A(S(x,z))A(x,z)v_1\| \leq k(M+1)|\tau'(\tau^{k-1}(x))\cdots\tau'(\tau(x))\tau'(x)|.$$

Since log is an increasing function, we get

$$\begin{aligned} l_1 &\leq \lim_{k\to\infty}\frac{1}{k}\log|(k(M+1))\tau'(\tau^{k-1}(x))\cdots\tau'(\tau(x))\tau'(x)| \\ &= \lim_{k\to\infty}\frac{\log(k(M+1))}{k} + \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1}\log|\tau'(\tau^i(x))| \\ &= \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1}\log|\tau'(\tau^i(x))| = \int_0^1\log|\tau'(x)|\,\mathrm{d}\lambda = h(\tau,\lambda). \end{aligned} \tag{3.2}$$

By (3.1) and (3.2) we get that $l_1 = h(\tau,\lambda)$. This proves the first part of the lemma. Now, let

$$l_2 = \lim_{k\to\infty}\frac{1}{k}\log\|A(S^{k-1}(x,z))\cdots A(S(x,z))A(x,z)v_2\|,$$

where $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We have

$$A(x,z)v_2 = \begin{pmatrix} 0 \\ T'(z) \end{pmatrix}$$

and

$$A(S(x,z))A(x,z)v_2 = \begin{pmatrix} 0 \\ T'(S(x,z))T'(z) \end{pmatrix}.$$

In general

$$A(S^k(x,z))\cdots A(S(x,z))A(x,z)v_2 = \begin{pmatrix} 0 \\ T'(S^k(x,z))\cdots T'(S(x,z))T'(z) \end{pmatrix}.$$

Therefore,

$$\begin{aligned} l_2 &= \lim_{k\to\infty}\frac{1}{k}\log|T'(S^k(x,z))\cdots T'(S(x,z))T'(z)| = \lim_{k\to\infty}\frac{1}{k}\sum_{i=0}^{k-1}\log|T'_x(S^i(z,x))| \\ &= \int_{[0,1]\times[0,1]}\log|T'_x(z)|\,\mathrm{d}\nu = \int_0^1\int_0^1\log|T'_x(z)|\,\mathrm{d}\mu_R(z)\,\mathrm{d}\lambda(x). \end{aligned} \tag{3.3}$$

□

Theorem 3.2. $h(S,\nu) = h(\tau,\lambda) + \int_0^1\int_0^1 \log|T'_x(z)|\,\mathrm{d}\mu_R(z)\,\mathrm{d}\lambda(x)$. *If all T_x preserve the same measure μ_R, then* $h(S,\nu) = h(\tau,\lambda) + \int_0^1 h(T_x,\mu_R)\,\mathrm{d}\lambda(x)$.

Proof. Using Pesin's formula and Lemma 3.1 we get

$$\begin{aligned} h(S,\nu) &= \int_{[0,1]\times[0,1]}\chi^u(x,z)\,\mathrm{d}\nu = \int_{[0,1]\times[0,1]}\left(h(\tau,\lambda) + \int_0^1\int_0^1\log|T'_x(z)|\,\mathrm{d}\mu_R(z)\,\mathrm{d}\lambda(x)\,\mathrm{d}\lambda(x)\right)\mathrm{d}\nu \\ &= h(\tau,\lambda) + \int_0^1\int_0^1\log|T'_x(z)|\,\mathrm{d}\mu_R(z)\,\mathrm{d}\lambda(x). \end{aligned} \tag{3.4}$$

□

4. Dynamical entropy estimation

In this section, we describe a method to filter the entropy of the chaotic map T from the total entropy of the noisy system. The second component of the skew product $S(x,z)$ is what we observe as a data contaminated with noise

$$S^n(x,z) = (\tau^n(x), T_{\tau^{n-1}(x)}\circ\cdots\circ T_{\tau(x)}\circ T_x(z)). \tag{4.1}$$

Let us define a new skew product whose second component represents only the noise.

Definition 4.1. Let $Q(x,z)$ be a transformation from $[0,1]\times[0,1]$ to itself given by

$$Q(x,z) = (\tau(x), I_x(z)),$$

where

$$I_x(z) = I(z) + g(x) \pmod 1$$

and I is the identity map from the unit interval [0, 1] to itself. Then

$$Q^n(x,z) = (\tau^n(x), I_{\tau^{n-1}(x)} \circ \cdots \circ I_{\tau(x)} \circ I_x(z)). \tag{4.2}$$

Corollary 4.2. $h(Q, \lambda \times \lambda) = h(\tau, \lambda)$

Proof. Using Theorem 3.2, we have

$$h(Q, \lambda \times \lambda) = h(\tau, \lambda) + \int_0^1 \int_0^1 \log|I'_x(z)|\, \mathrm{d}\lambda(z)\, \mathrm{d}\lambda(x). \tag{4.3}$$

Observe that $I'_x(z) = I'(z) = 1$. Thus, the double integral is equal to 0 and $h(Q, \lambda \times \lambda) = h(\tau, \lambda)$. □

Definition 4.3. We define the entropy of the noisy system by $h_{\text{total}} \equiv h(S, \nu)$, the entropy of the noise by $h_{\text{noise}} \equiv h(Q, \lambda \times \lambda)$ and the dynamical entropy as the difference between the total entropy of the noisy system and the entropy of the noise; i.e.

$$h_{\text{dyn}} = h_{\text{total}} - h_{\text{noise}} = h(S, \nu) - h(Q, \lambda \times \lambda) = \int_0^1 \int_0^1 \log|T'_x(z)|\, \mathrm{d}\mu_R(z)\, \mathrm{d}\lambda(x). \tag{4.4}$$

This definition of the dynamical entropy is based on the definition in [8]. Now, after two lemmas, we prove that the main conjecture of [8] holds for our model.

Lemma 4.4. *Let T be piecewise expanding, piecewise C^1 and $h(z) = 1/|(T(z))'|$ be a function of bounded variation. Then, for any $\alpha \in \mathbb{R}$, the map*

$$T_\alpha(z) = T(z) + \alpha \pmod 1$$

satisfies the same Lasota–Yorke estimate as T; i.e., let $\delta = \min_{i=1,\dots,q} \lambda(I_i)$, where $\mathcal{P} = \{I_i\}_{i=1}^q$ is the partition for which T is piecewise monotonic. Then for any $f \in BV_I$:

$$V_I P_{T_\alpha} f \le A V_I f + B||f||_1, \tag{4.5}$$

where $A = (\max_{1\le i\le q}(V_{I_i} h) + 2\sup_z h(z))$ and $B = (2(\sup_z h(z)/\delta) + (1/\delta)\max_{1\le i\le q} V_{I_i} h)$.

The proof of this lemma is in Appendix B.

Lemma 4.5. *Let T, T_x, μ, μ_R and $g(x)$ be as in Section 2. Suppose that T is piecewise expanding, piecewise C^1 and $h(z) = 1/|(T(z))'|$ is a function of bounded variation. Additionally, let us assume that μ is the unique absolutely continuous invariant measure of T. Then, $\mu_R \to \mu$ as $\sup_x |g(x)| \to 0$.*

Proof. The density of $\mu_{R,g}$, where index "g" shows dependence on the perturbation g, is a fixed point of operator $P_R(f) = \int_0^1 P_x(f)\, \mathrm{d}\lambda(x)$, where P_x is the Perron–Frobenius operator of T_x. By Lemma 4.4 all operators P_x satisfy the same Lasota–Yorke type estimates independently of the functions g. Thus, densities of $\mu_{R,g}$ have uniformly bounded variation and form a precompact set in $L^1([0,1],\lambda)$. Let g_n be a sequence of perturbations such that $\sup_x|g_n(x)| \to 0$, as $n \to +\infty$. The corresponding sequence of densities f_n of μ_{R,g_n} contains a convergent subsequence. We can assume it converges: $f_n \to f$ in $L^1([0,1],\lambda)$. We will show that f is the density of μ. It is enough to show that $P_T f = f$. Since $T_{x,n}$ (index "n" showing dependence on g_n) converges almost uniformly to T, as $n \to +\infty$, this can be proved in the same way as Lemma 11.2.2 of [4]. □

The following theorem generalizes the results of Corollary 2.1 of [7].

Theorem 4.6. *Let T, T_x, μ, μ_R and $g(x)$ be as in Section 2. Suppose that $T(z)$ is piecewise expanding, piecewise C^1 and $h(z) = 1/|(T(z))'|$ is a function of bounded variation. Additionally, let us assume that μ is the unique absolutely continuous invariant measure of T. Then*

$$h_{\rm dyn} \to h(T, \mu) \quad \textit{as} \ \sup_x |g(x)| \to 0.$$

Proof. First, observe that $T' = T'_x$ (derivative with respect to z) and

$$\mu_R \to \mu, \quad \text{as} \ \sup_x |g(x)| \to 0$$

in the sense of the convergence of densities in L^1 (Lemma 4.5), i.e.

$$\int_0^1 h \, \mathrm{d}\mu_R \to \int_0^1 h \, \mathrm{d}\mu, \quad \text{as} \ \sup_x |g(x)| \to 0 \tag{4.6}$$

for any h bounded on $[0, 1]$. Since $|T'| > 1$ (T is piecewise expanding), $\log|T'|$ is bounded. Therefore, by (4.6):

$$\int_0^1 \int_0^1 \log|T'_x(z)| \, \mathrm{d}\mu_R(z) \, \mathrm{d}\lambda(x) \to \int_0^1 \int_0^1 \log|T'(z)| \, \mathrm{d}\mu(z) \, \mathrm{d}\lambda(x) = h(T, \mu)$$

as $\sup_x |g(x)| \to 0$. This proves the theorem. It follows that, if $\sup_x |g(x)|$ is sufficiently small, then

$$h_{\rm dyn} \approx h(T, \mu).$$

□

Corollary 4.7. *Suppose that $T(z)$ is a piecewise monotonic map which is eventually expanding. Then*

$$h_{\rm dyn} \to h(T, \mu) \quad \textit{as} \ \sup_x |g(x)| \to 0.$$

Corollary 4.8. *Let $\bar{T} : [0, 1] \to [0, 1]$ differentiably conjugate to a piecewise expanding map T satisfying the assumptions of Theorem 4.6; i.e.*

$$\bar{T} = \phi \circ T \circ \phi^{-1},$$

where ϕ is a diffeomorphism. Let

$$\bar{T}_x(z) = \bar{T}(z) + \rho(x) \pmod 1.$$

Then

$$h_{\rm dyn}(\bar{T}) \to h(\bar{T}) \quad \textit{as} \ \sup_x |\rho(x)| \to 0.$$

Proof. Let $T_x(z) = T(z) + g(x) \pmod 1$ be such that

$$\phi \circ T_x \circ \phi^{-1}(z) = \bar{T}_x(z). \tag{4.7}$$

Observe that

$$\bar{T}_x \to \bar{T} \quad \text{as} \ \sup_x |\rho(x)| \to 0,$$

i.e.

$$\phi \circ T_x \circ \phi^{-1}(z) = \bar{T}_x(z) \to \bar{T} = \phi \circ T \circ \phi^{-1}(z) \quad \text{as } \sup_x |\rho(x)| \to 0$$

and

$$T_x \to T \quad \text{as } \sup_x |\rho(x)| \to 0.$$

But $T_x(z) = T(z) + g(x)$; therefore,

$$\sup_x |g(x)| \to 0 \quad \text{as } \sup_x |\rho(x)| \to 0.$$

It is known that (see [4]), if μ is T-invariant, then $\bar{\mu}$ is $\bar{T}$-invariant with $\bar{f} = (\phi \circ f)|(\phi^{-1}(z))'|$, where $\bar{f}$ and f are the densities of $\bar{\mu}$ and μ, respectively. Observe that

$$h(T, \mu) = h(\bar{T}, \bar{\mu}). \tag{4.8}$$

The density of the measure μ_R is a fixed point of the operator $\int_0^1 P_x(f)\,\mathrm{d}\lambda(x)$, where P_x is the Perron–Frobenius operator of T_x, and the density of the measure $\bar{\mu}_R$ is a fixed point of the operator $\int_0^1 \bar{P}_x(f)\,\mathrm{d}\lambda(x)$, where $\bar{P}_x$ is the Perron–Frobenius operator of $\bar{T}_x$. Therefore,

$$\int_0^1 \bar{P}_x(f)\,\mathrm{d}\lambda(x) = \int_0^1 \phi \circ P_x|(\phi^{-1}(z))'|(f)\,\mathrm{d}\lambda(x)$$

and $\bar{F} = (\phi \circ F)|(\phi^{-1}(z))'|$, where $\bar{F}$ and F are the densities of $\bar{\mu}_R$ and μ_R, respectively. Hence

$$h_{\text{dyn}}(T) = \int_0^1 \int_0^1 \log|T_x'|\,\mathrm{d}\mu_R\,\mathrm{d}\lambda(x) = \int_0^1 \int_0^1 \log|\bar{T}_x'|\,\mathrm{d}\bar{\mu}_R\,\mathrm{d}\lambda(x) = h_{\text{dyn}}(\bar{T}). \tag{4.9}$$

Now

$$\sup_x |\rho(x)| \to 0 \Rightarrow \sup_x |g(x)| \to 0;$$

by Theorem 4.6, this implies that

$$h_{\text{dyn}}(T) \to h(T, \mu)$$

and by (4.8) and (4.9)

$$h_{\text{dyn}}(\bar{T}) \to h(\bar{T}, \bar{\mu}).$$

It follows that, if $\sup_x |\rho(x)|$ is sufficiently small, then

$$h_{\text{dyn}}(\bar{T}) \approx h(\bar{T}, \bar{\mu}).$$

This proves the corollary. □

5. dynamical entropy computation

In general, the entropy of a noisy system and the entropy of the noise itself are very large. We now present examples where we compute the entropy of a chaotic map T and the dynamical entropy h_{dyn} of T contaminated by noise. Since, by Corollary 4.2, $h_{\text{noise}} = h(Q, \lambda \times \lambda) = h(\tau, \lambda)$, we calculate $h(\tau, \lambda)$ rather than $h(Q, \lambda \times \lambda)$. We use the algorithm of [10] to compute $h(S, \nu)$ and $h(\tau, \lambda)$ and then we find $h_{\text{dyn}} = h(S, \nu) - h(\tau, \lambda)$. To compute $h(S, \nu)$ and $h(\tau, \lambda)$, we consider the time series

$$w_{n+1} = T(w_n) + g(x_n) \pmod 1, \tag{5.1}$$

where

$$x_{n+1} = \tau(x_n). \tag{5.2}$$

$h(\tau, \lambda)$ is computed using time series (5.2), and to compute $h(S, \nu)$ we evaluate the entropy of the two-dimensional time series (x_n, w_n) as defined above. In our examples we verify that

$$h_{\text{dyn}} \approx h(T, \mu) \quad \text{as} \ \sup_x |g(x)| \to 0,$$

i.e., $h_{\text{dyn}} \approx h(T, \mu)$ if the effect of the noise on the data is sufficiently small. Note that in the following examples we could compute $h(S, \nu)$ and $h(\tau, \lambda)$ directly using the definitions of functions T and τ, but we actually use the time series (5.1) and (5.2) to compute the entropy. We chose this method since in real-world problems T and τ are not necessarily determined explicitly, and we only know the time series (5.1) and (5.2).

6. Example

In this section, we present an example in which we verify that

$$h_{\text{dyn}} \approx h(T, \mu) \quad \text{as} \ \sup_x |g(x)| \to 0,$$

i.e., $h_{\text{dyn}} \approx h(T, \mu)$ if the effect of the noise on the data is sufficiently small. We use the algorithm of [10] (see Appendix A) to compute the entropy of the noisy system and the entropy of the noise itself from the time series.

Example 6.1. We consider the piecewise linear transformation:

$$T(x) = 2x \pmod 1.$$

As the noise generator we use the map $x \mapsto 13x \pmod 1$, scaled and shifted to act on the interval $[-\epsilon/2, \epsilon/2]$:

$$\tau_\varepsilon(x) = \epsilon \left(\frac{13}{\epsilon} \left(x + \frac{\epsilon}{2} \right) \pmod 1 \right) - \frac{\epsilon}{2}.$$

Let $g(x) = x$. Using the algorithm introduced in [10], from time series (5.1) and (5.2) we compute values $h(S, \nu)$ and $h(\tau, \lambda)$, for different values of ϵ close to zero. Note that $-\epsilon/2 < x_n < \epsilon/2$ and if $\epsilon \to 0$, $\sup_x |g(x)| = \sup_x |x| \to 0$. Therefore for small values of ϵ we should have $h_{\text{dyn}} \approx \ln 2$. Table 1 shows the results of our computation for different values of ϵ. As expected (see Theorem 4.6) $h_{\text{dyn}} - h(T, \mu) \approx 0$ for small values of ε. The values of $h(S)$ and $h(\tau)$ in Table 1 are estimated by considering time series containing 100,000 iterations. We use initial partition size of 15 bins to compute $h(\tau, \lambda)$ in one dimension, and $15 \times 15 = 255$ bins of the initial partition size to compute $h(S)$ in the two-dimensional case, following the method of [10]. To evaluate each entropy value, we compute the slope of the best-fit line by using the first (at most 15) points (before the 20% separation limit is reached) on the plot of information against sequence length.

Table 1
Results for Example 6.1

ϵ	$h(S)$	$h(\tau)$	h_{dyn}	$h_{\text{dyn}} - h(T, \mu)$
0.2	3.465042157	2.564231195	0.900810962	0.2076637814
0.1	3.489686745	2.564231195	0.925455550	0.2323083694
0.06	3.453369315	2.564231195	0.889138120	0.1959909394
0.05	3.437043604	2.564231195	0.872812409	0.1796652284
0.005	3.265662893	2.564231195	0.701431698	0.0082845174
0.00001	3.229776885	2.564231195	0.665545690	−0.0276014906
0.0000001	3.229184915	2.564231195	0.664953720	−0.0281934606
0.000000001	3.228770468	2.564231195	0.664539273	−0.0286079076

We check the precision of our computations at $\epsilon = 0.000000001$, with respect to $\ln 2$. h_{dyn} calculated by this method is equal to 0.664539273. The percentage error is equal to

$$\left| \frac{0.664539273 - 0.6931471806}{0.6931471806} \right| = 0.04127248642 \approx 4\%. \tag{6.1}$$

Acknowledgements

The authors are deeply grateful to an anonymous reviewer for finding a serious error in a previous version of this paper. We would also like to remark that Proposition 3.6.5 in our book [4] is false.

Appendix A

We describe an algorithm developed in [10] which computes metric entropy of time series and which is used in Section 6. The approach is based directly on the definition of entropy. It includes methods of partitioning the data, computing sequential distributions, and compactifying results to reduce memory requirements. The underlying space (interval [0, 1] in our case) is partitioned into small rectangles called bins. Then, we compute the information:

$$I_n(B) = -\sum_{B_1, B_2, \ldots, B_n} \Pr(B_1, B_2, \ldots, B_n) \ln \Pr(B_1, B_2, \ldots, B_n),$$

where $\Pr(B_1, B_2, \ldots, B_n)$ is the probability that the dynamical system visits the sequence of boxes $B_1, B_2, \ldots, B_n$, at times $1, 2, \ldots, n$, respectively, and the sum is taken over all possible sequences of length n. Finally, the entropy of the system is estimated by the slope of the line asymptotic to the data on the graph of information I_n against sequence length n. Theoretically, for a infinite amount of data, such a line exists. In practice, with a finite amount of data, after a certain number of steps the information graph approaches the maximum information for the number of sequences under consideration. Therefore, we need to estimate the entropy by determining a linear region on the information graph. We say p% separation level has been reached when the number of realized sequences of length n is equal to p% of the total possible number of sequences of length n. Usually, one can find the best-fit line to the data of the information graph before 20% separation level is reached. In [10] algorithms for the one-dimensional case only were developed but one can easily apply the method for higher dimensions as we have done in our computations. It seems though that the entropy estimates in one dimension are more accurate than in higher dimensions.

Appendix B. Proof of Lemma 4.4

Proof. Let $a = x_0 < x_1 < \cdots < x_r = b$ be an arbitrary partition of I. Define $\phi_i = T_{\alpha,i}^{-1}$. We have

$$\begin{aligned} V_I P_{T_\alpha}(f)(x) &= \sum_{j=1}^{r} \left| \left(\sum_{i=1}^{q} h(\phi_i(x_j)) f(\phi_i(x_j)) \chi_{T_\alpha(I_i)}(x_j) - \sum_{i=1}^{q} h(\phi_i(x_{j-1})) f(\phi_i(x_{j-1})) \chi_{T_\alpha(I_i)}(x_{j-1}) \right) \right| \\ &\le \sum_{j=1}^{r} \sum_{i=1}^{q} |h(\phi_i(x_j)) f(\phi_i(x_j)) \chi_{T_\alpha(I_i)}(x_j) - h(\phi_i(x_{j-1})) f(\phi_i(x_{j-1})) \chi_{T_\alpha(I_i)}(x_{j-1})| \end{aligned} \tag{B.1}$$

We divide the sum on the right hand side into three parts:

(I) the summands for which $\chi_{T_\alpha(I_i)}(x_j) = \chi_{T_\alpha(I_i)}(x_{j-1}) = 1$,
(II) the summands for which $\chi_{T_\alpha(I_i)}(x_j) = 1$ and $\chi_{T_\alpha(I_i)}(x_{j-1}) = 0$,
(III) the summands for which $\chi_{T_\alpha(I_i)}(x_j) = 0$ and $\chi_{T_\alpha(I_i)}(x_{j-1}) = 1$.

First, we will estimate (I):

$$\begin{aligned} &\sum_{j=1}^{r} \sum_{i=1}^{q} |h(\phi_i(x_j)) f(\phi_i(x_j)) - h(\phi_i(x_{j-1})) f(\phi_i(x_{j-1}))| \\ &\quad \le \sum_{i=1}^{q} \sum_{j=1}^{r} |f(\phi_i(x_j))[h(\phi_i(x_j)) - h(\phi_i(x_{j-1}))]| + \sum_{i=1}^{q} \sum_{j=1}^{r} |h(\phi_i(x_{j-1}))[f(\phi_i(x_j)) - f(\phi_i(x_{j-1}))]| \\ &\quad \le \sum_{i=1}^{q} \left(\sup_{I_i} |f| V_{I_i} h \right) + \left(\sup_x h(x) \right) \sum_{i=1}^{q} V_{I_i} f \\ &\quad \le \max_{1\le i\le q} (V_{I_i} h) \sum_{i=1}^{q} \left(V_{I_i} |f| + \frac{1}{\lambda(I_i)} \int_{I_i} |f| \, \mathrm{d}\lambda \right) + \left(\sup_x h(x) \right) \sum_{i=1}^{q} V_{I_i} f \\ &\quad \le \max_{1\le i\le q} (V_{I_i} h) \left(V_I f + \frac{1}{\delta} \int_I |f| \lambda(\mathrm{d}x) \right) + \left(\sup_x h(x) \right) V_I f. \end{aligned} \tag{B.2}$$

We now consider the subsums (II) and (III) together. Notice that $\chi_{T_\alpha(I_i)}(x_j) = 1$ and $\chi_{T_\alpha(I_i)}(x_{j-1}) = 0$ occurs only if $x_j \in T_\alpha(I_i)$ and $x_{j-1} \notin T_\alpha(I_i)$. For this situation we have two cases:

(a) If T_α is monotonic on I_i, then it happens if x_j and x_{j-1} are on opposite sides of an end point of $T_\alpha(I_i)$, we can have at most one pair x_j, x_{j-1} like this and another pair $x'_j \notin T_\alpha(I_i)$ and $x_{j'-1} \in T_\alpha(I_i)$.

(b) If T_α is not monotonic on I_i, then T_α is monotonic on two subintervals of I_i, call them I_i^1 and I_i^2. Moreover, $T_\alpha(I_i^1) \cap T_\alpha(I_i^2) = \emptyset$. Then, the above situation happens if x_j and x_{j-1} are on opposite sides of an end point of $T_\alpha(I_i^1)$, we can have at most one pair x_j, x_{j-1} like this and another pair $x'_j \notin T_\alpha(I_i^2)$ and $x_{j'-1} \in T_\alpha(I_i^2)$. Thus, (II) and (III) can be estimated by

$$\sum_{i=1}^{q} (|h(\phi_i(x_j)) f(\phi_i(x_j))| + |h(\phi_i(x_{j'-1})) f(\phi_i(x_{j'-1}))|) \le \sup_x h(x) \sum_{i=1}^{q} (|f(\phi_i(x_j))| + f(\phi_i(x_{j'-1}))|). \tag{B.3}$$

Since $s_i = \phi_i(x_j)$ and $r_i = \phi_i(x_{j'-1})$ are both points in I_i, we can write

$$\sum_{i=1}^{q} (|f(s_i)| + |f(r_i)|) \le \sum_{i=1}^{q} (2|f(v_i)| + |f(v_i) - f(r_i)| + |f(v_i) - f(s_i)|),$$

where $v_i \in I_i$ is such a point that $|f(v_i)| \leq (1/\lambda(I_i)) \int_{I_i} |f|\lambda(\mathrm{d}x)$:

$$\sup_x h(x) \sum_{i=1}^{q} \left(V_{I_i} f + \frac{2}{\lambda(I_i)} \int_{I_i} |f|\lambda(\mathrm{d}x) \right) \leq \sup_x |h(x)| V_I f + \frac{2 \sup_x h(x)}{\delta} \int_I |f|\lambda(\mathrm{d}x). \tag{B.4}$$

Therefore,

$$V_I P_{T_\alpha}(f)(x) \leq \left(\max_{1\leq i\leq q} (V_{I_i} h) + 2 \sup_x h(x) \right) V_I f + \left(2 \frac{\sup_x h(x)}{\delta} + \frac{1}{\delta} \max_{1\leq i\leq q} V_{I_i} h \right) \int_I |f|\lambda(\mathrm{d}x). \tag{B.5}$$

□

References

[1] H.D.I. Abarbanel, Analysis of Observed Chaotic Data, Springer, Berlin, 1996.
[2] R. Adler, A note on the entropy of skew product transformations, Proc. Am. Math. Soc. 14 (1963) 665–669.
[3] Bogenschütz, Thomas; Crauel, Hans, The Abramov–Rokhlin formula, Ergodic theory and related topics, III (Güstrow, 1990), Lecture Notes in Mathematics, vol. 1514, Springer, Berlin, 1992, pp. 32–35.
[4] A. Boyarsky, P. Góra, Laws of Chaos, Birkhaüser, Boston, 1997.
[5] F. Ledrappier, Some properties of absolutely continuous invariant measures on an interval, Ergodic Theor. Dyn. Syst. 1 (1) (1981) 77–93.
[6] P.-D. Liu, Dynamics of random transformations: smooth ergodic theory, Ergodic Theor. Dyn. Syst. 21 (5) (2001) 1279–1319.
[7] P. Liu, Entropy formula of Pesin type for noninvertible random dynamical systems, Math. Z. 230 (2) (1999) 201–239.
[8] A. Ostruszka, P. Pakoński, W. Słomczyński, K. Życzkowski, Dynamical entropy for systems with stochastic perturbation, Phys. Rev. E 62 (2) (2000) 2018–2029.
[9] C. Robinson, Dynamical Systems, CRC Press, Boca Raton, FL, 1995.
[10] K.M. Short, Direct calculation of metric entropy from time series, J. Comput. Phys. 104 (1) (1993) 162–172.