Chapter 12
Computing Entropy

L'entropie est une loi générale de l'univers: la tendance naturelle des choses à passer de l'ordre au désordre sous l'effet d'un hasard calculable.

F. Jacob

12.1 Introduction

In general, computing the entropy of a measure-preserving transformation is a delicate problem. We will study a class of noninvertible maps for which the computation does not present too many problems.

A map T on a metric space X is called *uniformly dilating* if there exists a finite partition $\{X_i\}$ of X such that the restriction of T to each of the X_i dilates the metric in the following sense: there exists a constant $K > 1$ such that for every i,

$$\forall x,\ y \in X_i, \quad d\big(T(x), T(y)\big) \geqslant K\, d(x, y).$$

We have already come across several examples of dilating maps: the toral automorphisms whose eigenvalues all have absolute value greater than 1 and the piecewise C^1 maps on the interval whose derivatives are greater than some constant $K > 1$ satisfy this property. A few dilating transformations on the interval are shown in Fig. 12.1.

For these maps, we can give an explicit expression for the entropy of invariant probability measures. This expression uses the dilation factor of the measure under the action of the transformation. From an informal point of view, we could say that the inherent randomness of the system is proportional to this dilation factor.

This factor is easily computed when the transformation is regular and preserves an invariant measure that is absolutely continuous with respect to the Lebesgue measure. It can be obtained through a simple change of variables. The entropy then equals the integral of the logarithm of the Jacobian of the transformation. We thus link a measurable quantity, defined globally and measuring the uncertainty in the evolution of the system over time, to a quantity obtained by averaging the infinitesimal dilation observed in the neighborhood of each point of the space.

Y. Coudène, *Ergodic Theory and Dynamical Systems*, Universitext,
DOI 10.1007/978-1-4471-7287-1_12

The Bernoulli shifts defined on a finite alphabet I can be interpreted in terms of a dilating map. More generally, we can introduce a distance on the set of admissible sequences of a Markov chain with finite state space, for which the shift is dilating. It then becomes possible to compute the shift's entropy by evaluating its Jacobian. This computation method is not the most elementary one, but it illustrates well the concept of a dilating map.

12.2 The Rokhlin Formula

Let $(X, \mathcal{T}, \mu)$ be a probability space, and let $T : X \to X$ be a measurable map that preserves the measure μ. Suppose that T restricted to a measurable set $A \subset X$ is injective. We define the inverse of the Jacobian of T on A by the formula

$$\frac{1}{|T'_\mu|} = \frac{\mathrm{d}}{\mathrm{d}\mu}\big(T_*(\mu_{|A})\big) \circ T,$$

so that we have the usual change of variables:

$$\int_A g \circ T \, \mathrm{d}\mu = \int_{TA} \frac{g}{|T'_\mu| \circ T^{-1}_{|A}} \, \mathrm{d}\mu.$$

Proposition 12.1 *Let $(X, \mathcal{T}, \mu)$ be a probability space, and let $T : X \to X$ be a measurable map that preserves the measure μ. Suppose that there exists a finite partition X_i, for $1 \leqslant i \leqslant k$, such that the TX_i are measurable and $T : X_i \to TX_i$ is bijective with measurable inverse. Then*

$$\text{for almost all } x_0 \in X, \quad E(f \mid T^{-1}\mathcal{T})(x_0) = \sum_{Tx = Tx_0} \frac{f(x)}{|T'_\mu(x)|},$$

$$H(\{X_i\}_{i=1,\dots,k} \mid T^{-1}\mathcal{T}) = \int_X \log |T'_\mu| \, \mathrm{d}\mu.$$

Proof Denote by $T_i^{-1} : TX_i \to X_i$ the inverse of T on the domains in question; we extend this function to X arbitrarily. Note that for all $x_0 \in X_i$, we have

$$\{x \mid Tx = Tx_0\} = \{T_i^{-1}(Tx_0) \mid i \text{ such that } T(x_0) \in T(X_i)\}.$$

To compute the conditional expectation, we write $f = \sum_i \mathbf{1}_{X_i} f \circ T_i^{-1} \circ T$, which gives

$$E(f \mid T^{-1}\mathcal{T}) = \sum_i E(\mathbf{1}_{X_i} \mid T^{-1}\mathcal{T}) f \circ T_i^{-1} \circ T.$$

It therefore suffices to prove the equality $E(\mathbf{1}_{X_i} \mid T^{-1}\mathcal{T}) = \dfrac{\mathbf{1}_{TX_i} \circ T}{|T'_\mu| \circ T_i^{-1} \circ T}$:

$$\begin{aligned} \int g \circ T\, E(\mathbf{1}_{X_i} \mid T^{-1}\mathcal{T})\, \mathrm{d}\mu &= \int_{X_i} g \circ T\, \mathrm{d}\mu \\ &= \int \frac{\mathbf{1}_{TX_i}\, g}{|T'_\mu| \circ T_i^{-1}}\, \mathrm{d}\mu \\ &= \int \frac{\mathbf{1}_{TX_i} \circ T\, g \circ T}{|T'_\mu| \circ T_i^{-1} \circ T}\, \mathrm{d}\mu. \end{aligned}$$

It remains to compute the entropy of the partition $\{X_i\}$:

$$\begin{aligned} H(\{X_i\} \mid T^{-1}\mathcal{T}) &= \sum_i \int -\mathbf{1}_{X_i} \log E(\mathbf{1}_{X_i} \mid T^{-1}\mathcal{T})\, \mathrm{d}\mu \\ &= \sum_i \int_{X_i} \log |T'_\mu| \circ T_i^{-1} \circ T\, \mathrm{d}\mu \\ &= \sum_i \int_{X_i} \log |T'_\mu|\, \mathrm{d}\mu \\ &= \int_X \log |T'_\mu|\, \mathrm{d}\mu. \end{aligned}$$ □

The following formula, due to V. Rokhlin, will allow us to calculate the entropy of maps that are piecewise dilating.

Corollary 12.1 *The entropy is bounded from below by the integral of the Jacobian:*

$$h_\mu(T) \geqslant \int_X \log |T'_\mu|\, \mathrm{d}\mu,$$

with equality if the partition $\{X_i\}$ is a one-sided generator.

Remarks

- Suppose that we have $X \subset \mathbf{R}^n$, that μ is the Lebesgue measure, that T is C^1 on the interior of X_i (Lipschitz suffices), and that $\mu(\partial X_i) = 0$. Then $|T'_\mu| = |\det(DT)|$ almost everywhere.
- The quantity $E(\mathbf{1}_{X_i} \mid T^{-1}\mathcal{T})(x)$ can be seen as the probability that x is in X_i, given the value of $T(x)$.
- If T has a one-sided generator, then $h(T)$ can be seen as the average amount of information needed to know x, given that Tx is known.
- The operator $L_\mu f(y) = \sum_{Tx=y} f(x)/|T'_\mu(x)|$ is the *transfer operator* associated with T; it is the adjoint of the isometry $f \mapsto f \circ T$ defined on $L^2(X, \mu)$.

12.3 Entropy of Shifts

Let us compute the entropy of the shift on a Markov chain using the observations above. Let I be a finite alphabet. Let $X = I^{\mathbf{N}}$, and consider the shift on X given by $T(\{x_i\}_{i\in\mathbf{N}}) = \{x_{i+1}\}_{i\in\mathbf{N}}$.

We define a measure on X using a transition matrix $\{p_{i,j}\}$. For every $i, j \in I$, we take real numbers $p_{ij} \in [0,1]$ and $p_i \in [0,1]$ satisfying

$$\sum_i p_i = 1, \quad \sum_j p_{ij} = 1, \quad \text{and} \quad \sum_i p_i p_{ij} = p_j.$$

Let $a_1, \ldots, a_n$ be elements of I. Recall that the cylinder set $[a_1, \ldots, a_n] \subset X$ consists of the elements of X that begin with the sequence $a_1, \ldots, a_n$. By the Kolmogorov extension theorem, there exists a probability measure μ on X that satisfies

$$\mu([a_1, a_2, \ldots, a_n]) = p_{a_1} p_{a_1 a_2} \cdots p_{a_{n-1} a_n}.$$

This measure is T-invariant. The entropy of the transformation T relative to this measure is given by the following proposition.

Proposition 12.2 $h_\mu(T) = -\sum_{i,j} p_i\, p_{ij} \, \log p_{ij}$.

Proof Let N be the cardinality of I. Each point of X has exactly N preimages and T is bijective from $[i]$ to X. We therefore take $X_i = [i]$. The partition $\{[i] \mid i \in I\}$ is a generator because the partition generated by its first k inverse images under T consists of all cylinder sets of length k. Let us calculate the Jacobian of $T : [i] \to X$ restricted to the cylinder set $[i, j]$, when it has nonzero measure:

$$\mu(T^{-1}[j, a_1, \ldots, a_n] \cap [i]) = \mu([i, j, a_1, \ldots, a_n]) = \frac{p_i p_{ij}}{p_j}\, \mu([j, a_1, \ldots, a_n]).$$

This shows that $|T'_\mu|$ is constant on the cylinder set $[i, j]$, with value $p_j / p_i p_{ij}$. Finally,

$$h_\mu(T) = -\sum_{i,j} p_i\, p_{ij} \log(p_{ij}) - \sum_i p_i \log(p_i)\Big(\sum_j p_{ij}\Big) + \sum_j \log(p_j)\Big(\sum_i p_i p_{ij}\Big).$$

The last two terms cancel each other out. ◻

The space X can be endowed with the following distance:

$$d\big(\{x_i\}, \{y_i\}\big) = 2^{-\min\{j \in \mathbf{N} \mid x_j \neq y_j\}}.$$

The shift is dilating with respect to this distance when restricted to the cylinder sets $[a] = \{\, \{x_i\} \in I^{\mathbf{N}} \mid x_0 = a\}$, for every $a \in I$. This follows from the relation

$$\forall x, y \in [a], \quad d\big(Tx, Ty\big) \geqslant 2\, d(x, y).$$

The dynamical system we have just studied is therefore an example of a piecewise dilating map.

12.4 Entropy of Dilating Transformations

The trajectories of a dilating map T tend to separate over time, as illustrated by Fig. 12.2. This allows us to show that a partition $\{X_i\}$ is generating if the restriction of T to each piece of the partition is dilating.

Proposition 12.3 *Let X be a metric space, let μ be a Borel probability measure, and let $T : X \to X$ be a Borel map that preserves the measure μ. Let ξ be a finite partition whose elements have finite diameters; we suppose that for some $K > 1$ and for every $A \in \xi$,*

$$\forall x,\ y \in A, \quad d(Tx, Ty) \geqslant K\, d(x, y).$$

Then ξ is a one-sided generator.

Proof Let us show that the diameters of the sets $\big(\overset{n}{\underset{0}{\vee}} T^{-i}\xi\big)(x)$ tend to 0 when n tends to ∞.

If this is not the case, we can find $\delta > 0$ and, for every integer $n \in \mathbf{N}$, points x_n in $\overset{n}{\underset{0}{\vee}} T^{-i}\xi\ (x)$ such that $d(x_n, x) > \delta$. For every $i \in \{0, \dots, n\}$, the point $T^i x_n$ belongs to $\xi(T^i x)$, which gives

$$d(T^{i+1}x_n, T^{i+1}x) \geqslant K\, d(T^i x_n, T^i x).$$

Consequently, we have

$$\operatorname{diam} \xi(T^n x) \geqslant d(T^n x_n, T^n x) \geqslant K^n d(x_n, x) \geqslant K^n\, \delta.$$

The diameter of the $\xi(T^n x)$ is therefore not bounded. We conclude using the following lemma. ☐

Lemma 12.1 *Let X be a metric space, and let ξ_n be a sequence of countable partitions that satisfies, for every $x \in X$, $\operatorname{diam} \xi_n(x) \to 0$. Then the elements of the ξ_n for $n \in \mathbf{N}$ generate the Borel σ-algebra of X.*

Proof Let U be an open subspace of X. For every $x \in U$, there exists $n \in \mathbf{N}$ such that $\xi_n(x) \subset U$. We therefore have

$$U = \bigcup_{n\in\mathbf{N}} \bigcup_{\substack{A\in\xi_n \\ \text{and } A\subset U}} A.$$

The open subspace U is in the σ-algebra generated by the ξ_n. The proof is illustrated by Fig. 12.3. □

We have shown that the entropy of a piecewise dilating map can be obtained by integrating the Jacobian.

Theorem 12.1 *Let X be a metric space, and let $T : X \to X$ be a uniformly piecewise dilating Borel map: there exist a finite partition $\{X_i\}$ of X by bounded Borel sets and a constant $K > 1$ such that*

$$\forall i,\ \forall x,\ y \in X_i, \quad d\big(T(x), T(y)\big) \geqslant K\, d(x, y).$$

Let μ be a Borel probability measure that is invariant under T. Then the entropy of T relative to μ is given by

$$h_\mu(T) = \int_X \log |T'_\mu| \, \mathrm{d}\mu.$$

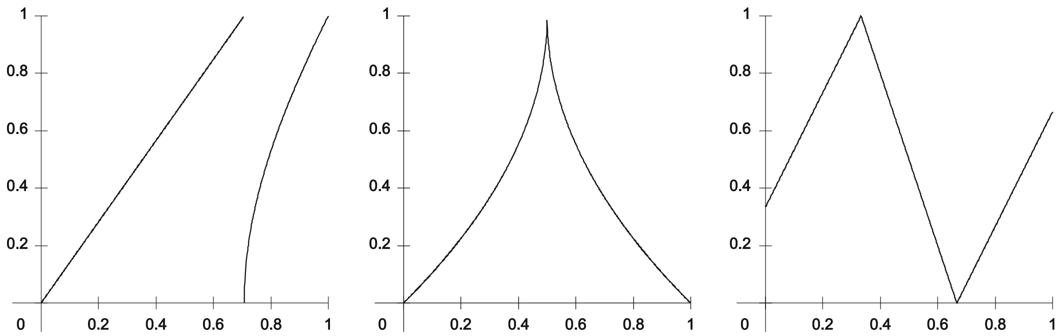

Fig. 12.1 A few dilating maps on the interval [0, 1]

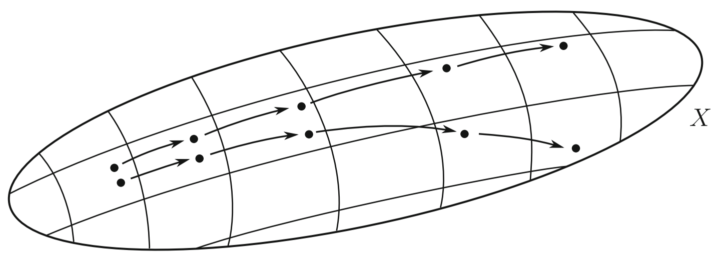

Fig. 12.2 Dilation and generating partition

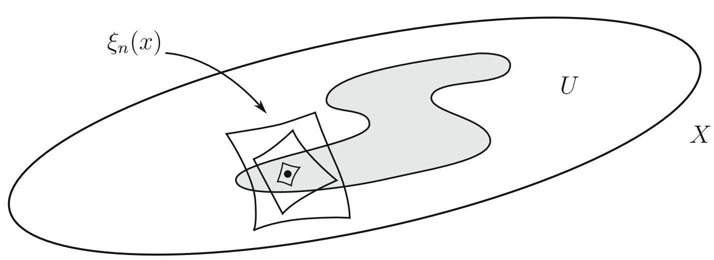

Fig. 12.3 Sequence of partitions with arbitrarily small diameter

12.5 Exercises

12.5.1 Basic Exercises

***Exercise* 1** Consider the map $T : [0, 1] \to [0, 1]$ given by

$$T(x) = \begin{cases} \sqrt{2}\,x & \text{if } x \in [0, 1/\sqrt{2}], \\ \sqrt{2x^2 - 1} & \text{if } x \in [1/\sqrt{2}, 1]. \end{cases}$$

Show that T preserves the measure $2x\mathrm{d}x$ and determine its entropy.

***Exercise* 2** Let A be an $n \times n$ matrix with integral coefficients and nonzero determinant, whose eigenvalues all have absolute value greater than 1. This matrix induces a map on the torus $\mathbf{T}^n$ by passing to the quotient. Show that this map preserves the Lebesgue measure, and then that it is injective and dilating on every set with sufficiently small diameter. Determine its entropy relative to the Lebesgue measure on the torus.

***Exercise* 3** Let U be an open subset of $\mathbf{R}^n$, and let $T : U \to \mathbf{R}^n$ be an injective C^1 map preserving a finite measure of the form $\mathrm{d}\mu = h\,\mathrm{d}x$, with h measurable. Compute $|T'_\mu|$.

Suppose that the function $\log(h)$ is μ-integrable. Prove the following formula:

$$\int_U \log|T'_\mu|\,\mathrm{d}\mu = \int_U \log|\det D_x T|\,\mathrm{d}\mu.$$

***Exercise* 4** Let X be a compact metric space, let $T : X \to X$ be a continuous map, and let μ be a Borel probability measure that is invariant under T. Let ξ be a finite partition satisfying the following property:

$$\forall A \in \xi,\ \forall x, y \in \overline{A} \text{ distinct}, \quad d(Tx, Ty) > d(x, y).$$

Show that the partition ξ is generating.

***Exercise* 5** Show that the following map $T : [0, 1] \to [0, 1]$ preserves the Lebesgue measure:

$$T(x) = \begin{cases} \sqrt{|2x - 1|} & \text{if } x \in [0, \frac{1}{2}], \\ 1 - \sqrt{|2x - 1|} & \text{if } x \in (\frac{1}{2}, 1]. \end{cases}$$

Compute its entropy.

12.5.2 More Advanced Exercises

***Exercise* 6** Let $T : [0,1] \to [0,1]$ be the map defined by

$$T(x) = \begin{cases} 2x + 1/3 & \text{if } x \in [0, \frac{1}{3}], \\ -3x + 2 & \text{if } x \in [\frac{1}{3}, \frac{2}{3}], \\ 2x - 4/3 & \text{if } x \in [\frac{2}{3}, 1]. \end{cases}$$

Show that T preserves a probability measure that is absolutely continuous with respect to the Lebesgue measure. Compute the entropy of T with respect to this measure. Does there exist a generating partition with two elements?

***Exercise* 7** Let $(X, \mathcal{T}, \mu)$ be a probability space, and let $T : X \to X$ be a measurable map that preserves μ. We suppose that there exists a finite partition $\{X_i\}$ such that the TX_i are measurable and that the restriction of T to each of the X_i is a bijective, bimeasurable map from X_i to TX_i. Prove the equality

$$\int_X |T'_\mu| \,\mathrm{d}\mu = \int_X \operatorname{Card}\bigl(T^{-1}(y)\bigr)\,\mathrm{d}\mu(y).$$

***Exercise* 8** Let $p, q \in [0,1]$ satisfy $p + q = 1$. Compute the Jacobian of the unilateral shift associated with the Markov chain with transition matrix $\left(\begin{smallmatrix} p & q \\ p & q \end{smallmatrix}\right)$. Then, do the same for the transition matrix $\left(\begin{smallmatrix} p & q \\ q & p \end{smallmatrix}\right)$. Deduce that the associated measure-preserving dynamical systems are not isomorphic.

12.6 Comments

For a measurable map $T : X \to X$ that is injective when restricted to the elements of a finite partition $\{X_i\}$, the Jacobian $|T'_\mu|$ as defined earlier a priori does not depend on the chosen partition. However, if $\{Y_j\}$ is another partition such that T is injective when restricted to its elements, then the Jacobians $T'_{\mu|X_i}$ and $T'_{\mu|Y_j}$ coincide almost everywhere when restricted to $X_i \cap Y_j$. The function $|T'_\mu|$ can therefore be considered well defined up to a set of measure 0, without needing to refer to a specific partition.

The inverse of the Jacobian admits a probabilistic interpretation. The quantity $1/|T'_\mu(x)|$ corresponds to the probability of obtaining the value x among all values of $T^{-1}(T(x))$ if we know the value of $T(x)$.

We have restricted ourselves to the case of a partition $\{X_i\}$ with finite cardinality because it is in this context that the concept of entropy was defined. The results proved earlier also hold if the partition $\{X_i\}$ is countable; the proofs are the same.

A Lebesgue space X on which there is a measurable map $T : X \to X$ that preserves the measure, and for which the cardinality of the fibers $T^{-1}(x)$ is countable for every $x \in X$, automatically admits a countable partition $\{X_i\}_{i \in \mathbf{N}}$ such that T restricted to each of the X_i is injective. This result is proved in the book by Parry [17].

We can relax the assumption of uniform dilation in the proposition showing that the partition $\{X_i\}$ is generating (Proposition 12.3). Indeed, it suffices that a power of the transformation be dilating. When the space X is compact, it suffices to have a bound of the form $d(Tx, Ty) > d(x, y)$ on the elements of the partition $\{X_i\}$. When the transformation is ergodic, we can settle for a dilation that is uniform on only one element X_i, and a bound of the form $d(Tx, Ty) \geqslant d(x, y)$ on all other elements. Almost all points in X have an orbit that passes infinitely many times through the element X_i on which we have the dilation, and this suffices to conclude.

Let X be a compact manifold, and let T be a C^1 differentiable map on X. The condition that $\det D_x T > 1$ at every point ensures that the transformation restricted to any set of sufficiently small diameter is uniformly dilating. Not all differential manifolds admit such a transformation. Only the manifolds that can be written as the quotient of a nilpotent Lie group by a discrete subgroup are likely to admit such maps.

The Jacobian is invariant by measurable conjugation. It can be used to distinguish between the unilateral shifts. For example, the unilateral shift associated with the Markov chain with transition matrix $\left(\begin{smallmatrix} p & q \\ p & q \end{smallmatrix}\right)$ and the one associated with the Markov chain with transition matrix $\left(\begin{smallmatrix} p & q \\ q & p \end{smallmatrix}\right)$ have the same entropy. However, the partitions given by the level sets of the Jacobian cannot be isomorphic: in the first case, this partition generates the Borel sets under the action of the shift, whereas in the second case, the generated partition is invariant under the permutation of the two symbols.