Chapter 5
Entropy for Continuous Maps

Measure-theoretic entropy was introduced to measure the complexity of measure-preserving transformations, as an invariant of measurable isomorphism, and to characterize distinguished measures. A continuous map T on a compact metric space has at least one invariant measure, so we start by analyzing the behavior of measure-theoretic entropy $h_\mu(T)$ as a function of the invariant probability measure $\mu \in \mathscr{M}^T$, where $\mathscr{M}^T = \mathscr{M}^T(X)$ denotes the space of T-invariant Borel probability measures on X. This problem is considered in Section 5.1, then in Section 5.2 a more intrinsic (purely topological) notion of entropy for continuous maps is introduced.

5.1 Continuity Properties of Entropy as a function of the measure

Suppose that (X, d) is a compact metric space and $T : X \to X$ is a continuous map. Then

$$h : \mu \longmapsto h_\mu(T)$$

is a function from $\mathscr{M}^T(X)$ to $[0, \infty]$. Recall from the discussion on [52, p. 454] that the space $\mathscr{M}^T(X)$ is weak* compact. The next example shows that even for the simplest of maps T, the map h is not (lower semi-)continuous.

Example 5.1. Let $X = \mathbb{T}$ and $T(x) = 2x \pmod 1$. Then $m_{\mathbb{T}}$, Lebesgue measure on $\mathbb{T}$ is T-invariant. Define a sequence of T-invariant probability measures (μ_k) by

$$\mu_k = \frac{1}{3^k} \sum_{\ell=0}^{3^k-1} \delta_{\ell/3^k}.$$

Then $h_{\mu_k}(T) = 0$ for all k (since $(\mathbb{T}, \mathscr{B}, \mu_k, T)$ is measurably isomorphic to a permutation of 3^k points), while $h_{m_{\mathbb{T}}}(T) = \log 2$. Even though $\mu_k \to m_{\mathbb{T}}$ as $k \to \infty$ in the weak* topology, the entropy $h_{\mu_k}(T)$ does not converge to $h_{m_{\mathbb{T}}}(T)$ as $k \to \infty$. It follows that the entropy of a continuous map with respect to a weak* convergent sequence of probability measures can jump up in the limit.

Example 5.2. In general $\mu \mapsto h_\mu(T)$ is also not upper semi-continuous. Let $X = \{(x,y) \mid 0 \leqslant y \leqslant x \leqslant 1\}$ and define a map T on X by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 2y \pmod{x} \end{pmatrix}.$$

For each $x \in [0,1]$ let $\mu_x = \delta_x \times m_{[0,x]}$ where $m_{[0,x]}$ is Lebesgue measure on $[0,x]$ normalized to have $m_{[0,x]}([0,x]) = 1$. Then

$$h_{\mu_x}(T) = \begin{cases} 0 & \text{if } x = 0; \\ \log 2 & \text{if } x > 0, \end{cases}$$

while $\mu_x \to \mu_0$ as $x \to 0$ by definition. Thus the entropy of a continuous map with respect to a weak* convergent sequence of probability measures can jump down in the limit.

Thus in general the entropy map $\mu \mapsto h_\mu(T)$ does not respect the topological structure of $\mathscr{M}^T(X)$ at all. However, we note that Example 5.1 is a quite natural dynamical system and sequence of weakly converging measures, while Example 5.2 seems artificially constructed to fail to be upper semi-continuous. And indeed, upper semi-continuity holds for many natural classes of dynamical systems. The most significant of the results in this direction that do not require expansiveness (as discussed below) are due to Newhouse [146], [147] who proved in particular that upper semi-continuity holds for C^∞ diffeomorphisms of compact manifolds.(22) One of the reasons upper semi-continuity is so important is that it ensures the existence of a *maximal measure*, a notion which will be discussed in detail in Chapter 8.

Definition 5.3 (Expansiveness). A homeomorphism $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ of a compact metric space is called *expansive*(23) if there is a $\delta > 0$ such that

$$\sup_{n \in \mathbb{Z}} \mathsf{d}(T^n x, T^n y) \leqslant \delta \implies x = y. \tag{5.1}$$

Any $\delta > 0$ that has the property in (5.1) is called an *expansive constant* for T.

Example 5.4. The automorphism of $\mathbb{T}^2$ associated to the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ from Section 1.6 and [52, Ex. 2.41] is expansive.

More generally, an automorphism of $\mathbb{T}^k$ associated to a matrix A is expansive if and only if the matrix A is hyperbolic, which means the eigenvalues all have modulus not equal to one.(24)

Lemma 5.5 (Generator). *Let $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ be an expansive homeomorphism with expansive constant δ. Then any finite partition ξ with the property that* $\operatorname{diam}(P) < \delta$ *for all $P \in \xi$ is a generator of the measure-preserving system $(X, \mathscr{B}, \mu, T)$ for any $\mu \in \mathscr{M}^T$.*

Proof. Let ξ be a finite partition as in the statement of the lemma, and suppose that $\eta = \max_{P\in\xi} \operatorname{diam}(P) < \delta$. Let

$$\mathscr{A}_0 = \bigcup_{n=1}^{\infty} \sigma\bigl(\xi_{-n}^{n}\bigr),$$

which is a countable algebra. We need to show that $\mathscr{A}_0$ generates the σ-algebra $\mathscr{B}$. For this it is enough to show that any open set $O \subseteq X$ can be written as a union (automatically a countable union) of elements of $\mathscr{A}_0$.

Let $O \subseteq X$ be open, and let $x \in O$. Then we claim that there exists some n with

$$[x]_{\bigvee_{k=-n}^{n} T^{-k}\xi} \subseteq O.$$

Suppose the opposite. Then for every n there would exist some $y_n \notin O$ with $\mathsf{d}(T^k x, T^k y_n) \leqslant \eta$ for $k = -n, \dots, n$. Let y be the limit of a convergent subsequence of the sequence (y_n), then continuity of T shows that

$$\mathsf{d}(T^k x, T^k y) \leqslant \eta < \delta$$

for any $k \in \mathbb{Z}$. Since $x \in O$ but $y \notin O$ this contradicts the expansiveness of T. The claim implies the lemma. □

In contrast to Example 5.2, we have the following result.

Theorem 5.6 (Upper semi-continuity). *If $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ is an expansive homeomorphism, then the map $\mu \mapsto h_\mu(T)$ is upper semi-continuous.*

What this means is that for any measure μ and for any $\varepsilon > 0$, there is a weak* neighbourhood U of μ with the property that

$$\nu \in U \implies h_\nu(T) < h_\mu(T) + \varepsilon.$$

For any set $B \subseteq X$, a metric space, write B^o for the interior of B, $\overline{B}$ for the closure and ∂B for the boundary. For the proof of Theorem 5.6 we will use Lemma 5.5 and two general lemmas.

Lemma 5.7 (Weak* continuity for sets). *If B is a Borel set in a compact metric space (X, d), $\mu, \mu_1, \mu_2, \dots$ probability measures on X satisfying $\mu(\partial B) = 0$ and $\mu_n \to \mu$ as $n \to \infty$ in the weak* topology, then also $\mu_n(B) \to \mu(B)$ as $n \to \infty$.*

Proof. Since $\mu(\partial B) = 0$, for any $\varepsilon > 0$ there is a compact set K and an open set U with $\mu(U \smallsetminus K) < \varepsilon$, $B^o \underset{\mu}{=} \overline{B}$, and $K \subseteq B^o \subseteq B \subseteq \overline{B} \subseteq U$.

Choose Urysohn functions f and g with $\mathbb{1}_K < f < \mathbb{1}_{B^o}$ and $\mathbb{1}_{\overline{B}} < g < \mathbb{1}_U$. Then $\int f \,\mathrm{d}\mu \leqslant \mu(B) \leqslant \int g \,\mathrm{d}\mu$ and $\int (g-f)\,\mathrm{d}\mu < \varepsilon$. Now by the weak* convergence, there is an N such that if $n > N$ then

$$\left| \int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu_n \right| < \varepsilon$$

and

$$\left| \int g \,\mathrm{d}\mu - \int g \,\mathrm{d}\mu_n \right| < \varepsilon.$$

It follows that for $n > N$,

$$\int f \,\mathrm{d}\mu - \varepsilon < \int f \,\mathrm{d}\mu_n < \mu_n(B) < \int g \,\mathrm{d}\mu_n < \int g \,\mathrm{d}\mu + \varepsilon,$$

so $|\mu(B) - \mu_n(B)| < 2\varepsilon$. □

As before we will use the abbreviation $\xi_k^\ell = \bigvee_{j=k}^{\ell} T^{-j}\xi$ whenever $k \leqslant \ell$ are integers† and ξ is a finite or countable partition of X.

Proof of Theorem 5.6. Apply Lemma 3.13 for X, μ, and an expansive constant δ. By Lemma 5.5 the partition ξ is a generator. Fix $\varepsilon > 0$ and choose N with

$$\frac{1}{N} H_\mu\left(\xi_0^{N-1}\right) < h_\mu(T) + \varepsilon.$$

The partition ξ_0^{N-1} contains finitely many atoms Q each with the property that $\mu(\partial Q) = 0$. By Lemma 5.7 there is a weak* neighbourhood U of μ with the property that $\nu \in U \implies |\nu(Q) - \mu(Q)| < \varepsilon'$ for all $Q \in \xi_0^{N-1}$. It follows, by continuity of the function ϕ defined by (1.6) on page 10, that if ε' is small enough, then

$$\frac{1}{N} H_\nu\left(\xi_0^{N-1}\right) \leqslant \frac{1}{N} H_\mu\left(\xi_0^{N-1}\right) + \varepsilon \leqslant h_\mu(T) + 2\varepsilon,$$

so $h_\nu(T) \leqslant h_\mu(T) + 2\varepsilon$, since

$$h_\nu(T) = \inf_{n \in \mathbb{N}} \left\{ \frac{1}{n} H_\nu\left(\xi_0^{n-1}\right) \right\}.$$

□

It is not difficult to generalize Definition 5.3 through Theorem 5.6 to give the notion of *forwardly expansive maps* and the result that the entropy map $\mu \mapsto h_\mu(T)$ is upper semi-continuous when T is a forwardly expansive map. Using this, Example 5.1 is a forwardly expansive endomorphism of $\mathbb{T}$. This shows that expansiveness cannot give more than upper semi-continuity (see also Exercises 5.1.1 and 5.1.2).

† The integers are assumed to be non-negative if T is not invertible and otherwise arbitrary.

Finally we note that even though the entropy function $\mu \in \mathscr{M}^T \mapsto h_\mu(T)$ is not continuous in most examples, it does nonetheless respect the convex structure of $\mathscr{M}^T$ (see Section 2.6).

Exercises for Section 5.1

Exercise 5.1.1. Show that $\mu \mapsto h_\mu(\sigma)$ is not lower semi-continuous for the full shift

$$\sigma : \{1, 2, \dots, k\}^{\mathbb{Z}} \longrightarrow \{1, 2, \dots, k\}^{\mathbb{Z}}$$

on an alphabet with $k \geqslant 2$ symbols.

Exercise 5.1.2. Show that $\mu \mapsto h_\mu(T_A)$ is not lower semi-continuous for the automorphism T_A of $\mathbb{T}^2$ associated to the matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ from Section 1.6.

Exercise 5.1.3. Let $T : X \to X$ be an expansive homeomorphism of a compact metric space. Show that for any $n \geqslant 1$ the set $\{x \in X \mid T^n x = x\}$ of points of period n under T is finite.

Exercise 5.1.4. Let $A \in \mathrm{GL}_r(\mathbb{Z})$ be a quasihyperbolic matrix (see Exercise 2.5.2). Let $T_A : \mathbb{T}^r \to \mathbb{T}^r$ be the corresponding quasihyperbolic automorphism. Show that T_A is not expansive but that the map

$$\mathscr{M}^T \ni \mu \longmapsto h_\mu(T_A)$$

is nonetheless upper semi-continuous.

Exercise 5.1.5. Let (X, d) be a compact metric space, let $T : X \to X$ be an expansive homeomorphism with expansive constant $\delta > 0$, and assume that $\mu \in \mathscr{M}^T(X)$ is an ergodic Borel probability measure.
(a) Use the methods of Section 3.4 to show that

$$-\frac{1}{n} \log \mu \left(D_n(x, \delta, T) \right) \longrightarrow h_\mu(T)$$

as $n \to \infty$, for μ-almost every $x \in X$.
(a) Assuming the result from Exercise 3.4.1, show that

$$-\frac{1}{2n} \log \mu \left(D_n^{\pm}(x, \delta, T) \right) \longrightarrow h_\mu(T)$$

as $n \to \infty$, for μ-almost every $x \in X$.

5.2 Topological Entropy

Entropy was originally defined purely in terms of measure-theoretic properties. As a continuous map may have many invariant measures, it became important to define a topological analog of entropy to reflect topological properties of a map. There are many ways in which the dependence on a specific measure may be removed in defining the entropy of a continuous map, of which two will be particularly useful. First, the set $\{h_\mu(T) \mid \mu \in \mathscr{M}^T\}$ may be considered as a whole. By Theorem 2.33, entropy is an affine function on the convex set $\mathscr{M}^T$, so the above set is an interval in $[0,\infty]$. Hence it is natural to associate the quantity $\sup\{h_\mu(T) \mid \mu \in \mathscr{M}^T\}$ to the continuous map T. Second, one may try to emulate the definition of measure-theoretic entropy using the metric or topological structure of a continuous map.

The second approach, which we develop now, was initiated by Adler, Konheim and McAndrew [5]. We shall see later that the two notions end up with the same quantity.

5.2.1 Cover entropy

Definition 5.8. For a cover $\mathscr{U}$ of a compact topological space X, define $N(\mathscr{U})$ to be the smallest cardinality of a subcover of $\mathscr{U}$, and define the entropy of $\mathscr{U}$ to be $H(\mathscr{U}) = \log N(\mathscr{U})$.

For covers $\mathscr{U}$ and $\mathscr{V}$ of X, we write $\mathscr{U} \vee \mathscr{V}$ for the cover

$$\mathscr{U} \vee \mathscr{V} = \{U \cap V \mid U \in \mathscr{U}, V \in \mathscr{V}\},$$

and notice that $N(\mathscr{U} \vee \mathscr{V}) \leqslant N(\mathscr{U})N(\mathscr{V})$. If T is continuous, then for any open cover $\mathscr{U}$, $T^{-1}(\mathscr{U})$ is also an open cover satisfying $N(T^{-1}\mathscr{U}) \leqslant N(\mathscr{U})$. We again introduce the shorthand

$$\mathscr{U}_k^\ell = \bigvee_{j=k}^{\ell} T^{-j}\mathscr{U}$$

for any integers $k \leqslant \ell$ (which are assumed to be non-negative if T is not a homeomorphism). Finally, the sequence (a_n) defined by

$$a_n = \log N\left(\mathscr{U}_0^{n-1}\right)$$

is sub-additive since

$$a_{m+n} = \log N(\mathscr{U}_0^{m+n-1}) \leqslant \log(N(\mathscr{U}_0^{m-1})N(\mathscr{U}_m^{m+n-1})) \leqslant a_m + a_n$$

for all $m, n \geqslant 1$. This proves the convergence statement implicit in Definition 5.9, by Lemma 1.13.

Definition 5.9 (Cover entropy). Let $T : X \to X$ be a continuous map on a compact topological space. The *cover entropy* of T with respect to an open cover $\mathscr{U}$ is defined to be

$$h_{\text{cover}}(T, \mathscr{U}) = \lim_{n\to\infty} \frac{1}{n} \log N\left(\mathscr{U}_0^{n-1}\right) = \inf_{n\geqslant 1} \frac{1}{n} \log N\left(\mathscr{U}_0^{n-1}\right),$$

and the *cover entropy* of T is defined to be

$$h_{\text{cover}}(T) = \sup_{\mathscr{U}} h_{\text{cover}}(T, \mathscr{U})$$

where the supremum is taken over all open covers of X.

As was the case with measure-theoretic entropy, it is not clear how to use this definition to actually compute topological entropy. The properties of a topological generator discussed below may help towards solving this problem.

We also note that the number $h_{\text{cover}}(T, \mathscr{U})$ is monotone with respect to $\mathscr{U}$ in the following way. We say that $\mathscr{V}$ is a refinement of $\mathscr{U}$, and write $\mathscr{U} \preccurlyeq \mathscr{V}$, if every set $V \in \mathscr{V}$ is contained in a set $U \in \mathscr{U}$. This implies $N(\mathscr{U}) \leqslant N(\mathscr{V})$, that $\mathscr{V}_0^{n-1}$ is a refinement of $\mathscr{U}_0^{n-1}$, and hence that

$$h_{\text{cover}}(T, \mathscr{U}) \leqslant h_{\text{cover}}(T, \mathscr{V}).$$

The results below make sense for continuous maps on compact spaces, but we restrict attention to the metric setting since that is simpler and covers what is needed for the applications considered here.

Definition 5.10. Let $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ be a homeomorphism on a compact metric space. A two-sided *topological generator* for T is a finite open cover $\mathscr{U} = \{U_i\}_{i \in I}$ with the property that for any sequence $(i_k)_{k \in \mathbb{Z}}$, the set

$$\bigcap_{k \in \mathbb{Z}} T^{-k} \left(\overline{U_{i_k}}\right)$$

contains at most one single point.

It is straightforward to modify the definition to obtain a notion of one-sided topological generator for continuous maps. The following results hold similarly for this notion.

Theorem 5.11 (Topological generators). *Let $\mathscr{U}$ be a two-sided generator for a homeomorphism*

$$T : (X, \mathsf{d}) \to (X, \mathsf{d})$$

on a compact metric space. Then $h_{\text{cover}}(T, \mathscr{U}) = h_{\text{cover}}(T)$.

For the proof we need a standard result about compact metric spaces.

Lemma 5.12 (Lebesgue number). *Let (X, d) be a compact metric space and $\mathscr{U}$ an open cover of X. Then there is some $\delta > 0$, called a* Lebesgue number *for $\mathscr{U}$, with the property that any subset of X with diameter less than δ is contained in some element of $\mathscr{U}$.*

Proof. If no Lebesgue number exists, then for any $\frac{1}{n} > 0$ there is an $x_n \in X$ with the property that $B_{1/n}(x_n) \nsubseteq U$ for any $U \in \mathscr{U}$. By compactness, we may choose a convergent subsequence (x_{n_j}) with $x_{n_j} \to x^* \in X$

as $j \to \infty$. Since $\mathscr{U}$ is an open cover, there is some $U^* \in \mathscr{U}$ and some $\delta > 0$ with $B_\delta(x^*) \subseteq U^*$. Choose j so that $\frac{1}{n_j} < \frac{\delta}{2}$ and $\mathsf{d}(x_{n_j}, x^*) < \frac{\delta}{2}$. Then $B_{1/n}(x_{n_j}) \subseteq U^*$, which is a contradiction. □

Theorem 5.11 is a direct analog of the Kolmogorov–Sinaĭ theorem (Theorem 1.21); it is also useful to have the analog of the sequence formulation of the Kolmogorov–Sinaĭ theorem (Exercise 1.3.1 or Theorem 2.20). For this we define the diameter of a family of sets $\mathscr{U}$ inside a metric space (X, d) by the formula

$$\operatorname{diam} \mathscr{U} = \sup_{O \in \mathscr{U}} \operatorname{diam} O$$

where $\operatorname{diam} B = \sup_{x,y \in B} \mathsf{d}(x, y)$ for any $B \subseteq X$.

Proposition 5.13. *Let $T : X \to X$ be a continuous map on a compact metric space (X, d). If $(\mathscr{U}_n)$ is a sequence of open covers of X with $\operatorname{diam}(\mathscr{U}_n) \to 0$ as $n \to \infty$, then*

$$\lim_{n\to\infty} h_{\text{cover}}(T, \mathscr{U}_n) = h_{\text{cover}}(T).$$

Proof. Assume first that $h_{\text{cover}}(T)$ is finite, fix $\varepsilon > 0$, choose an open cover $\mathscr{V}$ for which $h_{\text{cover}}(T, \mathscr{V}) > h_{\text{cover}}(T) - \varepsilon$, and let δ be a Lebesgue number for $\mathscr{V}$. Choose N so that $n \geqslant N$ implies that $\operatorname{diam}(\mathscr{U}_n) < \delta$, so $\mathscr{U}_n \preccurlyeq \mathscr{V}$ and hence

$$h_{\text{cover}}(T) \geqslant h_{\text{cover}}(T, \mathscr{U}_n) \geqslant h_{\text{cover}}(T, \mathscr{V}) > h_{\text{cover}}(T) - \varepsilon,$$

showing that $\lim_{n\to\infty} h_{\text{cover}}(T, \mathscr{U}_n) = h_{\text{cover}}(T)$. If $h_{\text{cover}}(T) = \infty$ then for any R there is an open cover $\mathscr{V}$ for which $h_{\text{cover}}(T, \mathscr{V}) > R$, and we may proceed as before. □

Proof of Theorem 5.11. We claim first that for any $\varepsilon > 0$ there is an M such that each set in the refinement $\mathscr{U}_{-M}^{M}$ has diameter less than ε. If not, there is some $\varepsilon > 0$ such that for each $k \geqslant 0$ we may find points $x_1^{(k)}, x_2^{(k)}$ with $\mathsf{d}(x_1^{(k)}, x_2^{(k)}) \geqslant \varepsilon$ that lie in the same element of $\mathscr{U}_{-k}^{k}$. By compactness, there is a subsequence k_j with $x_1^{(k_j)} \to x_1^*$ and $x_2^{(k_j)} \to x_2^*$, and $\mathsf{d}(x_1^*, x_2^*) \geqslant \varepsilon$. As $\mathscr{U}$ is a finite cover, we may choose the subsequence so that $T^n(x_1^{(k_j)})$ and $T^n(x_2^{(k_j)})$ belong to the same element $U_{i_n} \in \mathscr{U}$ for some sequence (i_n) for all $n \in \mathbb{Z}$. Therefore, the points x_1^* and x_2^* (with $\mathsf{d}(x_1^*, x_2^*) \geqslant \varepsilon$) lie in a single set

$$\bigcap_{n\in\mathbb{Z}} T^{-n}\left(\overline{U_{i_n}}\right),$$

contradicting the assumption that $\mathscr{U}$ is a two-sided generator.

By Proposition 5.13 we obtain

$$h_{\text{cover}}(T) = \lim_{M\to\infty} h_{\text{cover}}(T, \mathscr{U}_{-M}^{M}).$$

Using only the definition we also obtain

$$\begin{aligned} h_{\text{cover}}\left(T, \mathscr{U}_{-M}^{M}\right) &= \lim_{k\to\infty} \frac{1}{k} \log N\left(\bigvee_{j=0}^{k-1} T^{-j}(\mathscr{U}_{-M}^{M})\right) \\ &= \lim_{k\to\infty} \frac{2M+k}{k} \frac{1}{2M+k} \log N\left(\mathscr{U}_0^{2M+k-1}\right) = h_{\text{cover}}(T, \mathscr{U}), \end{aligned}$$

which proves the theorem. □

5.2.2 Entropy of Shift Maps

We recall that a *block* in a sequence $a = (\dots a_{-1}a_0a_1\dots) \in A^{\mathbb{Z}}$ is a finite segment $u = (a_j a_{j+1} \cdots a_k)$ for some $j \leqslant k$. We will say that the block *appears in the point* $a \in A^{\mathbb{Z}}$ in this case, and will also say that a block u *appears in a closed shift-invariant subset* $X \subseteq A^{\mathbb{Z}}$ if there is some $a \in X$ with the property that u appears in a. Notice that counting blocks of a given length is a measure of the combinatorial size of a shift-invariant subset. Two extreme examples are the full shift with alphabet A, which has $|A|^n$ blocks of length n, and the singleton $\{a^\infty\}$ containing the single point a^∞ with $a^\infty(j) = a$ for all $j \in \mathbb{Z}$, which has the single block $a^n = a \cdots a$ of length n for any $n \geqslant 1$.

Corollary 5.14 (Entropy of shift spaces). *Let X be a closed, shift-invariant subset of the full shift $\prod_{n\in\mathbb{Z}}\{0, 1, \dots, s-1\}$, and let $t_n(X)$ denote the number of blocks of length $n \geqslant 1$ that appear in any element of X. Then $\frac{1}{n}\log t_n(X)$ converges, and*

$$h_{\text{cover}}(\sigma|_X) = \lim_{n\to\infty} \frac{1}{n} \log t_n(X).$$

Proof. The open cover $\mathscr{U}$ by the sets $\{x \in X \mid x_0 = j\}$ for $1 \leqslant j \leqslant s-1$ is a generator, and $t_n(X) = N\left(\mathscr{U}_0^{n-1}\right)$; the result follows by Theorem 5.11. □

We refer to Appendix B for the result used implicitly in the next corollary, that an irreducible matrix has a unique largest positive eigenvalue which coincides with the spectral radius of the matrix.

Corollary 5.15 (Entropy of vertex shift). *Let $(X_G^{(v)}, \sigma)$ be an irreducible vertex shift (see Section A.4.1). Then*

$$h_{\text{cover}}(\sigma) = \log \lambda_A,$$

where λ_A is the largest positive eigenvalue of the adjacency matrix $A = A_G$.

Proof. A cylinder set $[i_0, \dots, i_{n-1}]_0^{n-1}$ is non-empty in $X_G^{(v)}$ if and only if $a_{i_0,i_1} \cdots a_{i_{n-2},a_{n-1}} = 1$, so

$$t_n\left(X_G^{(v)}\right) = \sum_{i_0,\dots,i_{n-1}} a_{i_0,i_1}\cdots a_{i_{n-2},i_{n-1}} = \sum_{i_0,i_{n-1}} \left(A^{n-1}\right)_{i_0,i_{n-1}} = \|A^{n-1}\|_1,$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm on the space of $|A|\times|A|$ matrices. Since all norms on this space are equivalent, we deduce by the spectral radius formula that

$$h_{\text{cover}}(\sigma) = \lim \frac{1}{n}\log t_n(X_G^{(v)}) = \log\lambda_A.$$

□

5.2.3 Seperating and spanning entropies

In topological dynamics, the behavior of individual points is visible, and this allows an alternative approach to entropy via measuring the asymptotic complexity of the metric space with respect to a sequence of metrics defined using the continuous map (equivalently, measuring the complexity of the orbits under the map). This approach is similar to the notion of name entropy in Section 1.4 and Exercise 3.1.2, and is due to Dinaburg [41] and Bowen. As we shall see in Section 6.3, Bowen in particular used this to extend the definition of topological entropy to uniformly continuous maps on locally compact metric spaces.

Definition 5.16. Let $T:(X,\mathsf{d})\to(X,\mathsf{d})$ be a continuous map on a compact metric space. A subset $F_{\text{span}}\subseteq X$ is (n,ε)-*spanning* if for every $x\in X$ there is a point $y\in F_{\text{span}}$ with

$$\max_{0\leqslant i\leqslant n-1}\mathsf{d}(T^ix,T^iy)\leqslant\varepsilon.$$

Define $s_{\text{span}}(n,\varepsilon)$ to be the smallest cardinality of an (n,ε)-spanning set. A subset $F_{\text{sep}}\subseteq X$ is (n,ε)-*separated* if for any two distinct points $x,y\in F_{\text{sep}}$ we have

$$\max_{0\leqslant i\leqslant n-1}\mathsf{d}(T^ix,T^iy)>\varepsilon.$$

Define $s_{\text{sep}}(n,\varepsilon)$ to be the largest cardinality of an (n,ε)-separated set.

Notice that both $s_{\text{span}}(n,\varepsilon)$ and $s_{\text{sep}}(n,\varepsilon)$ are finite by compactness. The *spanning set entropy* of T is

$$h_{\text{span}}(T) = \lim_{\varepsilon\searrow 0}\limsup_{n\to\infty}\frac{1}{n}\log s_{\text{span}}(n,\varepsilon),$$

and the *separated set entropy* of T is

$$h_{\text{sep}}(T) = \lim_{\varepsilon\searrow 0}\limsup_{n\to\infty}\frac{1}{n}\log s_{\text{sep}}(n,\varepsilon).$$

Notice that $\varepsilon < \varepsilon'$ implies that

$$s_{\text{span}}(n, \varepsilon) \geqslant s_{\text{span}}(n, \varepsilon')$$

and

$$s_{\text{sep}}(n, \varepsilon) \geqslant s_{\text{sep}}(n, \varepsilon'),$$

so the limit in ε exists in both cases. The following also provides the crucial step in showing that spanning set entropy and separated set entropy are equal.

Lemma 5.17. *Using the same assumptions and notations as in Definition 5.16 we have $s_{\text{span}}(n, \varepsilon) \leqslant s_{\text{sep}}(n, \varepsilon) \leqslant s_{\text{span}}(n, \varepsilon/2)$ for any $n \geqslant 1$ and $\varepsilon > 0$.*

This lemma is really an application of a more general statement about metric spaces. The link between the above and the following lemma is provided by studying for any $n \geqslant 1$ the *Bowen metric*

$$\mathsf{d}_n(x, y) = \max_{0 \leqslant i \leqslant n-1} \mathsf{d}(T^i x, T^i y)$$

for $x, y \in X$, that implicitly already appeared in Definition 5.16 (see also Exercise 5.2.2).

Lemma 5.18. *Let (X, d) be a compact metric space. For $\varepsilon > 0$ we let $s_{\text{span}}(\varepsilon)$ be the minimal number of ε-spanning points, that is the minimal number of closed ε-balls that are needed to cover X, and let $s_{\text{sep}}(\varepsilon)$ be the maximal number of ε-separated points (that is, points of distance $> \varepsilon$ apart) that can fit into X. Then $s_{\text{span}}(\varepsilon) \leqslant s_{\text{sep}}(\varepsilon) \leqslant s_{\text{span}}(\varepsilon/2)$ for any $\varepsilon > 0$.*

Proof of Lemmas 5.18 and hence 5.17. Fix some $\varepsilon > 0$. Let F_{sep} be a maximal set of ε-separated points with cardinality $s_{\text{sep}}(\varepsilon)$. If $y \in X$ then y must have distance no more than ε to some $x \in F_{\text{sep}}$ (for otherwise the set F_{sep} would not have been maximal). Hence the closed ε-balls with centers in F_{sep} cover and $s_{\text{span}}(\varepsilon) \leqslant s_{\text{sep}}(\varepsilon)$ follows.

Let F_{span} be now a set of cardinality $s_{\text{span}}(\varepsilon/2)$ and such that the closed ε-balls with centers in F_{span} cover all of X. Then for any $x \in F_{\text{span}}$ the closed ball of radius $\varepsilon/2$ can only contain one point of the ε-separated set F_{sep}. Hence

$$s_{\text{sep}}(\varepsilon) = |F_{\text{sep}}| \leqslant |F_{\text{span}}| = s_{\text{span}}(\varepsilon/2).$$

Lemma 5.17 follows by applying Lemma 5.18 to the Bowen metric d_n. $\square$

5.2.4 Topological entropy

Theorem 5.19 (Topological entropy). *Let $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ be a continuous map on a compact metric space. Then*

$$
\begin{aligned}
h_{\text{cover}}(T) = h_{\text{sep}}(T) &= h_{\text{span}}(T) \\
&= \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log s_{\text{sep}}(n,\varepsilon) \\
&= \lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n}\log s_{\text{span}}(n,\varepsilon).
\end{aligned}
$$

The common value is called the topological entropy *of T, denoted $h_{\text{top}}(T)$.*

Proof. Lemma 5.17 shows that $h_{\text{sep}}(T) = h_{\text{span}}(T)$.

Now let $\mathscr{U}$ be the cover of X by all open balls of radius 2ε, and $\mathscr{V}$ be any cover by open balls of radius $\varepsilon/2$. Then

$$
N\left(\mathscr{U}_0^{n-1}\right) \leqslant s_{\text{span}}(n,\varepsilon,X,\mathsf{d}) \leqslant s_{\text{sep}}(n,\varepsilon,X,\mathsf{d}) \leqslant N\left(\mathscr{V}_0^{n-1}\right). \tag{5.2}
$$

In fact, suppose F_{span} is an (n,ε)-spanning set of cardinality $s_{\text{span}}(n,\varepsilon)$. Then the sets $\bigcap_{k=0}^{n} T^{-k}B_{2\varepsilon}(T^k x) \in \mathscr{U}_0^{n-1}$ for $x \in F_{\text{span}}$ form a cover, which gives the first inequality (we use the cover $\mathscr{U}$ by open balls of radius 2ε since the spanning set is defined by closed balls in the metric d_n). The second inequality comes from Lemma 5.17. For the last inequality notice that no two points of a maximal (n,ε)-separated set F_{sep} can belong to the same element of $\mathscr{V}_0^{n-1}$.

Taking the logarithm of (5.2), dividing by n and taking the limit we obtain

$$
\begin{aligned}
h_{\text{cover}}(T,\mathscr{U}) &\leqslant \liminf_{n\to\infty}\frac{\log s_{\text{span}}(n,\varepsilon)}{n} \\
&\leqslant \liminf_{n\to\infty}\frac{\log s_{\text{sep}}(n,\varepsilon)}{n} \\
&\leqslant \limsup_{n\to\infty}\frac{\log s_{\text{sep}}(n,\varepsilon)}{n} \leqslant h_{\text{cover}}(T,\mathscr{V})
\end{aligned}
$$

Letting $\varepsilon \to 0$ and applying Proposition 5.13 gives the theorem. □

This result means that we can speak about ‘topological entropy’ and use whichever definition is most convenient for continuous maps on compact metric spaces.

Lemma 5.20. *If $T : X \to X$ is a homeomorphism of a compact metric space, then $h_{\text{top}}(T^{-1}) = h_{\text{top}}(T)$.*

Proof. Notice that for any continuous map T and open cover $\mathscr{U}$ we have

$$
N(T^{-1}\mathscr{U}) \leqslant N(\mathscr{U}),
$$

but if T is a homeomorphism we have $N(T^{-1}\mathscr{U}) = N(\mathscr{U})$. Thus

$$\begin{aligned}
h_{\mathrm{cover}}(T,\mathscr{U}) &= \lim_{n\to\infty}\frac{1}{n}\log N\left(\mathscr{U}_0^{n-1}\right)\\
&= \lim_{n\to\infty}\frac{1}{n}\log N\left(T^{n-1}\bigvee_{i=0}^{n-1}T^{-i}\mathscr{U}\right)\\
&= \lim_{n\to\infty}\frac{1}{n}\log N\left(\bigvee_{i=0}^{n-1}T^{i}\mathscr{U}\right) = h_{\mathrm{cover}}(T^{-1},\mathscr{U}).
\end{aligned}$$

□

Moreover, the topological entropy adds for iterates of a continuous map.

Lemma 5.21. *If $T : X \to X$ is a continuous map of a compact metric space, then $h_{\mathrm{top}}(T^k) = kh_{\mathrm{top}}(T)$ for any $k \geqslant 1$.*

PROOF.First, for any open cover $\mathscr{U}$ of X, we have

$$\begin{aligned}
h_{\mathrm{cover}}(T^k) &\geqslant h_{\mathrm{cover}}\left(T^k,\bigvee_{i=0}^{k-1}T^{-i}\mathscr{U}\right)\\
&= k\lim_{n\to\infty}\frac{1}{nk}\log N\left(\bigvee_{j=0}^{nk-1}T^{-j}\mathscr{U}\right) = kh_{\mathrm{cover}}(T,\mathscr{U}),
\end{aligned}$$

so $h_{\mathrm{cover}}(T^k) \geqslant kh_{\mathrm{cover}}(T)$. For the reverse inequality, notice that

$$\bigvee_{j=0}^{n-1}(T^k)^{-j}\mathscr{U} \preccurlyeq \bigvee_{j=0}^{nk-1}T^{-j}\mathscr{U}$$

for any open cover $\mathscr{U}$. Thus

$$\begin{aligned}
h(T,\mathscr{U}) &= \lim_{n\to\infty}\frac{1}{nk}\log N\left(\bigvee_{j=0}^{nk-1}T^{-j}\mathscr{U}\right)\\
&\geqslant \lim_{n\to\infty}\frac{1}{nk}\log N\left(\bigvee_{j=0}^{n-1}(T^k)^{-j}\mathscr{U}\right) = \frac{1}{k}h_{\mathrm{cover}}(T^k,\mathscr{U})
\end{aligned}$$

for any open cover $\mathscr{U}$, showing that $h_{\mathrm{cover}}(T^k) \leqslant kh_{\mathrm{cover}}(T)$. □

The following lemma has a purely topological proof, but it is by far more convenient to prove it as a corollary of the variational principle (see Exercise 5.3.3 on page 153 and Lemma 6.4).

Lemma 5.22. *If $T_i : X_i \to X_i$ are continuous maps of compact metric spaces for $i = 1, 2$, then $h_{\mathrm{top}}(T_1 \times T_2) = h_{\mathrm{top}}(T_1) + h_{\mathrm{top}}(T_2)$.*

Exercises for Section 5.2

Exercise 5.2.1. Show that Corollary 5.14 also holds for one-sided shifts.

Exercise 5.2.2. In the setting of Definition 5.16, let $n \geqslant 1$ and define

$$\mathsf{d}_n(x,y) = \max_{0 \leqslant j \leqslant n-1} \mathsf{d}(T^j x, T^j y)$$

for $x, y \in X$. Show that d_n is a metric on X that defines the same topology as the original topology defined by d for each $n \geqslant 1$, and formulate the notion of separated and spanning sets in terms of this *Bowen metric*.

Exercise 5.2.3. Let $T_i : X_i \to X_i$ be a continuous map of a compact metric space for each $i = 1, 2$. Prove that if (X_2, T_2) is a topological factor of (X_1, T_1) (that is, if there is a continuous onto map $\theta : X_1 \to X_2$ with $\theta \circ T_1 = T_2 \circ \theta$), then $h_{\mathrm{top}}(T_2) \leqslant h_{\mathrm{top}}(T_1)$. Deduce that topological entropy is an invariant of topological conjugacy for continuous maps of compact metric spaces.

Exercise 5.2.4. Show that an expansive homeomorphism of a compact metric space has finite topological entropy.

5.3 The Variational Principle

The variational principle[(25)] expresses a relationship between the topological entropy of a continuous map and the entropy with respect to invariant measures.

5.3.1 Periodic Points Producing Positive Entropy

To motivate the proof of Theorem 5.24 in Section 5.3.2 we state and prove a simpler result regarding the limiting behavior of periodic orbits for the map $T_p : x \mapsto px \pmod 1$ on the circle $\mathbb{T}$. As we will see later in the proof of Theorem 5.24, the argument relies on the discrete distribution and spacing properties of a sequence of measures before a weak* limit is taken.[(26)]

Proposition 5.23. *Let $T_p(x) = px$ for $x \in \mathbb{T}$ and some $p \geqslant 2$. Let $\alpha > 0$ and $C > 0$, and assume that for every n with* $\gcd(n, p) = 1$ *we choose a subset $S_n \subseteq \{0, \frac{1}{n}, \dots, \frac{n-1}{n}\}$ with $T_p(S_n) \subseteq S_n$ and $|S_n| \geqslant C n^{\alpha}$. Let*

$$\mu_n = \frac{1}{|S_n|} \sum_{x \in S_n} \delta_x$$

be the normalized counting measure on S_n for $n \geqslant 1$. Then any weak limit μ of (μ_n) is T_p-invariant and satisfies $h_\mu(T_p) \geqslant \alpha \log p$.*

PROOF. Note that $T_p(S_n) \subseteq S_n$ implies that $T_p(S_n) = S_n$ and that μ_n is T_p-invariant. In particular, any weak* limit μ of (μ_n) is also T_p-invariant.

Let $\xi = \{[0, \frac{1}{p}), [\frac{1}{p}, \frac{2}{p}), \dots, [\frac{p-1}{p}, 1)\}$ be the generator for $T = T_p$ from Example 1.28. Notice that

$$\mu(\{0\}) = \mu\bigl(T^{-1}\{0\}\bigr) = \mu\bigl(\{0\} \sqcup \{\tfrac{1}{p}, \tfrac{2}{p}, \dots, \tfrac{p-1}{p}\}\bigr)$$

for any T-invariant measure μ. It follows that $\{\frac{1}{p}, \frac{2}{p}, \dots, \frac{p-1}{p}\}$ is a μ-null set, and so $\mu(\partial P) = 0$ for all $P \in \xi$ unless $0 \in \partial P$ and $\mu(\{0\}) > 0$.

Assume first that $\mu_{n_k} \to \mu$ in the weak* topology, and $\mu(\{0\}) = 0$. Then, for any fixed $m \geqslant 1$, we have $\mu(\partial P) = 0$ for all $P \in \xi_0^{m-1} = \bigvee_{i=0}^{m-1} T^{-i}\xi$ and so

$$\mu(P) = \lim_{k\to\infty} \mu_{n_k}(P)$$

by Lemma 5.7. Just as in the proof of Theorem 5.6, this implies that

$$H_\mu\left(\xi_0^{m-1}\right) = \lim_{k\to\infty} H_{\mu_{n_k}}\left(\xi_0^{m-1}\right) \tag{5.3}$$

for any $m \geqslant 1$.

For each $k \geqslant 1$ write $\lfloor \log_p n_k \rfloor = d_k m + r_k$ with $d_k \geqslant 0$ and $0 \leqslant r_k \leqslant m-1$. By subadditivity of entropy (Proposition 1.7) and T-invariance of μ_{n_k} we have

$$\begin{aligned} H_{\mu_{n_k}}\left(\xi_0^{\lfloor \log_p n_k \rfloor}\right) &\leqslant d_k H_{\mu_{n_k}}\left(\xi_0^{m-1}\right) + H_{\mu_{n_k}}\left(\xi_0^{r_k}\right) \\ &\leqslant d_k H_{\mu_{n_k}}\left(\xi_0^{m-1}\right) + m \log p. \end{aligned} \tag{5.4}$$

Notice that the atoms of the partition $\xi_0^{\lfloor \log_p n_k \rfloor}$ are intervals of length

$$\frac{1}{p} p^{-\lfloor \log_p n_k \rfloor} \leqslant \frac{1}{n_k},$$

so that each such interval contains at most one member of S_{n_k}. In other words, as far as the measure μ_{n_k} is concerned, this partition is the partition into $|S_{n_k}|$ sets of equal measure (and some null sets), so that

$$H_{\mu_{n_k}}\left(\xi_0^{\lfloor \log_p n_k \rfloor}\right) = \log |S_{n_k}|.$$

Together with (5.4) we see that

$$\begin{aligned} \frac{1}{m} H_{\mu_{n_k}}\left(\xi_0^{m-1}\right) &\geqslant \frac{1}{m d_k} \log |S_{n_k}| - \frac{1}{d_k} \log p \\ &\geqslant \frac{\log_p n_k}{m d_k} \alpha \log p + \frac{1}{m d_k} \log C - \frac{1}{d_k} \log p, \end{aligned} \tag{5.5}$$

since, by assumption, $|S_{n_k}| \geqslant C n_k^\alpha$.

By the definition of d_k, we have $\frac{\log_p n_k}{md_k} \to 1$ and $\frac{1}{d_k} \to 0$ as $k \to \infty$. By (5.3) we conclude that

$$\frac{1}{m} H_\mu \left(\xi_0^{m-1}\right) \geqslant \alpha \log p$$

for any $m \geqslant 1$, which shows that $h_\mu(T) \geqslant \alpha \log p$ and proves the result under the assumption that $\mu(\{0\}) = 0$.

Assume now that $\mu(\{0\}) > 0$, so that we cannot use (5.3) for the partition ξ_0^{m-1}. However, only two partition elements $P \in \xi_m$ fail to have the property $\mu(\partial P) = 0$, which can only produce a bounded error as we now show. Let $P_0 = [0, \frac{1}{p^m})$ and $P_{-1} = [\frac{p^m-1}{p^m}, 1)$ be the two problematic elements of ξ_m, and define a new partition

$$\eta_m = \{P_0 \cup P_{-1}\} \cup \xi_m \smallsetminus \{P_0, P_{-1}\}.$$

Notice that ξ_0^{m-1} is a refinement of η_m, since

$$\xi_0^{m-1} = \eta_m \vee \{P_0, P_{-1}\}.$$

Thus

$$H_{\mu_{n_k}}(\xi_0^{m-1}) = H_{\mu_{n_k}}(\eta_m) + H_{\mu_{n_k}}\left(\{P_0, \mathbb{T} \smallsetminus P_0\} \big| \eta_m\right) \leqslant H_{\mu_{n_k}}(\eta_m) + \log 2.$$

Together with (5.5) this gives

$$\frac{1}{m} H_{\mu_{n_k}}(\eta_m) \geqslant \frac{\log_p n_k}{md_k} \alpha \log p + \frac{1}{md_k} \log C - \frac{1}{d_k} \log p - \frac{1}{m} \log 2,$$

and we may now use the weak* convergence of μ_{n_k} since $\mu(\partial P) = 0$ for all $P \in \eta_m$. Therefore

$$\frac{1}{m} H_\mu(\xi_0^{m-1}) \geqslant \frac{1}{m} H_\mu(\eta_m) \geqslant \alpha \log p - \frac{1}{m} \log 2,$$

which again implies that $h_\mu(T) \geqslant \alpha \log p$ by letting $m \to \infty$. □

5.3.2 The Variational Principle

The first inequality in the proof of Theorem 5.24 below was shown by Goodwyn [74]; Dinaburg [41] proved the full result under the assumption that X has finite covering dimension. Finally Goodman [72] proved the result without any assumptions. The first proof below is a simplification due to Misiurewicz [139], and the second is a strengthened version due to Blanchard, Glasner and Host [15].

Theorem 5.24 (Variational Principle). *Let* $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ *be a continuous map on a compact metric space. Then*

$$h_{\text{top}}(T) = \sup_{\mu \in \mathscr{M}^T(X)} h_\mu(T)$$

A measure $\mu \in \mathscr{M}^T(X)$ with $h_{\text{top}}(X) = h_\mu(T)$ is called a *maximal measure*. A maximal measure does not always exist (see Example 8.6), but does if T is expansive (see Corollary 8.1) or if the map $\mu \mapsto h_\mu(T)$ is upper semicontinuous for some other reason.

Write $\mathscr{E}^T$ for the subset of $\mathscr{M}^T$ consisting of those Borel probability measures on X that are ergodic with respect to T. It is enough to take the supremum over the ergodic measures rather than all the invariant measures in Theorem 5.24 by the ergodic decomposition of entropy and convexity of entropy (see Theorem 2.33 and Exercise 5.3.4).

5.3.3 First Proof of the Variational Principle

PROOF OF THEOREM 5.24: TOPOLOGICAL ENTROPY DOMINATES. Fix a measure $\mu \in \mathscr{M}^T(X)$, and let

$$\xi = \{P_1, \dots, P_k\}$$

be a measurable partition of X. Set $\varepsilon = \frac{1}{k \log k}$. For each $P_j \in \xi$ there is a compact set $Q_j \subseteq P_j$ with $\mu(P_j \smallsetminus Q_j) < \varepsilon$. Define a new partition

$$\eta = \{Q_0, Q_1, \dots, Q_k\}$$

where

$$Q_0 = X \smallsetminus \bigcup_{j=1}^{k} Q_j.$$

Notice that $\mu(Q_0) < k\varepsilon$ and

$$\mu(Q_i \cap P_j) = \begin{cases} \mu(Q_i) & \text{if } 0 < i = j; \\ 0 & \text{if } 0 < i \neq j; \\ < \varepsilon & \text{if } 0 = i < j. \end{cases}$$

It follows that

$$\begin{aligned}
H_\mu\left(\xi\middle|\eta\right) &= \sum_{i=0}^{k} \mu(Q_i) H_{\frac{1}{\mu(Q_i)}\mu|_{Q_i}}(\xi)\\
&= \mu(Q_0)\underbrace{H_{\frac{1}{\mu(Q_0)}\mu|_{Q_0}}(\xi)}_{\leqslant \log k}\\
&\leqslant \varepsilon k \log k = 1
\end{aligned}$$

since $x \in Q_i$ for $i > 0$ implies $x \in P_i$, by the trivial bound of entropy (Proposition 1.5) and by the choice of ε.

By the continuity bound of dynamical entropy (Proposition 1.17(3)) this gives

$$h_\mu(T,\xi) \leqslant h_\mu(T,\eta) + H_\mu(\xi\middle|\eta) < h_\mu(T,\eta) + 1. \tag{5.6}$$

In other words, it remains to estimate $h_\mu(T,\eta)$.

Now notice that the cover number $N\left(\eta_0^{n-1}\right)$ for a partition is simply the number of non-empty elements. Now $\mathscr{U} = \{Q_0 \cup Q_i \mid i = 1,\dots,k\}$ is an open cover of X, and so by the trivial bound of entropy (Proposition 1.5) we obtain

$$\begin{aligned}
H_\mu\left(\eta_0^{n-1}\right) &\leqslant \log N\left(\eta_0^{n-1}\right)\\
&\leqslant \log\left[N\left(\mathscr{U}_0^{n-1}\right)\cdot 2^n\right],
\end{aligned}\tag{5.7}$$

since every element of $\mathscr{U}_0^{n-1}$ is the union of at most 2^n elements of η_0^{n-1}.

By the inequalities (5.6) and (5.7), it follows that

$$h_\mu(T,\xi) \leqslant h_{\text{top}}(T) + \log 2 + 1.$$

Since it suffices to consider finite partitions (Lemma 1.19) we obtain

$$h_\mu(T) \leqslant h_{\text{top}}(T) + \log 2 + 1. \tag{5.8}$$

Now for any $n \geqslant 1$, $h_\mu(T^n) = nh_\mu(T)$ and $h_{\text{top}}(T^n) = nh_{\text{top}}(T)$ by Proposition 1.18(3) and Lemma 5.21, so (5.8) (used for T^n and letting $n \to \infty$) implies that

$$h_\mu(T) \leqslant h_{\text{top}}(T).$$

Since μ was arbitrary, this means that

$$\sup_{\mu\in\mathscr{M}^T(X)} h_\mu(T) \leqslant h_{\text{top}}(T). \tag{5.9}$$

□

The second half of the proof takes a bit longer, and uses the same idea as the proof of Proposition 5.23 does.

Continuation of proof of Theorem 5.24: constructing invariant measures with large entropy. Fix $\varepsilon > 0$ and let F_n be an (n,ε)-

separated set with cardinality $s_n = s_{\text{sep}}(n, \varepsilon)$. Define measures

$$\nu_n = \frac{1}{s_n} \sum_{x \in F_n} \delta_x$$

and

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T_*^j \nu_n.$$

Choose a sequence (n_k) with $n_k \to \infty$ for which

$$\limsup_{n\to\infty} \frac{1}{n} \log s_n = \lim_{k\to\infty} \frac{1}{n_k} \log s_{n_k},$$

and

$$\lim_{k\to\infty} \mu_{n_k} = \mu \in \mathscr{M}^T(X).$$

This may be done by finding a sequence with the first property, and then choosing a subsequence of that sequence with the second property using the weak* compactness property as in [52, Th. 4.1].

Let ξ be a finite measurable partition of X as in Lemma 3.13 with the property that $\mu(\partial P) = 0$ and $\operatorname{diam}(P) < \varepsilon$ for all $P \in \xi$. Then for x and y in the same atom

$$Q \in \bigvee_{i=0}^{n-1} T^{-i}\xi = \xi_0^{n-1}$$

we have

$$\mathsf{d}(T^i x, T^i y) < \varepsilon$$

for $i = 0, \dots, n-1$, so $|Q \cap F_n| \leqslant 1$. From the definition of ν_n, $\nu_n(Q)$ is either 0 or $\frac{1}{s_n}$. It follows that

$$H_{\nu_n}\left(\xi_0^{n-1}\right) = -\sum_Q \underbrace{\nu_n(Q)}_{0 \text{ or } 1/s_n} \log \nu_n(Q) = \log s_n. \tag{5.10}$$

However, to make use of this we need to start working with the (here only almost invariant) measure μ_n. Moreover, just as in the proof of Proposition 5.23 we will need to work with a fixed partition $\eta = \xi_0^{m-1}$ as $n \to \infty$ to make use of the weak* convergence. To this end, fix an integer $m \geqslant 1$ and let $n = dm + r$. Then

$$H_{\mu_n}(\eta) \geqslant \frac{1}{n} \sum_{j=0}^{n-1} H_{T_*^j \nu_n}(\eta) \geqslant \frac{1}{n} \sum_{k=0}^{m-1} \sum_{\ell=0}^{d-1} H_{\nu_n}\left(T^{-(\ell m+k)}\eta\right) \tag{5.11}$$

by convexity of $t \mapsto \phi(t)$ (Lemma 1.4), the formula

$$H_{T_*^j \nu_n}(\eta) = H_{\nu_n}(T^{-j}\eta),$$

and by dropping the last r terms. We wish to give a lower bound of the right-hand side of (5.11) that involves (5.10). For this we note that

$$[0, n-1] \cap \mathbb{Z} = \Bigl([0, k-1] \cup \bigcup_{\ell=0}^{d-2} (\ell m + k + [0, m-1]) \cup [(d-1)m, n-1]\Bigr) \cap \mathbb{Z}$$

for any integer k with $0 \leqslant k < m$. Therefore

$$\xi_0^{n-1} = \xi_0^{k-1} \vee \bigvee_{\ell=0}^{d-2} T^{-(\ell m+k)}\eta \vee \bigvee_{i=(d-1)m}^{n-1} T^{-i}\xi.$$

It follows by the subadditivity of entropy that

$$H_{\nu_n}\left(\xi_0^{n-1}\right) \leqslant \sum_{\ell=0}^{d-1} H_{\nu_n}\left(T^{-(\ell m+k)}\eta\right) + 3m \log |\xi|. \tag{5.12}$$

Thus

$$\begin{aligned}
H_{\mu_n}\left(\xi_0^{m-1}\right) &\geqslant \frac{1}{n} \sum_{k=0}^{m-1} \sum_{\ell=0}^{d-1} H_{\nu_n}\left(T^{-(\ell m+k)}\eta\right) && \text{(by (5.11))}\\
&\geqslant \frac{1}{n} \sum_{k=0}^{m-1} \left(H_{\nu_n}\left(\xi_0^{n-1}\right) - \frac{3m}{n} \log |\xi|\right) && \text{(by (5.12))}\\
&= \frac{m}{n} H_{\nu_n}\left(\xi_0^{n-1}\right) - \frac{3m^2}{n} \log |\xi| \\
&= \frac{m}{n} \log s_n - \frac{3m^2}{n} \log |\xi|. && \text{(by (5.10))}
\end{aligned}$$

Using the sequence (n_k) with $n_k \to \infty$ as before and recalling that

$$s_n = s_{\mathrm{sep}}(n, \varepsilon)$$

for some fixed ε we deduce that

$$H_\mu\left(\xi_0^{m-1}\right) = \lim_{k\to\infty} H_{\mu_{n_k}}\left(\xi_0^{m-1}\right) \geqslant m \limsup_{n\to\infty} \frac{1}{n} \log s_{\mathrm{sep}}(n, \varepsilon)$$

where we used Lemma 5.7 and $\mu(\partial Q) = 0$ for $Q \in \xi_0^{m-1}$. Now let $m \to \infty$ to see that

$$h_\mu(T) \geqslant h_\mu(T, \xi) = \lim_{m\to\infty} \frac{1}{m} H_\mu\left(\xi_0^{m-1}\right) \geqslant \limsup_{n\to\infty} \frac{1}{n} \log s_{\mathrm{sep}}(n, \varepsilon).$$

The measure μ potentially changes as ε changes, but nonetheless we can deduce that

$$\sup_{\mu \in \mathscr{M}^T(X)} h_\mu(T) \geqslant h_{\mathrm{top}}(T),$$

which with (5.9) proves the theorem. □

5.3.4 A Stronger Form of Variational Principle

The proof of the first inequality (5.9) in the variational principle is relatively straightforward. The result in this section gives a stronger local version of the more difficult reverse inequality, which constructs measures with entropy close to the topological entropy.(27)

The variational principle is a global statement about the collection of numbers $\{h_{\mathrm{top}}(T, \mathscr{U})\}$ over all open covers $\mathscr{U}$ and the collection of numbers $\{h_\mu(T, \xi)\}$ over all measures $\mu \in \mathscr{M}^T$ and partitions ξ, but says nothing about possible relationships between the topological entropy with respect to a specific cover and the measure-theoretic entropy with respect to a specific invariant measure and partition. In this section we present a result of Blanchard, Glasner and Host [15] (we follow their proof closely) which gives a local form of the variational principle in the following sense. Given an open cover $\mathscr{U}$, a Borel probability measure μ is constructed with the property that $h_\mu(T, \xi) \geqslant h_{\mathrm{top}}(T, \mathscr{U})$ for any partition ξ that *refines* U (that is, any partition ξ with the property that any atom of ξ is contained in an element of the cover $\mathscr{U}$).

Theorem 5.25 (Blanchard, Glasner, Host). *Let $T : (X, \mathsf{d}) \to (X, \mathsf{d})$ be a continuous map on a compact metric space, and let $\mathscr{U} = \{U_1, \dots, U_d\}$ be a finite open cover of X. Then there is a measure $\mu \in \mathscr{M}^T$ with the property that $h_\mu(T, \xi) \geqslant h_{\mathrm{top}}(T, \mathscr{U})$ for any partition ξ that refines $\mathscr{U}$.*

Corollary 5.26. *If there is a finite open over $\mathscr{U}$ with $h_{\mathrm{top}}(T, \mathscr{U}) = h_{\mathrm{top}}(T)$, then T has a maximal measure μ and a finite partition ξ with*

$$h_\mu(T) = h_\mu(T, \xi).$$

As in Proposition 5.23 and Theorem 5.24, the proof ends up by taking a weak* limit of measures, but in contrast to the situation of Proposition 5.23, we have no set of distinguished points to start with. Instead a counting argument is used to find points whose orbit behavior under the map is sufficiently complex, and the invariant measure is produced as a limit of convex combinations of measures supported on these points.

For the counting argument, we start with a finite alphabet $A = |\mathscr{U}|$ of symbols, and write $|w| = n$ for the length of a word $w = w_1 \dots w_n \in A^n$ of n

symbols. If u is a word of length $k \leqslant n$ and w is a word of length n, then write

$$\mathbf{d}_u(w) = \frac{1}{n-k+1} \bigl|\{i \mid 1 \leqslant i \leqslant n-k+1 \text{ and } w_i \dots w_{i+k-1} = u\}\bigr|$$

for the density of complete occurrences of u in w. Notice the two extreme cases: if u does not occur in w, then $\mathbf{d}_u(w) = 0$ and if (for example) $u = a$ and $w = a^n$ for a single symbol $a \in A$, then $\mathbf{d}_u(w) = 1$. In this way we associate to a given $w \in A^n$ a probability measure (that is, a probability vector) on A^k by assigning the probability $\mathbf{d}_u(w)$ to the word $u \in A^k$. We further define the k-entropy of the word w to be

$$H_k(w) = -\sum_{u \in A^k} \phi(\mathbf{d}_u(w)),$$

the entropy function applied to the probability vector $\bigl(\mathbf{d}_u(w) \mid u \in A^k\bigr)$.

To see why it is reasonable to view $H_k(w)$ as an entropy, notice that if w is a very long piece of a generic point in the full 2-shift with respect to the Bernoulli $(\frac{1}{2}, \frac{1}{2})$-measure then we expect a word u of length $k \ll n$ to appear in w with frequency approximately $\frac{1}{2^k}$. Thus, roughly speaking, we expect that

$$\frac{1}{k} H_k(w) \approx -\frac{1}{k} \sum_{u \in A^k} \phi(2^{-k}) = -\frac{2^k}{k} \phi(2^{-k}) = \log 2$$

as $n \to \infty$. In fact the k-entropy is more closely related to the material on compression and coding from Sections 1.5 and (3.2).

Lemma 5.27 (Counting lemma). *For any $h > 0$, $\varepsilon > 0$ and $k \geqslant 1$,*

$$|\{w \in A^n \mid H_k(w) \leqslant kh\}| \leqslant \mathrm{e}^{n(h+\varepsilon)}$$

for all sufficiently large n (depending on k, ε, and $|A|$).

Roughly speaking, Lemma 5.27 says that there cannot be too many words of length n whose complexity (as measured by the averaged k-entropy $\frac{1}{k} H_k(w)$) is bounded.

Proof of Lemma 5.27 for $k = 1$. Assume first that $k = 1$, so we are counting the appearance of single symbols in w. We may also think of the alphabet A as the set $\{1, 2, \dots, |A|\}$. By counting words w of length n with q_i appearances of the symbol i, we see that

$$|\{w \in A^n \mid H_1(w) \leqslant h\}| = \sum_{\mathbf{q}} \frac{n!}{q_1! \cdots q_{|A|}!}, \tag{5.13}$$

where the sum is taken over those vectors $\mathbf{q} \in \mathbb{N}^{|A|}$ with

$$\sum_{i=1}^{|A|} q_i = n \tag{5.14}$$

and

$$-\sum_{i=1}^{|A|} \phi(q_i/n) \leqslant h.$$

By Stirling's theorem (see [55, Th. 8.6]) there are constants $C_1 \in (0,1)$ and $C_2 > 0$ with

$$C_1\,(N/\mathrm{e})^N\sqrt{N} \leqslant N! \leqslant C_2\,(N/\mathrm{e})^N\sqrt{N} \tag{5.15}$$

for all $N \geqslant 1$. We claim that there is some constant C depending on $|A|$ but not on n with

$$\frac{n!}{q_1!\cdots q_{|A|}!} \leqslant C\mathrm{e}^{-n\sum_{i=1}^{|A|}\phi(q_i/n)} \leqslant C\mathrm{e}^{nh}. \tag{5.16}$$

To see this, we treat each of the factors appearing in (5.15) in turn.

(1) The constant coefficients contribute a factor of $\frac{C_2}{C_1^{|A|}}$.

(2) The term N^N contributes a factor

$$\begin{aligned}\frac{n^n}{\prod_{q_i>0} q_i^{q_i}} &= \exp\left(n\log n - \sum_{q_i>0} q_i \log q_i\right)\\ &= \exp\left(\sum_{q_i>0} q_i(\log n - \log q_i)\right) = \exp\left(-n\sum_{q_i>0}\frac{q_i}{n}\log\left(\frac{q_i}{n}\right)\right)\end{aligned}$$

by (5.14), giving $\mathrm{e}^{-n\sum_{i=1}^{|A|}\phi(q_i/n)}$ since $\phi(0)=0$.

(3) The term e^{-N} contributes a factor $\frac{\mathrm{e}^{-n}}{\mathrm{e}^{-q_1-\cdots-q_{|A|}}} = 1$ by (5.14).

(4) For the term $\sqrt{N}$, notice that

$$n = \sum_{q_i>0} q_i \leqslant |A| \max_{q_i>0} q_i \leqslant |A| \prod_{q_i>0} q_i,$$

giving a factor $\sqrt{|A|}$.

This gives the bound (5.16), with $C = \frac{C_2}{C_1^{|A|}}\sqrt{|A|}$.

The number of terms summed in (5.13) is no more than $(n+1)^{|A|}$, since each symbol in A appears no more than n times. Fixing some $\varepsilon > 0$ we obtain

$$\begin{aligned}|\{w \in A^n \mid H_1(w) \leqslant h\}| &\leqslant C(n+1)^{|A|}\mathrm{e}^{nh} && \text{(by (5.16))}\\ &\leqslant \mathrm{e}^{n(h+\varepsilon)},\end{aligned}$$

for sufficiently large n, as required. $\square$

As we will see, if $\frac{1}{k}H_k(w)$ is small, then it is possible to group the word w into blocks of length k (possibly using an offset j) so that the new word $w^{(j)}$ consisting of roughly $\frac{n}{k}$ blocks of length k has a small value for $H_1(w^{(j)})$, which will allow us to use the case considered above. This may be seen at (5.17) below.

Proof of Lemma 5.27 for $k > 1$. Now assume that $n > 2k$ and $k > 1$. We will think of the word w on two different scales. It is a sequence of symbols of length n; on the other hand it is also (up to remainder terms at the ends) a sequence of $m = \lfloor \frac{n}{k} \rfloor - 1$ words of length k.

For each j, $0 \leqslant j < k$, we define the word $w^{(j)}$ to be the word comprising m words of length k defined by

$$w^{(j)} = |w_{j+1} \dots w_{j+k}|w_{j+k+1} \dots w_{j+2k}| \cdots |w_{j+k(m-1)+1} \dots w_{j+mk}|,$$

viewed as a sequence of length m on the alphabet A^k. Here j is a shift determining which initial and final symbols are discarded when switching from w to $w^{(j)}$.

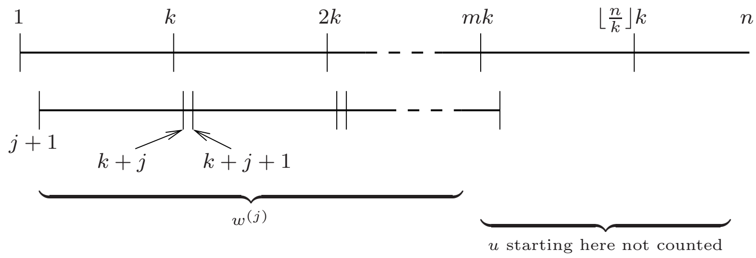

Fig. 5.1: The block $w^{(j)}$ and counting occurrences of u.

Now if $I_u(w)$ denotes the number of positions where an incidence of u starts in w, so that

$$I_u(w) = (n - k + 1)\mathbf{d}_u(w),$$

then we claim that

$$I_u(w) - 2k \leqslant \sum_{j=0}^{k-1} \mathbf{d}_u(w^{(j)})m \leqslant I_u(w).$$

To see the upper bound, notice that every occurrence of u in $w^{(j)}$ is an occurrence in w at some offset j; for the lower bound, notice that occurrences of u near the end of w may be missed (see Figure 5.1). Now $\frac{m}{n} \to \frac{1}{k}$ as $n \to \infty$, so it follows that

$$\left|\mathbf{d}_u(w) - \frac{1}{k}\sum_{j=0}^{k-1}\mathbf{d}_u(w^{(j)})\right| = \mathrm{O}\left(\tfrac{1}{n}\right)$$

as $n \to \infty$. Since ϕ is uniformly continuous, it follows that for large enough n (depending on k and ε) and for any word w of length n,

$$\sum_{u\in A^k}\left|\phi(\mathbf{d}_u(w)) - \phi\left(\frac{1}{k}\sum_{j=0}^{k-1}\mathbf{d}_u(w^{(j)})\right)\right| < \frac{\varepsilon}{2}.$$

By convexity of ϕ (Lemma 1.4), this gives

$$\begin{aligned}\frac{1}{k}\sum_{j=0}^{k-1}H_1\left(w^{(j)}\right) &= -\sum_{u\in A^k}\frac{1}{k}\sum_{j=0}^{k-1}\phi\left(\mathbf{d}_u(w^{(j)})\right)\\ &\leqslant -\sum_{u\in A^k}\phi\left(\sum_{j=0}^{k-1}\frac{1}{k}\mathbf{d}_u(w^{(j)})\right)\\ &\leqslant -\sum_{u\in A^k}\phi\left(\mathbf{d}_u(w)\right) + \frac{\varepsilon}{2} = H_k(w) + \frac{\varepsilon}{2}. \end{aligned}\tag{5.17}$$

It follows that if $H_k(w) \leqslant kh$ then there is some j with $H_1(w^{(j)}) \leqslant kh + \frac{\varepsilon}{2}$. Now for any $j \geqslant 1$ and word $\widetilde{w}$ of length m in the alphabet A^k, there are at most

$$|A|^{n-mk} \leqslant |A|^{2k}$$

words w of length n in the alphabet A for which $w^{(j)} = \widetilde{w}$ (see Figure 5.1). It follows from the case $k = 1$ that

$$\begin{aligned}|\{w \in A^n \mid H_k(w) \leqslant kh\}| &\leqslant |A|^{2k}\sum_{j=0}^{k-1}\left|\left\{\widetilde{w} \in \left(A^k\right)^m \mid H_1(\widetilde{w}) \leqslant kh + \frac{\varepsilon}{2}\right\}\right|\\ &\leqslant |A|^{2k}\sum_{j=0}^{k-1}\mathrm{e}^{m(kh+\varepsilon)}\end{aligned}$$

for sufficiently large n. We deduce that

$$|\{w \in A^n \mid H_k(w) \leqslant kh\}| \leqslant k|A|^{2k}\mathrm{e}^{n(h+\varepsilon/k)} \leqslant \mathrm{e}^{n(h+\varepsilon)}$$

for large n. □

In order to connect the behavior of partitions to the combinatorics of words, we will use the *names* associated to a partition (this idea was also used extensively also in Chapter 3). A finite partition $\xi = \{P_1, \dots, P_d\}$ defines, for each $N \geqslant 1$, a map

$$\mathbf{w}_N^{\xi} : X \to \{1, \dots, d\}^N$$

by requiring that the nth coordinate of $\mathbf{w}_N^{\xi}(x)$ is j if $T^{n-1}x \in P_j$, $1 \leqslant n \leqslant N$.

We will also reduce Theorem 5.25 to the case of a zero-dimensional dynamical system, that is a continuous map on a zero-dimensional compact metric space.[†]

Definition 5.28. Let $T : X \to X$ be a continuous map on a compact metric space. Another system (Y, S), where $S : Y \to Y$ is also a continuous map on a compact metric space, is called a *topological factor* of (X, T) if there is a continuous surjective map $\pi : Y \to X$ with $\pi \circ S = T \circ \pi$. If π is a homeomorphism, then (X, T) is *topologically conjugate* to (Y, S).

Lemma 5.29 (Existence of symbolic cover). *Any topological dynamical system is a topological factor of a zero-dimensional dynamical system.*

Proof. Let $T : X \to X$ be a continuous map on a compact metric space. Since X is compact, we may find a sequence of finite open covers

$$\mathscr{U}_1 \leqslant \mathscr{U}_2 \leqslant \cdots$$

with $\operatorname{diam}(\mathscr{U}_n) \to 0$ as $n \to \infty$. Let a_n be the number of elements in $\mathscr{U}_n$, and fix an enumeration $\mathscr{U}_n = \{O_n^1, \dots, O_n^{a_n}\}$. Define a map θ by sending an element $(x_1, x_2, \dots) \in K_0 = \prod_{n \geqslant 1} \{1, 2, \dots, a_n\}$ (which is a compact metric space in the product topology) to the unique element $x \in X$ with the property that x belongs to the closure $\overline{O_n^{x_n}}$ of the x_nth element of $\mathscr{U}_n$ for all $n \geqslant 1$, if there is such an element. By a straightforward compactness argument, this procedure defines a surjective map from a closed subset $K \subseteq K_0$ onto X.

Now let

$$Y = \{y \in K^{\mathbb{Z}} \mid \theta(y_{n+1}) = T(\theta(y_n)) \text{ for all } n \in \mathbb{Z}\},$$

and define $\pi : Y \to X$ by $\pi(y) = \theta(y_0)$. Since T and θ are continuous, Y is a closed subset of the compact set $K^{\mathbb{Z}}$ with the product topology, and it is clear that π is continuous and onto. Finally, if we write $S : Y \to Y$ for the left shift map defined by $(S(y))_k = y_{k+1}$ for all $k \in \mathbb{Z}$, we have $\pi \circ S = T \circ \pi$. □

Proof of Theorem 5.25: reduction to zero-dimensional case. Assume first that we have proved the theorem for any zero-dimensional topological dynamical system, and let $\pi : Y \to X$ be a topological factor map from a zero-dimensional system $S : Y \to Y$, which exists by Lemma 5.29. Let $\mathscr{V} = \pi^{-1}(\mathscr{U})$ be the pre-image of $\mathscr{U}$, so that $h_{\text{top}}(S, \mathscr{V}) = h_{\text{top}}(T, \mathscr{U})$. By

[†] A topological space is said to be zero-dimensional if there is a basis for the topology comprising sets that are both open and closed (clopen sets). Any discrete space is zero-dimensional, but a zero-dimensional space need not have any isolated points. Examples include $\mathbb{Q}$ in the subspace topology induced from the reals. The examples *ne plus ultra* for dynamics come from shift spaces: for any finite set A with the discrete topology, the spaces $A^{\mathbb{Z}}$, $A^{\mathbb{N}}$, or any closed subset of them, are zero-dimensional.

Theorem 5.25 for zero-dimensional systems, there is a measure $\nu \in \mathscr{M}^S(Y)$ with the property that $h_\nu(S,\eta) \geqslant h_{\mathrm{top}}(S,\mathscr{V})$ for every measurable partition η finer than $\mathscr{V}$. Let $\mu = \pi_*\nu$; then $\mu \in \mathscr{M}^T(X)$, and for any measurable partition ξ finer than $\mathscr{U}$, $\pi^{-1}(\xi)$ is a measurable partition of Y that refines $\mathscr{V}$, so

$$h_\mu(T,\xi) = h_\nu(S,\pi^{-1}(\xi)) \geqslant h_{\mathrm{top}}(S,\mathscr{V}) = h_{\mathrm{top}}(T,\mathscr{U})$$

as required. □

All that remains is to prove the theorem for a zero-dimensional system. We start with a lemma which relates partitions to names and the combinatorial entropy of blocks in those names.

Lemma 5.30 (Blocks of large complexity). *If $\mathscr{U}$ is a finite cover of X and $\{\xi_\ell \mid 1 \leqslant \ell \leqslant K\}$ is a finite list of finite measurable partitions of X each of which refines $\mathscr{U}$, then for any $\varepsilon > 0$ and sufficiently large n, there is an $x \in X$ with*

$$H_k\left(\mathbf{w}_N^{\xi_\ell}(x)\right) \geqslant k\left(h_{\mathrm{top}}(T,\mathscr{U}) - \varepsilon\right)$$

if $1 \leqslant k, \ell \leqslant K$.

Proof of Lemma 5.30. By allowing empty sets and taking unions of atoms that lie in the same element of $\mathscr{U}$, we may assume that all the partitions ξ_ℓ have $d = |\mathscr{U}|$ elements; write $A = \{1,\dots,d\}$. Then, if n is large enough, we have by Lemma 5.27

$$|\Sigma(n,k)| \leqslant \mathrm{e}^{n(h_{\mathrm{top}}(T,\mathscr{U})-\varepsilon/2)}$$

for $1 \leqslant k \leqslant K$, where

$$\Sigma(n,k) = \{w \in A^n \mid H_k(w) < k\left(h_{\mathrm{top}}(T,\mathscr{U}) - \varepsilon\right)\}.$$

By further increasing n if necessary, also assume that

$$\mathrm{e}^{n\varepsilon/2} > K^2.$$

Write

$$\Upsilon(k,\ell) = \left\{x \in X \mid \mathbf{w}_n^{\xi_\ell}(x) \in \Sigma(n,k)\right\};$$

since $\Upsilon(k,\ell)$ is the union of $|\Sigma(n,k)|$ elements of $\bigvee_{j=0}^{n-1} T^{-j}\xi_\ell$, a partition finer than $\bigvee_{j=0}^{n-1} T^{-j}\mathscr{U}$, the set $\Upsilon(k,\ell)$ is covered by

$$|\Sigma(n,k)| \leqslant \mathrm{e}^{n(h_{\mathrm{top}}(T,\mathscr{U})-\varepsilon/2)}$$

elements of $\bigvee_{j=0}^{n-1} T^{-j}\mathscr{U}$. It follows that $\bigcup_{1\leqslant k,\ell\leqslant K}\Upsilon(k,\ell)$ is covered by

$$K^2\mathrm{e}^{n(h_{\mathrm{top}}(T,\mathscr{U})-\varepsilon/2)} < \mathrm{e}^{nh_{\mathrm{top}}(T,\mathscr{U})}$$

elements of $\bigvee_{j=0}^{n-1} T^{-j}\mathscr{U}$. Since

$$h_{\mathrm{top}}(T, \mathscr{U}) \leqslant \frac{1}{n} \log N \left(\bigvee_{j=0}^{n-1} T^{-j} \mathscr{U} \right),$$

by Fekete's lemma (Lemma 1.13, as used in Definition 5.9), any subcover of

$$\bigvee_{j=0}^{n-1} T^{-j} \mathscr{U}$$

has at least $\mathrm{e}^{n h_{\mathrm{top}}(T, \mathscr{U})}$ elements, so we deduce that

$$\bigcup_{1 \leqslant k, \ell \leqslant K} \Upsilon(k, \ell) \neq X,$$

and the point x may be found in the complement. □

Proof of Theorem 5.25: zero-dimensional case. Assume now that X is zero-dimensional and let $\mathscr{U} = \{U_1, \dots, U_d\}$ be the open cover of X. Consider initially the collection Ξ of all partitions ξ of X with the property that

$$\xi = \{P_1, \dots, P_d\}$$

comprises d clopen sets with $P_i \subseteq U_i$ for all $i = 1, \dots, d$ (we will see later how to extend the result to all partitions). The collection Ξ is countable,† so we may enumerate it as $\Xi = \{\xi_\ell \mid \ell \geqslant 1\}$. Using Lemma 5.30, we may find a sequence of integers $n_K \to \infty$ and a sequence (x_K) in X for which

$$H_k \left(\mathbf{w}_{n_K}^{\xi_\ell}(x_K) \right) \geqslant k \left(h_{\mathrm{top}}(T, \mathscr{U}) - \tfrac{1}{K} \right) \tag{5.18}$$

for $1 \leqslant k, \ell \leqslant K$. Define a measure μ_K by

$$\mu_K = \frac{1}{n_K} \sum_{i=0}^{n_K - 1} \delta_{T^i x_K},$$

and (by passing to a subsequence of (n_K) and using the corresponding subsequence of (x_K) with the property (5.18); for brevity we use K again to index the resulting sequences) we can assume that $\mu_K \to \mu$ in the weak* topology.

By [52, Th. 4.1], $\mu \in \mathscr{M}^T(X)$. Fix k and ℓ and let E be any atom of the partition $\bigvee_{j=0}^{k-1} T^{-j} \xi_\ell$ with name $u \in \{1, \dots, d\}^k$. For every K,

$$\left| \mu_K(E) - \mathbf{d}_u(\mathbf{w}_{n_K}^{\xi_\ell}(x_K)) \right| = \mathrm{O}\left(\tfrac{k}{n_K} \right)$$

† To see this, notice that for any $\varepsilon = \frac{1}{n}$ the whole space can be covered by finitely many clopen sets of diameter less than ε. Denote by Ξ_0 the countable collection of finite unions of such sets. Then Ξ_0 is the collection of all clopen sets: any open set is a union of elements of Ξ_0 and a compact open set is a finite union.

by definition of μ_K and of $\mathbf{d}_u$. Since E is clopen, $\mathbb{1}_E$ is continuous, so

$$\mu(E) = \lim_{K\to\infty} \mu_K(E) = \lim_{K\to\infty} \mathbf{d}_u(\mathbf{w}^{\xi_\ell}_{n_K}(x_K)),$$

and therefore

$$\phi\left(\mu(E)\right) = \lim_{K\to\infty} \phi\left(\mathbf{d}_u(\mathbf{w}^{\xi_\ell}_{n_K}(x_K)\right). \tag{5.19}$$

Summing (5.19) over all $u \in \{1,\dots,d\}^k$ and using (5.18) gives

$$H_\mu\left(\bigvee_{j=0}^{k-1} T^{-j}\xi_\ell\right) = \lim_{K\to\infty} H_k\left(\mathbf{w}^{\xi_\ell}_{n_K}(x_K)\right) \geqslant k h_{\mathrm{top}}(T,\mathscr{U}),$$

and letting $k\to\infty$ then gives $h_\mu(T,\xi_\ell) \geqslant h_{\mathrm{top}}(T,\mathscr{U})$ for all $\ell \geqslant 1$.

Finally, since X is zero-dimensional the family Ξ of partitions using clopen sets is dense (with respect to the L^1_μ metric on partitions) in the family of partitions with d atoms each of which is a subset of an atom of $\mathscr{U}$. Together with the continuity of entropy (continuity bound in Proposition 1.17(3) together with Exercise 1.1.5) we also have

$$h_\mu(T,\xi) \geqslant h_{\mathrm{top}}(T,\mathscr{U})$$

for any partition of this shape, proving the theorem. □

Exercises for Section 5.3

Exercise 5.3.1. Let $T_p(x) = px \mod 1$ for $x \in \mathbb{T}$ for some $p \geqslant 2$, and assume that for every n with $\gcd(n,p) = 1$ we choose a subset $S_n \subseteq \{0, \frac{1}{n}, \dots, \frac{n-1}{n}\}$ with $T_p(S_n) \subseteq S_n$ and $|S_n| \geqslant n^{1-\mathrm{o}(1)}$ (that is, with $|S_n| \geqslant n^{1-a_n}$ for some sequence $a_n \to 0$ as $n\to\infty$). Use Proposition 5.23 to show that the sequence of sets (S_n) is equidistributed: for any continuous function $f : \mathbb{T} \to \mathbb{R}$,

$$\frac{1}{|S_n|}\sum_{s\in S_n} f(s) \longrightarrow \int f \,\mathrm{d}m_{\mathbb{T}}$$

as $n \to \infty$.

Exercise 5.3.2. Fill in the details of the compactness argument in the proof of Lemma 5.29.

Exercise 5.3.3. Use the variational principle to prove Lemma 5.22.

Exercise 5.3.4. Strengthen the variational principle above (Theorem 5.24) by showing that if $T : X \to X$ is a continuous map on a compact metric space, then

$$h_{\mathrm{top}}(T) = \sup_{\mu\in\mathscr{E}^T(X)} h_\mu(T).$$

Notes to Chapter 5

(21) Accessible accounts of the Krieger generator theorem from different points of view may be found in the monographs of Rudolph [180] and Parry [161].
(22) (Page 126) Newhouse [146], [147] demonstrates, for any compact manifold M and a C^r map $f : M \to M$, an upper bound on the possible failure of upper semi-continuity in terms of volume growth of pieces of smooth submanifolds of M. These are then combined with work of Yomdin [210], [211] to deduce that the function $f \mapsto h_{\text{top}}(f)$ is upper semi-continuous with respect to the C^∞ topology on the space of C^∞ diffeomorphisms of M, and that the function $\mu \longmapsto h_\mu(f)$ is upper semi-continuous on $\mathscr{M}(M)$ for a fixed $f \in C^\infty(M)$.
(23) (Page 126) The notion of expansiveness (under the name *instability*) was introduced by Utz [200], who found many of its basic consequences. An alternative and equally natural definition is the following: A homeomorphism $T : X \to X$ of a compact metric space (X, d) is called *pointwise expansive* if there is a map $\delta : X \to \mathbb{R}_{>0}$ such that $\mathsf{d}(T^n x, T^n y) \leqslant \delta(x)$ for all $n \in \mathbb{Z}$ implies $x = y$. This shares many of the properties of expansiveness. For example, if T is pointwise expansive then T has finitely many fixed points and at most countably many periodic points; there are no pointwise expansive homeomorphisms of a circle. These observations are due to Reddy [171], who also showed that there are pointwise expansive homeomorphisms that are not expansive.
(24) (Page 126) The definition of expansiveness does not require the space to be compact, and the local isometry $\mathbb{R}^k \to \mathbb{R}^k/\mathbb{Z}^k \cong \mathbb{T}^k$ may be used to show that the toral automorphism corresponding to $A \in \mathrm{GL}_k(\mathbb{Z})$ is expansive if and only if the action of A on $\mathbb{R}^k$ is expansive. Studying the geometry of the action of the Jordan form of the complexification of A gives the result (see Eisenberg [54]). A similar argument will be used in Section 6.3 to compute the topological entropy of toral automorphisms.
(25) (Page 138) The inequality $h_\mu(T) \leqslant h_{\text{top}}(T)$ for all $\mu \in \mathscr{M}^T(X)$ was shown by Goodwyn [74]; Dinaburg [41] then proved Theorem 5.24 under the assumption that $\dim(X) < \infty$, and finally Goodman [72] proved the general case. The variational principle is generalized significantly by Walters [201] using the notion of topological pressure. It is extended in a different direction by Blanchard, Glasner and Host [15], and their result will be described in Section 5.3.4.
(26) (Page 138) The result below is a simple version of a similar phenomena used by Einsiedler, Lindenstrauss, Michel and Venkatesh [48] in the study of periodic orbits on homogeneous spaces arising through a number-theoretic construction.
(27) (Page 145) This result is also presented in Glasner's monograph [69]. A similar local strengthening of the reverse inequality, namely the result that

$$\sup_{\mu \in \mathscr{M}^T} \inf_{\xi \succcurlyeq U} h_\mu(T, \xi) = h_{\text{top}}(T, U)$$

(where the infimum is taken over all partitions that refine the open cover U) was found by Glasner and Weiss [70]. In [15, Prop. 4] an example is given of a dynamical system (X, T) and an open cover U of X with the property that $h_{\text{top}}(T, U) > 0$, but for every ergodic measure $\mu \in \mathscr{M}^T$ we have $h_\mu(T, \xi) = 0$ for some partition ξ that refines U. There are subsequent developments, many of which may be found in a review paper of Glasner and Ye [71].