BEYOND2019 - SPECIFICATION - NOTES FOR LECTURE 1

VAUGHN CLIMENHAGA

Disclaimer. These notes are my preliminary attempt to provide a little more detail for my first lecture during the 2019 *Dynamics Beyond Uniform Hyperbolicity* conference at CIRM. There will be three lectures by myself and then three by Dan Thompson on our joint project to study existence and uniqueness of measures of maximal entropy and equilibrium states in non-uniform hyperbolicity using weakened versions of specification and expansivity. Eventually our notes from all six lectures will be combined into a coherent document that tells the whole story as we understand it so far. In the meantime, these hastily prepared notes will have to do. In particular I stress that these notes make no attempt to survey the vast literature on the use of specification and its relatives to study other properties besides existence and uniqueness, or on the use of other techniques to study existence and uniqueness.

The goal of this first lecture is to describe the general principles behind the use of "decompositions" for shift spaces, which help quantify "obstructions to specification", and to give an application to β-shifts. The discussion of β-shifts did not appear in the lecture itself but is useful reading for anyone new to this area. The outline of the proof of Theorem 3.5 was also omitted from the live lecture, and I will sketch it at the start of the second lecture. That second lecture will also describe "obstructions to expansivity" for partially hyperbolic systems, and formulate a more general uniqueness result. The third lecture will describe further applications, including billiards and geodesic flows.

Preamble: entropy of probability vectors

Consider a probability vector $\vec{p} = (p_1, \dots, p_N)$, where $p_i \geq 0$ and $\sum p_i = 1$. The *entropy* of $\vec{p}$ is $H(\vec{p}) = \sum_i -p_i \log p_i$. It is a calculus exercise to show that $\max_{\vec{p}} H(\vec{p}) = \log N$, and that this is achieved if and only if $p_i = \frac{1}{N}$ for all i; equivalently, if $p_i = p_j$ for all i, j.

Although our results will be in a more complicated dynamical setting, the general principle from this simple example carries through: there is a function called 'entropy' that we wish to maximize; it is maximized at a unique point; and that point is characterized by an equidistribution property. Later, this will appear as the fact that for a broad class of dynamical systems, there is a unique measure of maximal entropy, which satisfies a Gibbs property (equidistribution).

Date: 2019-05-14 15:48:20-05:00.

1. Thermodynamic formalism

Let X be a compact metric space and $f\colon X \to X$ a continuous map. This gives a discrete-time topological dynamical system. Later we will also consider continuous-time systems given by a flow $f_t\colon X \to X$. In both discrete and continuous time, we are often interested in the case when X is a smooth manifold. In discrete time, we will also consider the case when (X, f) is a shift space.

For now we focus on the discrete-time case; the continuous-time case is largely analogous, though there are a few subtleties that we mention later on. Let $\mathcal{M}_f(X)$ denote the space of Borel f-invariant probability measures on X. When f exhibits some hyperbolic behavior, $\mathcal{M}_f(X)$ is typically extremely large – an infinite-dimensional simplex – and it becomes important to identify certain "distinguished measures" in $\mathcal{M}_f(X)$. This includes SRB measures, measures of maximal entropy, and more generally, equilibrium measures.

Now we need some standard definitions; we refer to [DGS76, Wal82, Pet89, VO16] for further details and properties.

Definition 1.1 (Measure-theoretic Kolmogorov–Sinai entropy)**.** Fix $\mu \in \mathcal{M}_f(X)$. Given a countable partition α of X into Borel sets, write

$$H_\mu(\alpha) := \sum_{A\in\alpha} -\mu(A)\log\mu(A) = \int -\log\mu(\alpha(x))\,d\mu(x) \tag{1.1}$$

for the *static entropy* of α, where we write $\alpha(x)$ for the element of α containing x.[1] Given $j \leq k$, the corresponding *dynamical refinement* of α records which elements of α the iterates $f^j x, \dots, f^k x$ lie in:

$$\alpha_j^k = \bigvee_{i=j}^k f^{-i}\alpha \quad\Leftrightarrow\quad \alpha_j^k(x) = \bigcap_{i=j}^k f^{-i}(\alpha(f^i x)). \tag{1.2}$$

A standard short argument shows that

$$H_\mu(\alpha_0^{n+m-1}) \leq H_\mu(\alpha_0^{n-1}) + H_\mu(\alpha_n^{n+m-1}) = H_\mu(\alpha_0^{n-1}) + H_\mu(\alpha_0^{0+m-1}), \tag{1.3}$$

so that the sequence $c_n = H_\mu(\alpha_0^{n-1})$ is subadditive: $c_{n+m} \leq c_n + c_m$. Then Fekete's lemma[2] implies that $\lim \frac{c_n}{n}$ exists. Thus we can define the *dynamical entropy* of α with respect to f to be

$$h_\mu(f,\alpha) := \lim_{n\to\infty} \frac{1}{n} H_\mu(\alpha_0^{n-1}) = \inf_{n\in\mathbb{N}} \frac{1}{n} H_\mu(\alpha_0^{n-1}). \tag{1.4}$$

The *measure-theoretic (Kolmogorov–Sinai) entropy* of (X, f, μ) is

$$h_\mu(f) = \sup_\alpha h_\mu(f,\alpha), \tag{1.5}$$

where the supremum is taken over all partitions α as above for which $H_\mu(\alpha) < \infty$.

[1] One can interpret $H_\mu(\alpha)$ as the expected amount of information gained by observing which partition element a point $x \in X$ lies in.

[2] This result dates back at least to [Fek23] and states that subadditivity implies $\lim \frac{c_n}{n} = \inf \frac{c_n}{n}$; the proof is a short exercise.

The *variational principle* states that

$$\sup_{\mu\in\mathcal{M}_f(X)} h_\mu(f) = h_{\text{top}}(X,f), \tag{1.6}$$

where $h_{\text{top}}(X,f)$ is the *topological entropy* of $f\colon X\to X$, which we will define more carefully below (Definition **??**).

Now we define the central object of our study (at least for the first three lectures).

Definition 1.2 (MMEs)**.** A measure $\mu\in\mathcal{M}_f(X)$ is a *measure of maximal entropy (MME)* for (X,f) if $h_\mu(f) = h_{\text{top}}(X,f)$; equivalently, if $h_\nu(f)\le h_\mu(f)$ for every $\nu\in\mathcal{M}_f(X)$.

For uniformly hyperbolic systems, the following (classical) theorem gives a fairly complete picture regarding equilibrium measures.

Theorem 1.3: Existence and Uniqueness

Suppose that we are in one of the following situations.

(1) $(X, f=\sigma)$ is a transitive shift of finite type (SFT).
(2) $f\colon M\to M$ is a C^1 diffeomorphism and $X\subset M$ is a compact f-invariant topologically transitive locally maximal hyperbolic set.[a]

Then there exists a unique measure of maximal entropy μ for (X,f).

[a] In particular, this holds if $X=M$ is compact and f is a transitive Anosov diffeomorphism.

Remark 1.4. The unique MME can be thought of as the 'most complex' invariant measure for a system, and often encodes dynamically relevant information such as the distribution and asymptotic behavior of the set of periodic points. It is also interesting to maximize the quantity $h_\mu(f)+\int\varphi\,d\mu$ for some *potential function* $\varphi\colon X\to\mathbb{R}$; in the setting of Theorem 1.3, there is a unique maximizing measure, called an *equilibrium state*, whenever φ is Hölder continuous. For example, if f is a $C^{1+\alpha}$ Anosov diffeomorphism (or if X is an Axiom A attractor) then the unique equilibrium state for the *geometric potential* $\varphi(x) = -\log|\det Df|_{E^u(x)}|$ is the physically relevant Sinai–Ruelle–Bowen (SRB) measure.

Remark 1.5. One can also show that the unique MME (or more generally, the unique equilibrium state) has strong ergodic and statistical properties such as Bernoullicity, large deviations, central limit theorem, exponential decay of correlations, etc., but we will not pursue this for the time being.

2. Uniqueness for shift spaces with specification

Following Bowen, we outline a proof of Theorem 1.3 in the first case, when (X,σ) is a transitive SFT. The original construction of the MME in this setting is due to Parry and uses the transition matrix. Bowen's proof works for a broader class of systems, which we now describe.

Fix a finite set A (the *alphabet*), let $\sigma\colon A^{\mathbb{N}} \to A^{\mathbb{N}}$ be the shift map $\sigma(x_1x_2\dots) = x_2x_3\dots$, and let $X \subset A^{\mathbb{N}}$ be closed and σ-invariant: $\sigma(X) = X$. Here $A^{\mathbb{N}}$ (and hence X) is equipped with the metric $d(x,y) = 2^{-\min\{n : x_n \neq y_n\}}$. We refer to X as a *one-sided shift space*.[3] Note that so far we do not assume that X is an SFT or anything of the sort.

Given $x \in A^{\mathbb{N}}$ and $i < j$, we write $x_{[i,j]} = x_i x_{i+1} \cdots x_j$ for the *word* that appears in positions i through j. We use similar notation to denote subwords of a word $w \in A^* := \bigcup_n A^n$. Given w, we write $[w] = \{x \in X : x_{[1,n]} = w\}$ for the *cylinder* of $[w]$. We write

$$\mathcal{L}_n := \{w \in A^n : [w] \neq \emptyset\}, \quad \mathcal{L} := \bigcup_{n \geq 0} \mathcal{L}_n, \tag{2.1}$$

and refer to $\mathcal{L}$ as the *language* of X.

Definition 2.1. The *topological entropy* of X is $h_{\mathrm{top}}(X) = \lim_{n\to\infty} \frac{1}{n} \log \#\mathcal{L}_n$.[4]

It is a simple exercise to verify that every transitive SFT has the following property: there is $\tau \in \mathbb{N}$ such that for every $v, w \in \mathcal{L}$ there is $u \in \mathcal{L}$ with $|u| \leq \tau$ such that $vuw \in \mathcal{L}$. Iterating this, we see that

$$\begin{aligned} &\text{for every } w^1, \dots, w^k \in \mathcal{L} \text{ there are } u^1, \dots, u^{k-1} \in \mathcal{L} \\ &\text{such that } |u^i| \leq \tau \text{ for all } i, \text{ and } w^1u^1w^2u^2 \cdots u^{k-1}w^k \in \mathcal{L}. \end{aligned} \tag{2.2}$$

A shift space whose language satisfies (2.2) is said to have the *specification property*.

Theorem 2.2: Shift spaces with specification (Bowen)

Let (X, σ) be a shift space with the specification property. Then there is a unique measure of maximal entropy on X.

In §§2.1–2.2 we outline the two main steps in the proof of Theorem 2.2.

2.1. **The lower Gibbs bound as the mechanism for uniqueness.** It follows from the Shannon–McMillan–Breiman theorem that if μ is an ergodic shift-invariant measure, then for μ-a.e. x we have

$$-\frac{1}{n} \log \mu[x_{[1,n]}] \to h_\mu(\sigma) \text{ as } n \to \infty.$$

This can be rewritten as

$$-\frac{1}{n} \log \left(\frac{\mu[x_{[1,n]}]}{e^{nh_\mu(\sigma)}} \right) \to 0;$$

in other words, the ratio $\mu[x_{[1,n]}]/e^{nh_\mu(\sigma)}$ behaves subexponentially in n for μ-a.e. x; it does not approach either 0 or ∞ exponentially quickly. The crucial observation vis-à-vis uniqueness is that if this ratio is actually uniformly bounded away from 0,

[3] One could just as well consider two-sided shift spaces by replacing $\mathbb{N}$ with $\mathbb{Z}$, and all the results below would be the same.

[4] In the lecture I often write $h(X)$ for brevity. The limit exists by Fekete's lemma using the fact that $\log \#\mathcal{L}_n$ is subadditive, which we prove in Lemma 2.6 below.

then μ is the unique MME for $(\operatorname{supp}\mu, \sigma)$. The following argument goes back (in a mildly different form) to Bowen [Bow74].

Proposition 2.3. *Let μ be an ergodic σ-invariant measure on $A^{\mathbb{N}}$, and suppose that there is $K \geq 1$ such that for μ-a.e. x and every $n \in \mathbb{N}$, we have the* lower Gibbs bound

$$\mu[x_{[1,n]}] \geq K^{-1} e^{-nh_\mu(\sigma)}. \tag{2.3}$$

Let $X = \operatorname{supp}\mu$. Then μ is the unique measure of maximal entropy for (X, σ).

Remark 2.4. It is more standard to state the lower Gibbs bound as "$\mu[x_{[1,n]}] \geq K^{-1} e^{-nh_{\mathrm{top}}(X,\sigma)}$ for all $x \in X$"; see Proposition 2.5 below. In fact the two formulations are equivalent, as we will see in the proof. I have stated the bound in this form to emphasize the general principle that existence and uniqueness results in this area typically go hand-in-hand with the following phenomenon: there is some quantity that grows or decays at most subexponentially by general principles (such as an ergodic theorem), which in the case of the unique MME can be shown to be bounded away from 0 and ∞.

Proof of Proposition 2.3. We prove that given any $\nu \in \mathcal{M}_\sigma(X)$ with $\nu \neq \mu$, we have $h_\nu(\sigma) < h_\mu(\sigma)$; by the variational principle, this implies that μ is an MME, and thus it is the unique MME. First observe that if $\nu \ll \mu$ then $\nu = \mu$ by ergodicity. Thus by the Lebesgue decomposition theorem if $\nu \neq \mu$ then $\nu = t\nu_1 + (1-t)\nu_2$ where $\nu_1, \nu_2 \in \mathcal{M}_f(X)$ and $\nu_1 \perp \mu$. Since $h_\nu(\sigma) = th_{\nu_1}(\sigma) + (1-t)h_{\nu_2}(\sigma)$, it suffices to prove that $h_\nu(\sigma) < h_\mu(\sigma)$ whenever $\nu \perp \mu$.

To this end, choose $D \subset X$ such that $\mu(D) = 1$ and $\nu(D) = 0$. Since cylinders generate the σ-algebra, there is $\mathcal{D} \subset \mathcal{L}(X)$ such that $\mu(\mathcal{D}_n) \to 1$ and $\nu(\mathcal{D}_n) \to 0$, where $\mu(\mathcal{D}_n) := \mu\big(\bigcup_{w \in \mathcal{D}_n} [w]\big)$. Then writing α for the (generating) partition into 1-cylinders, we have

$$nh_\nu(\sigma) = h_\nu(\sigma^n) = h_\nu(\sigma^n, \alpha_0^{n-1}) \leq H_\nu(\alpha_0^{n-1}) = \sum_{w \in \mathcal{L}_n} -\nu[w] \log \nu[w]. \tag{2.4}$$

We break the sum into two pieces, one over $\mathcal{D}_n$ and one over $\mathcal{D}_n^c = \mathcal{L}_n \setminus \mathcal{D}_n$. Observe that

$$\begin{aligned}
\sum_{w \in \mathcal{D}_n} -\nu[w] \log \nu[w] &= \sum_{w \in \mathcal{D}_n} -\nu[w] \Big(\log \frac{\nu[w]}{\nu(\mathcal{D}_n)} + \log \nu(\mathcal{D}_n) \Big) \\
&= \Big(\nu(\mathcal{D}_n) \sum_{w \in \mathcal{D}_n} -\frac{\nu[w]}{\nu(\mathcal{D}_n)} \log \frac{\nu[w]}{\nu(\mathcal{D}_n)} \Big) - \nu(\mathcal{D}_n) \log \nu(\mathcal{D}_n) \\
&\leq (\nu(\mathcal{D}_n) \log \#\mathcal{D}_n) + 1,
\end{aligned}$$

where the last line uses the fact that $\sum_{i=1}^k -p_i \log p_i \leq \log k$ whenever $p_i \geq 0$, $\sum p_i = 1$, as well as the fact that $-t \log t \leq 1$ for all $t \in [0,1]$. A similar computation holds for $\mathcal{D}_n^c$, and together with (2.4) this gives

$$nh_\nu(\sigma) \leq 2 + \nu(\mathcal{D}_n) \log \#\mathcal{D}_n + \nu(\mathcal{D}_n^c) \log \#\mathcal{D}_n^c. \tag{2.5}$$

Let $Z = \{x : (2.3) \text{ holds for all } n\}$. Since $X = \operatorname{supp}\mu$, for every $w \in \mathcal{L}_n$ there exists $x_w \in Z \cap [w]$, and thus Condition (2.3) gives $\mu[w] \geq K^{-1} e^{-nh_\mu(\sigma)}$. Summing over $\mathcal{D}_n$ gives

$$\mu(\mathcal{D}_n) = \sum_{w \in \mathcal{D}_n} \mu[w] \geq K^{-1} e^{-nh_\mu(\sigma)} \#\mathcal{D}_n \quad \Rightarrow \quad \#\mathcal{D}_n \leq K e^{nh_\mu(\sigma)} \mu(\mathcal{D}_n),$$

and similarly for $\mathcal{D}_n^c$, so (2.5) gives

$$\begin{aligned} nh_\nu(\sigma) &\leq 2 + \nu(\mathcal{D}_n)\big(\log K + nh_\mu(\sigma) + \log \mu(\mathcal{D}_n)\big) \\ &\qquad + \nu(\mathcal{D}_n^c)\big(\log K + nh_\mu(\sigma) + \log \mu(\mathcal{D}_n^c)\big) \\ &= 2 + \log K + nh_\mu(\sigma) + \nu(\mathcal{D}_n) \log \mu(\mathcal{D}_n) + \nu(\mathcal{D}_n^c) \log \mu(\mathcal{D}_n^c). \end{aligned}$$

Rewriting this as

$$n(h_\nu(\sigma) - h_\mu(\sigma)) \leq 2 + \log K + \nu(\mathcal{D}_n) \log \mu(\mathcal{D}_n) + \nu(\mathcal{D}_n^c) \log \mu(\mathcal{D}_n^c),$$

we see that the right-hand side goes to $-\infty$ as $n \to \infty$, since $\nu(\mathcal{D}_n) \to 0$ and $\mu(\mathcal{D}_n) \to 1$, so the left-hand side must be negative for large enough n, which implies that $h_\nu(\sigma) < h_\mu(\sigma)$ and completes the proof. □

2.2. **Building a Gibbs measure.** Now the question becomes how to build an ergodic measure satisfying the lower Gibbs bound. It turns out that when (X, σ) is a shift space with the specification property, the construction of an MME in Proposition **??** does the job. In this setting, that construction takes the following form: let ν_n be any measure on X such that $\nu_n[w] = 1/\#\mathcal{L}_n$ for every $w \in \mathcal{L}_n$, and then consider the measures

$$\mu_n := \frac{1}{n} \sum_{k=0}^{n-1} \sigma_*^k \nu_n = \frac{1}{n} \sum_{k=0}^{n-1} \nu_n \circ \sigma^{-k}. \tag{2.6}$$

A general argument (which appears in the proof of the variational principle, see for example [Wal82, Theorem 8.6]) shows that any weak* limit point of the sequence μ_n is an MME. In fact, one can prove more.

Proposition 2.5. *Let (X, σ) be a shift space with the specification property, let μ_n be given by* (2.6)*, and suppose that $\mu_{n_j} \to \mu$ in the weak* topology. Then μ is σ-invariant, ergodic, and there is $K \geq 1$ such that μ satisfies the following* Gibbs property:[5]

$$K^{-1} e^{-nh_{\text{top}}(X)} \leq \mu[w] \leq K e^{-nh_{\text{top}}(X)} \text{ for all } w \in \mathcal{L}_n. \tag{2.7}$$

We omit the full proof of Proposition 2.5, and highlight only the most important part of the associated counting estimates.

Lemma 2.6. *Let (X, σ) be a shift space with the specification property, with gap size τ. Then for every $n \in \mathbb{N}$, we have*

$$e^{nh_{\text{top}}(X)} \leq \#\mathcal{L}_n \leq Q e^{nh_{\text{top}}(X)}, \quad \text{where } Q = (\tau + 1) e^{\tau h_{\text{top}}(X)}. \tag{2.8}$$

[5]Note that (2.7) implies that $\operatorname{supp}\mu = X$.

Proof. For every $m, n \in \mathbb{N}$, there is an injective map $\mathcal{L}_{m+n} \to \mathcal{L}_m \times \mathcal{L}_n$ defined by $w \mapsto (w_{[1,m]}, w_{[m+1,m+n]})$, so $\#\mathcal{L}_{m+n} \leq \#\mathcal{L}_m \#\mathcal{L}_n$. Iterating this gives

$$\#\mathcal{L}_{kn} \leq (\#\mathcal{L}_n)^k \quad \Rightarrow \quad \frac{1}{kn} \log \#\mathcal{L}_{kn} \leq \frac{1}{n} \log \#\mathcal{L}_n,$$

and sending $k \to \infty$ we get $h_{\text{top}}(X) \leq \frac{1}{n} \log \#\mathcal{L}_n$ for all n, which proves the lower bound. For the upper bound we observe that specification gives a map $\mathcal{L}_m \times \mathcal{L}_n \to \mathcal{L}_{m+n+\tau}$ defined by mapping (v, w) to $vuwu'$, where $u = u(v, w) \in \mathcal{L}$ with $|u| \leq \tau$ is the 'gluing word' provided by the specification property, and u' is *any* word of length $\tau - |u|$ that can legally follow vuw. This map may not be injective because w can appear in different positions, but each word in $\mathcal{L}_{m+n}$ can have at most $(\tau + 1)$ preimages, since v, w are completely determined by $vuwu'$ and the length of u. This shows that

$$\#\mathcal{L}_{m+n+\tau} \geq \frac{1}{\tau + 1} \#\mathcal{L}_m \#\mathcal{L}_n \quad \Rightarrow \quad \#\mathcal{L}_{k(n+\tau)} \geq \Big(\frac{\#\mathcal{L}_n}{\tau + 1} \Big)^k.$$

Taking logs and dividing by $k(n + \tau)$ gives

$$\frac{1}{k(n + \tau)} \#\mathcal{L}_{k(n+\tau)} \geq \frac{1}{n + \tau} \big(\log \#\mathcal{L}_n - \log(\tau + 1) \big).$$

Sending $k \to \infty$ and rearranging gives $\log \#\mathcal{L}_n \leq \log(\tau + 1) + (n + \tau) h_{\text{top}}(X)$. Taking an exponential proves the upper bound. □

With Lemma 2.6 in hand, the idea of Proposition 2.5 is to first prove the bounds on $\mu[w]$ by estimating, for each $n \gg |w|$ and $k \in \{1, \dots, n - |w|\}$, the number of words $u \in \mathcal{L}_n$ for which w appears in position k; see Figure 1. By considering the subwords of u lying before and after w, one sees that there are at most $(\#\mathcal{L}_k)(\#\mathcal{L}_{n-k-|w|})$ such words, as in the proof of Lemma 2.6, and thus the bounds from that lemma give

$$\nu_n(\sigma^{-k}[w]) \leq \frac{(\#\mathcal{L}_k)(\#\mathcal{L}_{n-k-|w|})}{\#\mathcal{L}_n} \leq \frac{Q e^{k h_{\text{top}}(X)} Q e^{(n-k-|w|) h_{\text{top}}(X)}}{e^{n h_{\text{top}}(X)}} = Q^2 e^{-|w| h_{\text{top}}(X, \sigma)};$$

averaging over k gives the upper Gibbs bound, and the lower Gibbs bound follows from a similar estimate that uses the specification property.

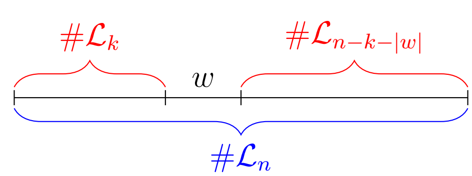

FIGURE 1. Estimating $\nu_n(\sigma^{-k}[w])$.

Next, one can use similar arguments to produce $c > 0$ such that, for each pair of words v, w, there are arbitrarily large $j \in \mathbb{N}$ such that $\mu([v] \cap \sigma^{-j}[w]) \geq c \mu[v] \mu[w]$; this is once again done by counting the number of long words that have v, w in the appropriate positions.

Since any measurable sets V and W can be approximated by unions of cylinders, one can use this to prove that $\overline{\lim}_n \mu(V \cap \sigma^{-n} W) \geq c\mu(V)\mu(W)$. Considering the case when $V = W$ is σ-invariant demonstrates that μ is ergodic.

Together, Propositions 2.3 and 2.5 prove Theorem 2.2, that a shift space with specification has a unique MME.

In later lectures we will discuss extensions of Theorem 2.2 to non-symbolic systems and to equilibrium measures for nonzero potential functions. Now, though, we turn to the central topic of these lectures by describing a non-uniform version of the specification property that also implies uniqueness.

3. Decompositions

Let X be a shift space on a finite alphabet, and $\mathcal{L}$ its language. We consider the following more general version of (2.2).

Definition 3.1. A collection of words $\mathcal{G} \subset \mathcal{L}$ has *specification* if there exists $\tau \in \mathbb{N}$ such that for every finite set of words $w^1, \dots, w^k \in \mathcal{G}$, there are $u^1, \dots, u^{k-1} \in \mathcal{L}$ with $|u^i| \leq \tau$ such that $w^1 u^1 w^2 u^2 \cdots u^{k-1} w^k \in \mathcal{L}$.

The only difference between this definition and (2.2) is that here we only require the gluing property to hold for words in $\mathcal{G}$, not for all words.

Remark 3.2. In particular, $\mathcal{G}$ has specification if there is $\tau \in \mathbb{N}$ such that for every $v, w \in \mathcal{G}$, there is $u \in \mathcal{L}$ with $|u| \leq \tau$ and $vuw \in \mathcal{G}$, because iterating this property gives the one stated above. The property above, which is sufficient for our uniqueness results, is a priori more general because the concatenated word is not required to lie in $\mathcal{G}$.

Now we need a way to say that a collection $\mathcal{G}$ on which specification holds is sufficiently large.

Definition 3.3. A *decomposition* of the language $\mathcal{L}$ consists of three collections of words $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$ with the property that

$$\text{for every } w \in \mathcal{L}, \text{ there are } u^p \in \mathcal{C}^p, v \in \mathcal{G}, u^s \in \mathcal{C}^s \text{ such that } w = u^p v u^s. \tag{3.1}$$

Given a decomposition of $\mathcal{L}$, we also consider for each $M \in \mathbb{N}$ the collection of words

$$\mathcal{G}^M := \{u^p v u^s \in \mathcal{L} : u^p \in \mathcal{C}^p, v \in \mathcal{G}, u^s \in \mathcal{C}^s, |u^p|, |u^s| \leq M\}. \tag{3.2}$$

If each $\mathcal{G}^M$ has specification, then the set $\mathcal{C}^p \cup \mathcal{C}^s$ can be thought of as the set of *obstructions* to the specification property.

Definition 3.4. The *entropy* of a collection of words $\mathcal{C} \subset \mathcal{L}$ is

$$h(\mathcal{C}) = \overline{\lim_{n\to\infty}} \frac{1}{n} \log \#\mathcal{C}_n. \tag{3.3}$$

Theorem 3.5: Uniqueness using a decomposition [CT12]

Let X be a shift space on a finite alphabet, and suppose that the language $\mathcal{L}$ of X admits a decomposition $\mathcal{C}^p\mathcal{G}\mathcal{C}^s$ such that

(I) every collection $\mathcal{G}^M$ has specification, and

(II) $h(\mathcal{C}^p \cup \mathcal{C}^s) < h(X)$.

Then (X, σ) has a unique MME μ.

Remark 3.6. Note that $\mathcal{L} = \bigcup_{M\in\mathbb{N}} \mathcal{G}^M$; the sets $\mathcal{G}^M$ play a similar role to the regular level sets that appear in Pesin theory. The gap size τ appearing in the specification property for $\mathcal{G}^M$ is allowed to depend on M, just as the constants appearing in the definition of hyperbolicity are allowed to depend on which regular level set a point lies in. Similarly, for the unique MME μ one can prove that $\lim_{M\to\infty} \mu(\mathcal{G}^M) = 1$, which mirrors a standard result for hyperbolic measures and Pesin sets.

Remark 3.7. In fact we do not quite need *every* $w \in \mathcal{L}$ to admit a decomposition as in (3.1). It is enough to have $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$ such that $h(\mathcal{L} \setminus (\mathcal{C}^p\mathcal{G}\mathcal{C}^s)) < h(X)$, in addition to the conditions above [Cli18].

We outline the proof of Theorem 3.5. The idea is to mimic Bowen's proof using Propositions 2.3 and 2.5 by completing the following steps.

(1) Prove uniform counting bounds as in Lemma 2.6.

(2) Use these to establish the following *non-uniform* Gibbs property for any limit point μ of the sequence of measures in (2.6): there are constants $K, K_M \geq 1$ such that

$$K_M^{-1} e^{-|w|h_{\text{top}}(X)} \leq \mu[w] \leq K e^{-|w|h_{\text{top}}(X)} \text{ for all } M \in \mathbb{N} \text{ and } w \in \mathcal{G}^M. \tag{3.4}$$

(3) Give a similar argument for ergodicity, and then prove that the non-uniform lower Gibbs bound in (3.4) still gives uniqueness as in Proposition 2.3.

Once the uniform counting bounds are established, the proof of (3.4) follows the same approach as before. The third step, establishing ergodicity and uniqueness, involves some unilluminating technicalities and we will not discuss it further.

For the counting bounds in the first step, we start by observing that the bound $\#\mathcal{L}_n \geq e^{nh_{\text{top}}(X)}$ did not require any hypotheses on X and thus continues to hold. The argument for the upper bound in Lemma 2.6 can be easily adapted to show that there is a constant Q such that $\#\mathcal{G}_n \leq Qe^{nh_{\text{top}}(X)}$ for all n. Then the desired upper bound for $\#\mathcal{L}_n$ is a consequence of the following.

Lemma 3.8. *For any $r \in (0, 1)$, there is M such that $\#\mathcal{G}_n^M \geq r\#\mathcal{L}_n$ for all n.*

Proof. Let $a_i = \#(\mathcal{C}_i^p \cup \mathcal{C}_i^s)e^{-ih_{\text{top}}(X)}$, so that in particular $\sum a_i < \infty$ by (II). Since any $w \in \mathcal{L}_n$ can be written as $w = u^p v u^s$ for some $u \in \mathcal{C}_i^p$, $v \in \mathcal{G}_j$, and $w \in \mathcal{C}_k^s$ with $i + j + k = n$, we have

$$\#\mathcal{L}_n \leq \#\mathcal{G}_n^M + \sum_{\substack{i+j+k=n \\ \max(i,k)>M}} (\#\mathcal{C}_i^p)(\#\mathcal{G}_j)(\#\mathcal{C}_k^s) \leq \#\mathcal{G}_n^M + \sum_{\substack{i+j+k=n \\ \max(i,k)>M}} a_i a_k Q e^{nh_{\text{top}}(X)},$$

where the second inequality uses the upper bound $\#\mathcal{G}_j \leq Qe^{jh_{\mathrm{top}}(X)}$. Since $\sum a_i < \infty$, there is M such that

$$\sum_{\substack{i+j+k=n \\ \max(i,k)>M}} a_i a_k Q e^{nh_{\mathrm{top}}(X)} < (1-r)e^{nh_{\mathrm{top}}(X)} \leq (1-r)\#\mathcal{L}_n,$$

where the second inequality uses the lower bound $\#\mathcal{L}_n \geq e^{nh_{\mathrm{top}}(X)}$. Combining these estimates gives $\#\mathcal{L}_n \leq \#\mathcal{G}_n^M + (1-r)\#\mathcal{L}_n$, which proves the lemma. □

The same specification argument that gives the upper bound on $\#\mathcal{G}_n$ gives a corresponding upper bound on $\mathcal{G}_n^M$ (with a different constant), and thus we deduce the following consequence of Lemma 3.8.

Corollary 3.9. *There are constants $a, A > 0$ and $M \in \mathbb{N}$ such that*

$$e^{nh_{\mathrm{top}}(X)} \leq \#\mathcal{L}_n \leq Ae^{nh_{\mathrm{top}}(X)} \quad \text{and} \quad \#\mathcal{G}_n^M \geq ae^{nh_{\mathrm{top}}(X)} \text{ for all } n \in \mathbb{N}.$$

Remark 3.10. In fact, the proof of Lemma 3.8 can easily be adapted to show a stronger result: given any $\gamma > 0$ and $r \in (0,1)$, there is M such that if $\mathcal{D}_n \subset \mathcal{L}_n$ has $\#\mathcal{D}_n \geq \gamma e^{nh_{\mathrm{top}}(X)}$, then $\#(\mathcal{D}_n \cap \mathcal{G}_n^M) \geq r\#\mathcal{D}_n$. These types of estimates are what lie behind the claim in Remark 3.6 that the (non-uniform) Gibbs property implies $\mu(\mathcal{G}^M) \to 1$ as $M \to \infty$.

4. An example: beta shifts

Given a real number $\beta > 1$, the corresponding β-transformation $f \colon [0,1) \to [0,1)$ is $f(x) = \beta x \pmod 1$. Let $A = \{0, 1, \dots, \lceil \beta \rceil - 1\}$; then every $x \in [0,1)$ admits a coding $y = \pi(x) \in A^{\mathbb{N}}$ defined by $y_n = \lfloor \beta f^{n-1}(x) \rfloor$, and we have $\pi \circ f = \sigma \circ \pi$, where $\sigma \colon A^{\mathbb{N}} \to A^{\mathbb{N}}$ is the left shift. Observe that $\pi(x)_n = a$ if and only if $f^{n-1}(x) \in I_a$, where the intervals I_a are as shown in Figure 2.[6] Given $n \in \mathbb{N}$ and $w \in A^n$, let

$$I(w) := \bigcap_{k=1}^{n} f^{-(k-1)}(I_{w_k})$$

be the interval in $[0,1)$ containing all points x for which the first n iterates are coded by w. The figure shows an example for which $f^n(I(w))$ is not the whole interval $[0,1)$; it is worth checking some other examples and seeing if you can tell for which words $f^n(I(w))$ *is* equal to the whole interval. Observe that if β is an integer then this is true for every word.

Definition 4.1. The *β-shift* X_β is the closure of the image of π, and is σ-invariant. Equivalently, X_β is the shift space whose language $\mathcal{L}$ is the set of all $w \in A^*$ such that $I(w) \neq \emptyset$; thus $y \in A^{\mathbb{N}}$ is in X_β if and only if $I(y_1 \cdots y_n) \neq \emptyset$ for all $n \in \mathbb{N}$.

[6] Formally, $I_a = \{x \in [0,1) : \lfloor \beta x \rfloor = a\}$, so $I_a = [\frac{a}{\beta}, \frac{a+1}{\beta})$ if $a < \lceil \beta \rceil - 1$, and $I_a = [\frac{a}{\beta}, 1)$ if $a = \lceil \beta \rceil - 1$.

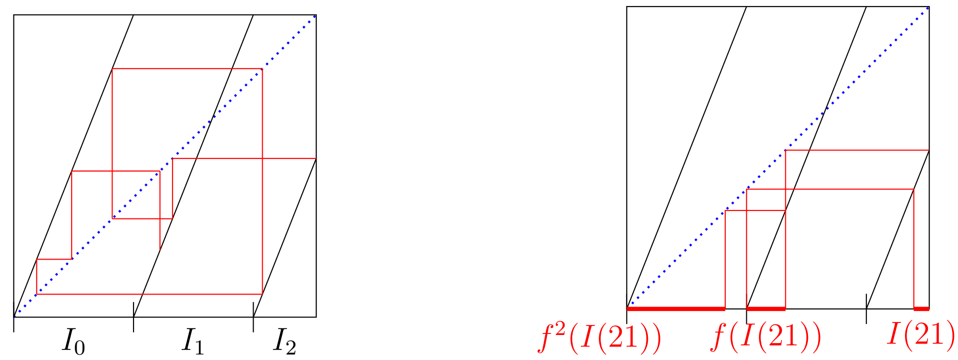

FIGURE 2. Coding a β-transformation.

For further background on the β-shifts, see [Rén57, Par60, Bla89]. We summarize the properties relevant for our purposes.

Write $\preceq$ for the lexicographic order on $A^{\mathbb{N}}$ and observe that π is order-preserving. Let $\mathbf{z} = \lim_{x \nearrow 1} \pi(x)$ denote the supremum of X_β in this ordering. It will be convenient to extend $\preceq$ to A^*, writing $v \preceq w$ if for $n = \min(|v|, |w|)$ we have $v_{[1,n]} \preceq w_{[1,n]}$.

Remark 4.2. Observe that on $A^* \cup A^{\mathbb{N}}$, $\preceq$ is only a pre-order, because there are $v \neq w$ such that $v \preceq w$ and $w \preceq v$; this occurs whenever one of v, w is a prefix of the other.

The β-shift can be described in terms of the lexicographic ordering, or in terms of the following countable-state graph:

- the vertex set is $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$;
- the vertex n has $1 + \mathbf{z}_{n+1}$ outgoing edges, labeled with $\{0, 1, \dots, \mathbf{z}_{n+1}\}$; the edge labeled $\mathbf{z}_{n+1}$ goes to $n + 1$, and the rest go to the 'base' vertex 0.

Figure 3 shows (part of) the graph when $\mathbf{z} = 2102001\dots$, as in Figure 2.

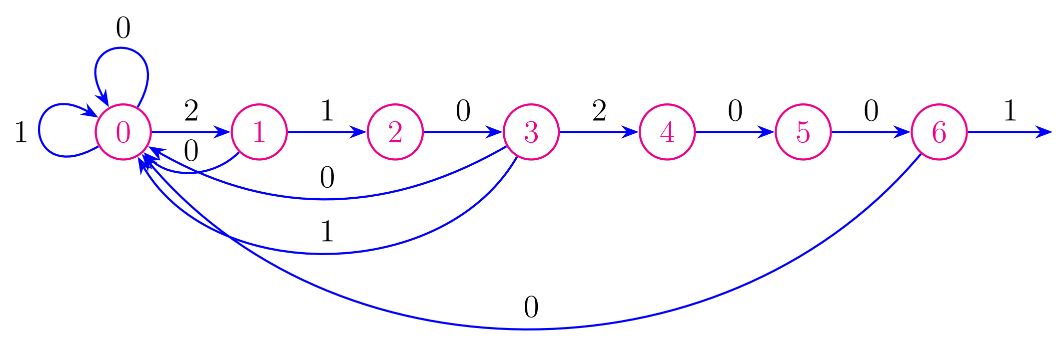

FIGURE 3. A graph representation of X_β.

Proposition 4.3. *Given $n \in \mathbb{N}$ and $w \in A^n$, the following are equivalent.*

(1) $I(w) \neq \emptyset$ (which is equivalent to $w \in \mathcal{L}(X_\beta)$ by definition).

(2) $w_{[j,n]} \preceq \mathbf{z}$ *for every* $1 \leq j \leq n$.
(3) w *labels the edges of a path on the graph that starts at the base vertex* 0.

Idea of proof. Using induction, check that the following are equivalent for every $n \in \mathbb{N}$, $0 \leq k \leq n$, and $w \in A^n$.

(1) $f^n(I(w)) = f^k(I(\mathbf{z}_{[1,k]})$, where we write $I(\mathbf{z}_{[1,0]}) := [0, 1)$.
(2) $w_{[j,n]} \preceq \mathbf{z}$ for every $1 \leq j \leq n$, and k is maximal such that $w_{[n-k+1,n]} = \mathbf{z}_{[1,k]}$.
(3) w labels the edges of a path on the graph that starts at the base vertex 0 and ends at the vertex k. □

Corollary 4.4. *Given* $x \in A^{\mathbb{N}}$*, the following are equivalent.*

(1) $x \in X_\beta$.
(2) $\sigma^n(x) \preceq \mathbf{z}$ *for every* n.
(3) x *labels the edges of an infinite path of the graph starting at the vertex* 0.

Exercise 4.5. Prove that X_β has the specification property if and only if $\mathbf{z}$ does not contain arbitrarily long strings of 0s.

In fact, Schmeling showed [Sch97] that for Lebesgue-a.e. $\beta > 1$, the β-shift X_β does *not* have the specification property. Nevertheless, every β-shift has a unique MME. This was originally proved by Hofbauer [Hof78] and Walters [Wal78] using techniques not based on specification. Theorem 3.5 gives an alternate proof: writing $\mathcal{G}$ for the set of words that label a path starting *and ending* at the base vertex, and $\mathcal{C}^s$ for the set or words that label a path starting at the base vertex *and never returning to it*, one quickly deduces the following.

- $\mathcal{G}\mathcal{C}^s$ is a decomposition of $\mathcal{L}$.
- $\mathcal{G}^M$ is the set of words labeling a path starting at the base vertex and ending somewhere in the first M vertices; writing τ for the maximum graph distance from such a vertex to the base vertex, $\mathcal{G}^M$ has specification with gap size τ.
- $\#\mathcal{C}^s_n = 1$ for every n, and thus $h(\mathcal{C}^s) = 0 < h_{\text{top}}(X_\beta) = \log \beta$.

This verifies the conditions of Theorem 3.5 and thus provides another proof of uniqueness of the MME.

Remark 4.6. Because the earlier proofs of uniqueness did not pass to subshift factors of β-shifts, it was for several years an open problem (posed by Klaus Thomsen) whether such factors still had a unique MME. The inclusion of this problem in Mike Boyle's article "Open problems in symbolic dynamics" [Boy08] was our original motivation for studying uniqueness using non-uniform versions of the specification property, which led us to formulate the conditions in Theorem 3.5; these can be shown to pass to factors, providing a positive answer to Thomsen's question [CT12].

5. Periodic points

It is often the case that one can prove a stronger version of specification, for example, when X is a mixing SFT.

Definition 5.1. Say that $\mathcal{G} \subset \mathcal{L}$ has *periodic specification* if there exists $\tau \in \mathbb{N}$ such that for all $w^1, \dots, w^k \in \mathcal{G}$, there are $u^1, \dots, u^k \in \mathcal{L}_\tau$ such that $v := w^1 u^1 \cdots w^k u^k \in \mathcal{L}$, and moreover $x = vvvvv \cdots \in X$.

There are two strengthenings of specification here; first, we are assuming that the gap size is equal to τ, not just $\leq \tau$, and second, we are assuming that the "glued word" can be extended periodically after adding τ more symbols.

If we replace specification in Theorem 3.5 with periodic specification for each $\mathcal{G}^M$, then the counting estimates in Lemma 2.6 immediately lead to the following estimates on the number of periodic points: writing $\mathrm{Per}_n = \{x \in X : \sigma^n x = x\}$, we have[7]

$$C^{-1} e^{n h_{\mathrm{top}}(X)} \leq \# \mathrm{Per}_n \leq C e^{n h_{\mathrm{top}}(X)}. \tag{5.1}$$

Using this fact and the construction of the unique MME given just before Proposition 2.5, one can also conclude that the unique MME μ is the limiting distribution of periodic orbits in the following sense:

$$\frac{1}{\# \mathrm{Per}_n} \sum_{x \in \mathrm{Per}_n} \delta_x \xrightarrow{\text{weak*}} \mu \text{ as } n \to \infty. \tag{5.2}$$

References

[Bla89] F. Blanchard, *β-expansions and symbolic dynamics*, Theoret. Comput. Sci. **65** (1989), no. 2, 131–141. MR 1020481

[Bow74] Rufus Bowen, *Maximizing entropy for a hyperbolic flow*, Math. Systems Theory **7** (1974), no. 4, 300–303. MR 0385928

[Boy08] Mike Boyle, *Open problems in symbolic dynamics*, Geometric and probabilistic structures in dynamics, Contemp. Math., vol. 469, Amer. Math. Soc., Providence, RI, 2008, pp. 69–118. MR 2478466

[Cli18] Vaughn Climenhaga, *Specification and towers in shift spaces*, Comm. Math. Phys. **364** (2018), no. 2, 441–504. MR 3869435

[CT12] Vaughn Climenhaga and Daniel J. Thompson, *Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors*, Israel J. Math. **192** (2012), no. 2, 785–817. MR 3009742

[DGS76] Manfred Denker, Christian Grillenberger, and Karl Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Mathematics, Vol. 527, Springer-Verlag, Berlin-New York, 1976. MR 0457675

[Fek23] M. Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Math. Z. **17** (1923), no. 1, 228–249. MR 1544613

[Hof78] Franz Hofbauer, *β-shifts have unique maximal measure*, Monatsh. Math. **85** (1978), no. 3, 189–198. MR 0492180

[Par60] W. Parry, *On the β-expansions of real numbers*, Acta Math. Acad. Sci. Hungar. **11** (1960), 401–416. MR 0142719

[Pet89] Karl Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989, Corrected reprint of the 1983 original. MR 1073173

[Rén57] A. Rényi, *Representations for real numbers and their ergodic properties*, Acta Math. Acad. Sci. Hungar **8** (1957), 477–493. MR 0097374

[7]Of course, this is also true in the classical Theorem 2.2.

[Sch97] Jörg Schmeling, *Symbolic dynamics for β-shifts and self-normal numbers*, Ergodic Theory Dynam. Systems **17** (1997), no. 3, 675–694. MR 1452189

[VO16] Marcelo Viana and Krerley Oliveira, *Foundations of ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 151, Cambridge University Press, Cambridge, 2016. MR 3558990

[Wal78] Peter Walters, *Equilibrium states for β-transformations and related transformations*, Math. Z. **159** (1978), no. 1, 65–88. MR 0466492

[Wal82] ———, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR 648108