Strong approximations of shifts and the characteristic measures problem

Van Cyr[1] · Bryna Kra[2] · Samuel Petite[3]

Abstract

Every symbolic system supports a Borel measure that is invariant under the shift, but it is not known if every such systems supports a measure that is invariant under all of its automorphisms, known as a characteristic measure. We give sufficient conditions to find a characteristic measure, additionally showing when it can be taken to be a measure of maximal entropy. The class of systems to which these sufficient conditions apply is large, containing a dense G_δ set in the space of all shifts on a given alphabet, and is also large in the sense that it is closed under taking factors. We also investigate natural systems to which these sufficient conditions apply.

Keywords Subshift · Block complexity · Automorphism · Invariant measure

Mathematics Subject Classification 37B10 · 68R15 · 37A15

1 Introduction

By the Kryloff–Bogoliouboff Theorem [17], any topological dynamical system (X, T) supports a T-invariant Borel probability measure. A natural question is when such a system supports a measure, called characteristic by Frisch and Tamuz [13], that is invariant under the symmetries associated to the system, namely invariant under every automorphism of the system. We note that this question is topological in nature. Given a topological system and a given Borel measure, the property of being a characteristic measure is not a measure theoretic

✉ Samuel Petite
samuel.petite@u-picardie.fr

Van Cyr
van.cyr@bucknell.edu

Bryna Kra
kra@math.northwestern.edu

[1] Bucknell University, Lewisburg, PA 17837, USA

[2] Northwestern University, Evanston, IL 60208, USA

[3] Université de Picardie Jules Verne, Amiens 80039, France

isomorphism invariant: the Jewett–Krieger Theorem implies that any ergodic measure preserving system is measure theoretically isomorphic to a uniquely ergodic topological system, and hence to a system having a characteristic measure.

Frisch and Tamuz [13] proved that every zero entropy symbolic system (on a finite alphabet) has such a measure, while also observing that not every topological dynamical system does. Specifically, the identity map acting on the Cantor set does not. Symbolic systems, however, have special features not shared by all topological systems and results in the literature show that symbolic systems often have a characteristic measure: any mixing shift of finite type does [19], any shift with zero entropy does [13], and any shift that contains periodic points supports such a measure as well. With these observations in mind, Frisch and Tamuz asked if every symbolic system has a characteristic measure [13]. The goal of the present work is to make progress on this problem.

In an approach to the characteristic measures problem, a new class of shifts, the language stable shifts, is defined in [10], where it was shown that each such shift supports a characteristic measure of maximal entropy. The class of language stable shifts is large in the sense that it forms a dense G_δ in the space of all shifts on a fixed alphabet, as well as in the subspace of shifts whose entropy is bounded from below by some $h \geq 0$. It also contains all shifts of finite type. One might ask for other senses in which the class of language stable shifts is large and, specifically we consider if this class is closed under taking factors, and a first result proved in Sect. 3 shows that it is not.

Theorem 1.1 *Any shift that is both language stable and is sofic is a shift of finite type.*

Thus any sofic shift that is not a shift of finite type is not language stable, but is a factor of a language stable shift. This motivates our introduction in Sect. 4 of a class of shifts that are sufficiently well approximated by shifts of finite type in order that properties of characteristic measures are preserved.

Returning to the characteristic measures question we note that sofic shifts are known to have such measures for other reasons (for example, because they have periodic points or, in the irreducible case, because they are intrinsically ergodic). But these properties are not shared by all factors of language stable shifts, and even language stable shifts themselves need not have periodic points or only finitely many ergodic measures of maximal entropy. This naturally leads us to the question of whether every subshift factor of a language stable shift has a characteristic measure and whether this measure can be taken to be a measure of maximal entropy. To investigate these questions, in Sect. 4 we define a notion that we call well-approximable language stable shifts, quantifying how well the shift can be approximated by shifts of finite type, loosely analogous to how well a real number can be approximated by rationals of a given denominator. We show that well-approximable language stable shifts remain a large set, in the sense that they continue to contain a dense G_δ in the space of all shifts (see Proposition 5.2), and that symbolic factors of such shifts have characteristic measures of maximal entropy (see Theorem 4.6 for the precise statement).

Theorem 1.2 *Every subshift factor of a well-approximable language stable shift has a characteristic measure that is a measure of maximal entropy.*

We give an effective rate of approximation by subshifts of finite type that guarantees the existence of a characteristic measure (but not necessarily a characteristic measure of maximal entropy) in Theorem 4.6. We also give a faster (and ineffective) rate of approximation that guarantees the existence of a characteristic measure of maximal entropy in Proposition 4.12.

If every shift were a factor of a well-approximable language stable shift, we would then be able to conclude that every symbolic system supports a characteristic measure. Although

we do not rule this possibility out entirely, in Sect. 5.1, we construct an example of a shift that is not a factor of any sufficiently well-approximable language stable shift. However, we do not answer the question if every subshift factor of a language stable shift has a characteristic measure.

Theorem 1.1 illustrates that the class of language stable shifts are not closed under passage to factors. However, this class is still large in other dynamical senses and this is the content of Sect. 6. For example, we show that the class of language stable shifts is closed under several natural operations, such as passing to a power of the shift or more generally passing to any speed-up. We continue the section by exhibiting language stable shifts in other well-studied classes of shifts, and include a characterization of which β-shifts are language stable. Most of these examples have positive entropy, and in Sect. 7, we show that the class of language stable shifts is also large in zero entropy. In particular, we show that any aperiodic linear complexity shift is language stable.

This leaves open numerous questions about language stable shifts and their characterizations, including Question 7.5 on the complexity threshold for a shift that is not language stable and questions in Sect. 5.2 on quantitative bounds for well-approximable shifts.

2 Classes of shifts and their properties

2.1 Symbolic systems and subshifts

Assume that $\mathcal{A}$ is a finite set, the *alphabet*, and we denote $x \in \mathcal{A}^{\mathbb{Z}}$ by $x = (x_n)_{n \in \mathbb{Z}}$. The space $\mathcal{A}^{\mathbb{Z}}$ is a compact metric space when endowed with the metric d defined by

$$d\big((x_n), (y_n)\big) = 2^{-\inf\{|n| : x_n \neq y_n\}}.$$

The *left shift* $\sigma \colon \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ defined as $(\sigma x)_n = x_{n+1}$ for all $n \in \mathbb{Z}$ is a homeomorphism of $\mathcal{A}^{\mathbb{Z}}$. If $X \subset \mathcal{A}^{\mathbb{Z}}$ is a (non-empty) closed and σ-invariant, then the pair (X, σ) is a *shift*, sometimes also referred to as a *subshift*. We also sometimes omit the transformation σ from the notation and refer to X as a shift.

The *cylinder sets*, meaning the sets where finitely many coordinates are determined, form a basis for the topology on the space X.

2.2 Presenting a shift via its language

If $\mathcal{A}$ is a finite alphabet, $\mathcal{A}^*$ denotes all finite words in the alphabet $\mathcal{A}$. The concatenation of two words $u, v \in \mathcal{A}^*$ is denoted by uv and the word u is called a *prefix* of uv and the word v is a *suffix* of uv. If u is nonempty, then we say that it is a *strict* prefix, and similarly when v is nonempty it is a *strict* suffix. We say that the word v *occurs* in the finite word uv and use the same terminology in infinite words. For $u \in \mathcal{A}^*$, we let $|u|$ denote the length of the word u, and for $n \geq 1$, the word u^n denotes the word of length $n|u|$ obtained by concatenating u with itself n times.

If $\mathcal{F} \subset \mathcal{A}^*$, then the shift $X_{\mathcal{F}}$ associated to the collection $\mathcal{F}$ of *forbidden words* is defined by

$$X_{\mathcal{F}} = \{x \in \mathcal{A}^{\mathbb{Z}} \colon \text{ no subword of } x \text{ belongs to } \mathcal{F}\}.$$

Conversely, any shift X is defined by a collection of forbidden words: if

$$\mathcal{F}(X) := \{w \in \mathcal{A}^* : w \text{ does not occur in any element of } X\}$$

then $X = X_{\mathcal{F}(X)}$.

For the shift (X, σ), the set $\mathcal{A}^* \setminus \mathcal{F}(X)$ of all allowable words is the *language* of the shift and is denoted by $\mathcal{L}(X)$. The word $w \in \mathcal{L}(X)$ is *left special* if there exist distinct letters $a, b \in \mathcal{A}$ such that both $aw, bw \in \mathcal{L}(X)$, and similarly if there exist distinct letters $a, b \in \mathcal{A}$ such that $wa, wb \in \mathcal{L}(X)$, we say that it is *right special*. A word is *bispecial* if it is both left and right special.

We write $\mathcal{L}_n(X)$ for the words of length n in the language $\mathcal{L}(X)$, meaning that $\mathcal{L}_n(X) = \mathcal{L}(X) \cap \mathcal{A}^n$. The *complexity* of a shift X is defined to be function $p_X \colon \mathbb{N} \to \mathbb{N}$ given by $p_X(n) = |\mathcal{L}_n(X)|$ and the exponential growth rate of $p_X(n)$ is the *(topological) entropy* of the shift (X, σ) and is denoted by $h_{\text{top}}(X)$.

Any word $w \notin \mathcal{L}(X)$ is said to be *not allowed* in X. When the set $\mathcal{F}$ of forbidden words is finite, the shift $(X_{\mathcal{F}}, \sigma)$ is known as a *shift of finite type*. More generally, a *sofic shift* (X, σ) is a shift that admits a *rational set* $\mathcal{F} \subset \mathcal{A}^*$ of forbidden words, meaning that the collection is recognized by a finite state deterministic automaton (see [18, Chap. 3] for further background). This means that the allowable words of X can be described in terms of paths on the follower set graph.

Given a shift (X, σ), define $\mathcal{F}_n := \mathcal{F}(X) \cap \mathcal{A}^n$ (the forbidden words in X with length at most n) and let $X_n := X_{\mathcal{F}_n}$ be the shift of finite type defined by the set of forbidden words $\mathcal{F}_n$. Then we can write $X = \bigcap_{n=1}^{\infty} X_n$ as the intersection of a canonical, descending chain of nested shifts of finite type and we call the sequence of shifts $(X_n)_{n \in \mathbb{N}}$ the *SFT cover* of X. Note that X is a shift of finite type if and only if $X = X_n$ for all sufficiently large $n \in \mathbb{N}$.

For a word $w \in \mathcal{L}(X)$, we let $[w]$ denote the cylinder set determined by the word $w = w_1 \cdots w_{|w|}$, meaning the collection of all $x \in X$ such that $x_i = w_i$ for all $1 \leq i \leq |w|$.

2.3 Language stable shifts

We recall a notion introduced in [2]. If X is a shift with forbidden words $\mathcal{F}(X)$, let

$$\mathcal{M}(X) = \{w \in \mathcal{F}(X) : \text{no proper subword of } w \text{ lies in } \mathcal{F}(X)\}$$

denote the *minimal forbidden words* in the language $\mathcal{L}(X)$ of the shift X. We note that $\mathcal{L}(X) = \mathcal{A}^* \setminus \mathcal{A}^* \mathcal{M}(X) \mathcal{A}^*$, meaning that the allowed words are all the words that do not contain a minimal forbidden word. So the sets $\mathcal{F}(X)$ and $\mathcal{M}(X)$ define the same shift, meaning that $X_{\mathcal{F}(X)} = X_{\mathcal{M}(X)}$.

We note that if $a, b \in \mathcal{A}$ and the word aub is a minimal forbidden word of X, then the word u is bispecial (see [2, Remark 3]).

For each $n > 0$, set $\mathcal{M}_n(X) = \mathcal{M}(X) \cap \mathcal{A}^n$ to be the set of minimal forbidden word of length n. A shift (X, σ) is *language stable* if the set of lengths

$$LS(X) = \{n \in \mathbb{N} : \mathcal{M}_n(X) \neq \emptyset\} \tag{1}$$

has zero lower uniform density, meaning that

$$\lim_{k \to \infty} \inf_{n \in \mathbb{N}} \frac{|LS(X) \cap [n+1, n+k]|}{k} = 0.$$

Equivalently, if X is a shift and $(X_n)_{n \in \mathbb{N}}$ is its SFT cover, then X is language stable if and only if for all $k \in \mathbb{N}$ there exists $n \in \mathbb{N}$ with $X_n = X_{n+k}$.

We introduce a new definition to quantify the relationship between the parameters n and k in the definition of language stable.

Definition 2.1 Suppose $\alpha \colon \mathbb{N} \to \mathbb{N}$ is non-decreasing and unbounded. A shift X, with SFT cover $(X_n)_{n \in \mathbb{N}}$ is *well-approximable at rate α* if there are infinitely many n such that $X_n = X_{n+\alpha(n)}$.

Thus the set of language stable shifts is the union over all non-decreasing and unbounded functions α of the shifts that are well-approximable at rate α. In theorems about well-approximable shifts, we are explicit about the rate α, but when speaking more informally about results we sometimes just call a shift well-approximable with the understanding that it is well-approximable at some (typically fast) rate.

We include an example showing that this notion of well-approximable is not vacuous, as up to finitely many elements, any set of lengths $LS(X)$ arises.

Example 2.2 For any set $L \subset \mathbb{N}$, there is a subshift X such that the set of lengths of its minimal forbidden words $LS(X)$ is all but finitely many elements of L. To check this, it suffices to consider the subshift on 3 letters $\{0, 1, 2\}$ with words of the form $\{01^n0 : n \in L \cap 2\mathbb{N}\} \cup \{02^n0 : n \in L \cap (2\mathbb{N}+1)\}$ forbidden. Since the words of the form 01^n0 for odd n are allowed, such words with even n are minimal forbidden words; the analogous property holds for words of the form 02^n0.

2.4 Topological properties of shifts and automorphisms

The shift (Y, σ) is a *factor* of the shift (X, σ) if there exists a surjective and continuous map $\pi \colon X \to Y$ satisfying $\pi \circ \sigma = \sigma \circ \pi$, and if additionally the map π is injective, we say that the systems are *conjugate*. This framework gives us the useful dynamical characterization that a shift is sofic if and only if is a factor of a shift of finite type (see for example [18, Chap. 3]).

We note that the class of language stable shifts is stable under conjugacy, meaning that any subshift conjugate to a language stable shift is also language stable (see [2, 10]). However being language stable is not preserved under passage to a factor, as shown for sofic shifts in Theorem 1.1. The shift (X, σ) is *mixing* if for all words $u, v \in \mathcal{L}(X)$, there exists some $N \in \mathbb{N}$ such that for all $n \geq N$, there is some $w \in \mathcal{L}_n(X)$ such that $uwv \in \mathcal{L}(X)$.

An *automorphism* of the shift (X, σ) is a self conjugacy of the system and we denote the group of all automorphisms of the system by $\mathrm{Aut}(X, \sigma)$, or just $\mathrm{Aut}(X)$ when the shift is clear from the context. The classic theorem of Curtis, Hedlund, and Lyndon (see [14]) states that any factor ϕ of a shift (X, σ) is a block code of some range R for some $R \in \mathbb{N}$, meaning that for all $x \in X$, the symbol that $\phi(x)$ assigns to 0 is determined by the word $x_{-R} \cdots x_R$ in $x = (x_n)_{n \in \mathbb{Z}}$. A block code of range R is a block code of range S for any $S > R$. The same holds for any automorphism, and without loss we assume throughout that the range of any block code arising from an automorphism ϕ is also the range for ϕ^{-1}, and we refer to such a range as being *symmetric*.

2.5 Invariant measures

By the Kryloff–Bogoliouboff Theorem [17], there is always an invariant measure on a shift (X, σ). If the entropy of this measure is equal to the topological entropy of the system, then this is a *measure of maximal entropy*.

A measure that is invariant under the full automorphism group $\mathrm{Aut}(X, \sigma)$ is said to be a *characteristic measure*.

3 Strictly sofic shifts are not language stable

It is shown in [10] that every language stable shift has a characteristic measure of maximal entropy. This motivates the question of which shifts are not language stable, as these are the shifts for which the characteristic measures problem remains open. Our first result shows that the familiar class of sofic shifts is comprised almost entirely of shifts that are not language stable. More precisely, we show that the intersection of the classes of sofic shifts and language stable shifts is exactly the shifts of finite type.

***Proof of Theorem* 1.1** Let (X, σ) be a sofic shift and let $\mathcal{F}$ be a regular language of forbidden words of the shift X.

For contradiction, assume that (X, σ) is not a shift of finite type. Then for any $\ell_0 > 0$, there exist $\ell \geq \ell_0$ and a word $w \in \mathcal{F}$ of length ℓ that does not contain some proper word in $\mathcal{F}$ as a subword. Thus we can pick some such word w whose length is greater than the number of states of the finite state deterministic automaton. The word w corresponds to some vertex path $v_0, \ldots, v_\ell$, starting with some initial vertex v_0 and ending with the vertex v_ℓ, and by choice of the length of the word it passes at least twice through some vertex. Thus there exist $i \neq j$ such that $v_i = v_j$. However, no v_k is an accepting state of the automaton, since the word $v_0 \cdots v_{k-1}$ is not one of the forbidden words in the collection $\mathcal{F}$. Similarly, the path associated to the word $v_1 \cdots v_k$ does not end in an accepting state. It follows that for any $n \geq 0$, the edge path associated the path $v_0, \ldots, v_{i-1}, (v_i, \ldots, v_j)^n, v_{j+1}, \ldots, v_\ell$ belongs to $\mathcal{F}$ but no subword belongs to $\mathcal{F}$. In other words, the set of lengths of minimal forbidden words has positive density and so the shift (X, σ) is not language stable. □

4 Characteristic measures for well-approximable language stable shifts

4.1 Obtaining a characteristic measure using periodic points

The goal of this section is to show that factors of well-approximable shifts have characteristic measures, provided the shifts are well-approximable at a sufficiently fast rate.

Definition 4.1 Let $a, f \in \mathbb{N}$ and let $\mathcal{X}_{a,f}$ denote the set of all subshifts of finite type $X \subseteq \{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ that can be defined using only forbidden words of length at most f. Let $R \in \mathbb{N}$ and let $\mathcal{Y}_{a,f,R}$ denote the set of all sofic shifts $Y \subseteq \{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ for which there exist $X \in \mathcal{X}_{a,f}$ and a factor map $\varphi \colon X \to Y$ that can be implemented using a block code of range at most R.

It follows immediately from the definitions that up to renaming the letters of the alphabet, every shift of finite type is an element of $\mathcal{X}_{a,f}$ for some choice of $a, f \in \mathbb{N}$. We check that the same holds for any sofic shift, and this allows us to assume in this definition, that both the shift of finite type X and the sofic shift Y are defined on the same alphabet. In particular, this shows that any sofic shift lies in some $Y_{a,f,R}$ for appropriate choices of the parameters.

Lemma 4.2 *Up to renaming the letters of the alphabet, for any sofic shift Y there exist $a, f, R \in \mathbb{N}$ such that $Y \in \mathcal{Y}_{a,f,R}$.*

Proof By renaming the letters of the alphabet if necessary, we can assume that $Y \subseteq \{0, 1, \ldots, a_Y - 1\}^{\mathbb{Z}}$ for some $a_Y \in \mathbb{N}$. Since Y is sofic, there exists a shift of finite type X and a factor map $\varphi \colon X \to Y$. Fixing some finite set of forbidden words such that X is the shift of finite type obtained by forbidding these words, define f to be the maximal length of the words in this set. Again renaming the letters of the alphabet if necessary, we can choose $a_X \in \mathbb{N}$ such that $X \subseteq \{0, 1, \ldots, a_X - 1\}^{\mathbb{Z}}$. Setting $a := \max\{a_X, a_Y\}$, we have that $X, Y \subseteq \{0, 1, \ldots, a - 1\}^{\mathbb{Z}}$. Then, by the Curtis-Hedlund-Lyndon Theorem, there exists $R \in \mathbb{N}$ such that φ can be implemented by a range R block code and so $Y \in \mathcal{Y}_{a,f,R}$. $\square$

Note that $\mathcal{Y}_{a,f,R} \subseteq \mathcal{Y}_{a',f',R'}$ provided $a \leq a'$, $f \leq f'$, and $R \leq R'$. Combining this and Lemma 4.2, it follows that the set $\bigcup_{n=1}^{\infty} \mathcal{Y}_{n,n,n}$ contains an isomorphic copy of every sofic shift (where the isomorphism only changes the names of the letters in the alphabet).

Notation 4.3 For a subshift Y and $p \in \mathbb{N}$, let

$$\mathrm{Per}_p(Y) = \{y \in Y \colon \sigma^p y = y\}$$

denote the set of all periodic points in Y of (not necessarily minimal) period p. Let

$$\mathrm{Per}_{\leq p}(Y) := \bigcup_{q=1}^{p} \mathrm{Per}_q(Y)$$

denote the set of all periodic points in Y of minimal period at most p.

We check that if two sofic shifts have the same language up to some large scale, then they have the same periodic points of small period.

Lemma 4.4 *Fix $p \in \mathbb{N}$. Suppose that $Y_1, Y_2 \subseteq \mathcal{Y}_{a,f,R}$ for some a, f, $R \in \mathbb{N}$ and that there exists $N > p \cdot a^{\max\{f, 2R+1\}}$ such that $\mathcal{L}_N(Y_1) = \mathcal{L}_N(Y_2)$. Then $\mathrm{Per}_{\leq p}(Y_1) = \mathrm{Per}_{\leq p}(Y_2)$.*

Proof For $i = 1, 2$, choose shifts of finite type $X_i \in \mathcal{X}_{a,f}$ and factor maps $\varphi_i \colon X_i \to Y_i$. Let Φ_i be a range R block code that implements φ_i, meaning that

$$\varphi_i(x)_t = \Phi_i(x_{t-R}, \ldots, x_t, \ldots, x_{t+R})$$

for all $t \in \mathbb{Z}$. Let $\tilde{R} = \max\{\lceil (f-1)/2 \rceil, R\}$ and let $\tilde{\Phi}_i$ be a range $\tilde{R}$ block code that also implements φ_i (only making use of $(x_{t-R}, \ldots, x_{t+R})$ in the case that $(f-1)/2 > R$).

We prove $\mathrm{Per}_{\leq p}(Y_1) \subseteq \mathrm{Per}_{\leq p}(Y_2)$, the other case being analogous. Let $y \in \mathrm{Per}_{\leq p}(Y_1)$ and find $q \leq p$ such that $\sigma^q y = y$. Since $N > q \cdot a^{2\tilde{R}+1}$ and $\mathcal{L}_N(Y_2) = \mathcal{L}_N(Y_1)$, we can find $x \in X_2$ such that $\varphi_2(x)$ satisfies

$$y_t = \varphi_2(x)_t \quad \text{for all } 0 \leq t \leq q \cdot a^{2\tilde{R}+1}. \tag{2}$$

For each $0 \leq k \leq a^{2\tilde{R}+1}$, set

$$w_k := (x_{qk-\tilde{R}}, \ldots, x_{qk}, \ldots, x_{qk+\tilde{R}}).$$

It follows from Eq. (2) that $\tilde{\Phi}_2(w_k) = y_{qk}$ for all $0 \leq k \leq a^{2\tilde{R}+1}$. By definition of $\tilde{R}$, we have that $|w_k| \geq f$. Since $|\mathcal{L}_{2\tilde{R}+1}(X_2)| \leq a^{2\tilde{R}+1}$, there exist $0 \leq k_1 < k_2 \leq a^{2\tilde{R}+1}$ such that $w_{k_1} = w_{k_2}$. Since X_2 can be defined using a set of forbidden words that all have length at most f, there is a periodic point $z \in X_2$ of period $q(k_2 - k_1)$ such that

$$z_t = x_t \quad \text{for all } qk_1 - \tilde{R} \leq t \leq qk_2 - \tilde{R}.$$

Note that this automatically also holds for $qk_2 - \tilde{R} < t \leq qk_2 + \tilde{R}$ in z by periodicity, and in x since $w_{k_1} = w_{k_2}$. Therefore, it follows that

$$\varphi_2(z)_t = \varphi_2(x)_t \quad \text{for all } qk_1 \leq t \leq qk_2. \tag{3}$$

Thus $\varphi_2(z)$ is periodic, with (not necessarily minimal) period $q(k_2 - k_1)$, and because $\varphi_2(x)$ is a periodic word with period q, Eq. (3) implies that $\varphi_2(z)$ is also periodic with period q. It follows that $y = \varphi_2(z) \in \text{Per}_q(Y_2) \subseteq \text{Per}_{\leq p}(Y_2)$. Since $y \in \text{Per}_{\leq p}(Y_1)$ is arbitrary, it follows that $\text{Per}_{\leq p}(Y_1) \subseteq \text{Per}_{\leq p}(Y_2)$. $\qquad\square$

For a sofic shift Y, we are interested in constructing a measure supported on the periodic orbits of small period. Of course Y might not have any periodic points of very low period, and we use a lemma to provide an upper bound on the smallest period of any periodic point in Y.

Lemma 4.5 *Let $Y \in \mathcal{Y}_{a,f,R}$ for some $a, f, R \in \mathbb{N}$. There exists $p \leq a^f$ such that $\text{Per}_p(Y) \neq \emptyset$.*

Proof Find $X \in \mathcal{X}_{a,f}$ and a factor map $\varphi \colon X \to Y$. Since $|\mathcal{L}_f(X)| \leq a^f$, for any allowed word $x_1 \cdots x_{a^f + f} \in \mathcal{L}_{a^f + f}(X)$ there exist $0 \leq i < j \leq a^f$ such that

$$x_i \cdots x_{i+f-1} = x_j \cdots x_{j+f-1}.$$

Since X can be defined using minimal forbidden words that all have length at most f, there is a periodic point $y \in X$ of period $j - i$ such that $x_t = y_t$ for all $i \leq t \leq j + f$. It follows that $\varphi(y) \in Y$ is a periodic point of period dividing $j - i \leq a^f$. In particular, Y has a periodic point of period at most a^f. $\qquad\square$

Theorem 4.6 *Let X be a shift and $(X_n)_{n \in \mathbb{N}}$ be its SFT cover. If there exist infinitely many $n \in \mathbb{N}$ such that $X_n = X_{n+\tau(n)}$ for the function $\tau(n) = 2n + (1 + n^n) \cdot n^{4n+1}$, then any factor of X has a characteristic measure.*

We have not made any effort to optimize the growth rate $\tau(n)$, as this does not lead to stronger results in our setting. We note that this result gives a direct proof that any sofic shift admits a characteristic measure.

Proof Let Y be a factor of X and $\varphi \colon X \to Y$ be the associated a factor map. Let R_φ be a range for φ and let Φ be a range R_φ block code that implements it. The domain of Φ naturally extends to any shift Z satisfying $\mathcal{L}_{2R_\varphi+1}(Z) = \mathcal{L}_{2R_\varphi+1}(X)$. In particular, it extends to X_n for all $n \geq 2R_\varphi + 1$. Making a small abuse of notation, we use φ to denote the map determined by Φ on each such X_n. For $n \geq 2R_\varphi + 1$, define $Y_n := \varphi(X_n)$. It follows immediately from the definitions that $\mathcal{L}_k(Y_n) = \mathcal{L}_k(Y)$ for all $k \leq n - 2R_\varphi$.

Let $\beta \in \text{Aut}(Y)$ and let R_β be a symmetric range for β. Again, the domain of β naturally extends to any shift Z satisfying $\mathcal{L}_{2R_\beta+1}(Z) = \mathcal{L}_{2R_\beta+1}(Y)$. In particular, the domain extends to Y_n for all $n \geq 2R_\varphi + 2R_\beta + 1$ (recall that we assume that the range of β is symmetric). For such n, since $\beta(Y) = Y$ it follows that $\mathcal{L}_k(\beta(Y_n)) = \mathcal{L}_k(Y)$ for all $k \leq n - 2R_\beta$. Thus, the domain of the block code implementing β^{-1} extends naturally to $\beta(Y_n)$ for all $n \geq 4R_\beta + 1$, provided n is also at least $2R_\varphi + 2R_\beta + 1$ so that Y_n is defined. Composing the block codes implementing β and β^{-1}, the resulting block code of range $2R_\beta$ implements the identity map on any shift Z satisfying $\mathcal{L}_{4R_\beta+1}(Z) = \mathcal{L}_{4R_\beta+1}(Y)$. In particular, these block codes implement the identity map on Y_n for all $n \geq 2R_\beta + \max\{2R_\varphi, 2R_\beta\} + 1$. It follows that, for

such n, the map β is a topological conjugacy between Y_n and $\beta(Y_n)$. Moreover, for all such n,

$$\mathcal{L}_k(Y) = \mathcal{L}_k(Y_n) = \mathcal{L}_k(\beta(Y_n)) \quad \text{for all } k \leq n - 2R_\beta. \tag{4}$$

Finally we observe that Y_n is a factor of X_n implemented by some factor map of range R_φ and $\beta(Y_n)$ is a factor of X_n implemented by some factor map of range $R_\varphi + R_\beta$. Setting $a = \max\{|\mathcal{L}_1(X)|, |\mathcal{L}_1(Y)|\}$ we have, up to renaming the letters of the alphabets in X and Y, that both Y_n and $\beta(Y_n)$ are in $\mathcal{Y}_{a,n,R_\varphi+R_\beta}$.

Suppose $T, n \in \mathbb{N}$ are such that $X_n = X_{n+T}$ (and thus we also have that $Y_n = Y_{n+T}$). Then applying Eq. (4) to Y_{n+T}, it follows that

$$\mathcal{L}_k(Y) = \mathcal{L}_k(Y_n) = \mathcal{L}_k(\beta(Y_n)) \quad \text{for all } k \leq n + T - 2R_\beta \tag{5}$$

(note that we are using $Y_n = Y_{n+T}$ and $\beta(Y_n) = \beta(Y_{n+T})$). Applying Lemma 4.4, we have that $\text{Per}_{\leq p}(Y_n) = \text{Per}_{\leq p}(\beta(Y_n))$ for any p such that $n + T - 2R_\beta \geq p \cdot a^{\max\{n, 2R_\varphi+2R_\beta+1\}}$. In other words, this holds so long as

$$T \geq 2R_\beta + p \cdot a^{\max\{n, 2R_\varphi+2R_\beta+1\}} - n.$$

Moreover, by Lemma 4.5, we have that $\text{Per}_{\leq p}(Y_n) \neq \emptyset$ so long as $p \geq 1 + a^n$.

Summarizing, if $n, T \in \mathbb{N}$ are such that $X_n = X_{n+T}$ and

$$T \geq 2R_\beta + (1 + a^n) \cdot a^{\max\{n, 2R_\varphi+2R_\beta+1\}} - n$$

then there exists $p \leq 1 + a^n$ such that $\text{Per}_{\leq p}(Y_n) = \text{Per}_{\leq p}(\beta(Y_n))$ and these sets are nonempty. It follows that in this case we have

$$\beta\left(\text{Per}_{\leq(1+a^n)}(Y_n)\right) = \text{Per}_{\leq(1+a^n)}(\beta(Y_n)) \tag{6}$$

and so β preserves the measure

$$\nu_{a,n,R_\beta,R_\varphi} = \frac{1}{|\text{Per}_{\leq(1+a^n)}(Y_n)|} \sum_{z \in \text{Per}_{\leq(1+a^n)}(Y_n)} \delta_z.$$

In fact, this measure is preserved by any automorphism in $\text{Aut}(Y)$ of range at most R_β. But by the assumption on the SFT cover of X, there are infinitely many $n \in \mathbb{N}$ for which $Y_n = Y_{n+\tau(n)}$ and $\tau(n) \geq 2R_\beta + (1 + a^n) \cdot a^{\max\{n, 2R_\varphi+2R_\beta+1\}} - n$ for all but finitely many such n. Therefore, for all but finitely many such n, the measure $\nu_{n,n,n,n}$ is preserved by any automorphism of Y whose range is at most R_β. By taking n sufficiently large (depending on the radius of the factor map ϕ), the measure $\nu_{n,n,n,n}$ is preserved by ϕ. Since $\beta \in \text{Aut}(Y)$, and hence R_β is arbitrary, for any automorphism of Y the measure $\nu_{n,n,n,n}$ is preserved for all sufficiently large n. Any weak* limit of the sequence $\{\nu_{n,n,n,n}\}_{n=1}^\infty$ must be $\text{Aut}(Y)$-invariant and since $Y = \bigcap_{n=1}^\infty Y_n$, the limiting measure is supported on Y. Thus the factor Y has a characteristic measure. $\qquad\square$

It does not follow from the proof of Theorem 4.6 that the resulting characteristic measure is necessarily a measure of maximal entropy. Moreover, even though the set of measures of maximal entropy is compact, convex and invariant under all automorphisms, as the automorphism group of a shift need not be amenable, it is not clear *a priori* that there exists a measure of maximal entropy invariant under any automorphism. However, we show that with further assumptions on the growth of $\tau(n)$, we can guarantee this condition as well. To help us do this, we need to better understand the measures of maximal entropy on a (not necessarily transitive) sofic shift Y and this is carried out in the next section.

4.2 Obtaining a characteristic measure that is of maximal entropy

A shift (X, σ) is *forward transitive* if for some $x \in X$, the forward orbit $\{\sigma^n x : n \in \mathbb{N}\}$ is dense in X. The next lemma states that the support of a measure of maximal entropy of a sofic shift lies on a forward transitive sofic shift, including a quantitative version that holds for non-irreducible sofic shift. See [16, Chap. 6] for standard results on irreducible subshifts of finite type.

Lemma 4.7 *Let $Y \in \mathcal{Y}_{a,f,R}$ for some $a, f, R \in \mathbb{N}$, $f > 2R + 1$. There exists a forward transitive sofic shift $Z \subseteq Y$ such that $Z \in \mathcal{Y}_{a,f,R}$ and $h_{\text{top}}(Z) = h_{\text{top}}(Y)$. Moreover, if $X \in \mathcal{X}_{a,f}$ and $\varphi \colon X \to Y$ is a range R block code, then there exists $X' \subseteq X$ such that $X' \in \mathcal{X}_{a,f}$, X' is forward transitive and $Z = \varphi(X')$. In addition, if μ_Y is any ergodic measure of maximal entropy on Y, then Z can be chosen such that Z contains the support of μ_Y.*

Proof Let $X \in \mathcal{X}_{a,f}$ and let $\varphi \colon X \to Y$ be a factor map that can be implemented by a range R block code (note that by Lemma 4.2, we can assume that X and Y are defined over the same alphabet). Since Y is a subshift, it has an ergodic measure of maximal entropy μ_Y. By the pointwise ergodic theorem, for μ_Y-almost every $y \in Y$ we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi(\sigma^k y) = \int_Y \psi \, d\mu_Y$$

for all $\psi \in L^1(\mu_Y)$. Fix one such $y \in Y$ and observe that μ_Y is supported on the ω-limit set $\omega(y) \subseteq Y$. In particular, $h_{\text{top}}(\omega(y)) \geq h_{\mu_Y}(\sigma) = h_{\text{top}}(Y)$ and thus $h_{\text{top}}(\omega(y)) = h_{\text{top}}(Y)$.

Choose $x \in X$ such that $y = \varphi(x)$. Define

$$\mathcal{W}_x := \{w \in \mathcal{L}_f(X) \colon \sigma^k(x) \in [w] \text{ for infinitely many } k \in \mathbb{N}\}$$

and $\mathcal{F}_x := \{0, 1, \dots, a-1\}^f \setminus \mathcal{W}_x$. Let X' be the shift of finite type defined by forbidding the words in $\mathcal{F}_x$. So we have that $X' \in \mathcal{X}_{a,f}$ and $X' \subset X$. Since they are only finitely many words of length f, ultimately, the sequence x is a concatenation of words in $\mathcal{W}_x$. More precisely, there is some $k_0 \in \mathbb{N}$ such that $x_k x_{k+1} \cdots x_{k+f-1} \in \mathcal{L}_f(X')$ for all $k \geq k_0$. Hence any word $w \in \mathcal{W}_x$ occurs in an infinite sequence in X' and a word $w \in \mathcal{L}(X)$ with $|w| \geq f$ is in the language $\mathcal{L}(X')$ if and only if all of its subwords of length f are in the language $\mathcal{L}_f(X')$.

Since X' is defined by forbidden words of length f and for any $u, v \in \mathcal{L}_f(X')$ there exists $w \in \mathcal{L}(X')$ such that $\sigma^k(x) \in [uwv]$ for some $k \geq k_0$, it follows that X' is forward transitive. Setting $Z := \varphi(X') \in \mathcal{Y}_{a,f,R}$, we have that Z is a forward transitive sofic shift in Y. Furthermore, since X' contains the ω-limit set of x, it follows that $\omega(y) \subseteq Z$. Thus we have that $h_{\text{top}}(Y) = h_{\text{top}}(\omega(y)) \leq h_{\text{top}}(Z)$, which implies that $h_{\text{top}}(Z) = h_{\text{top}}(Y)$. □

Recall that any forward transitive sofic shift has a unique measure of maximal entropy [23] and is *entropy minimal* [11], meaning that all proper subshifts have strictly lower entropy.

Lemma 4.8 *Let $Y \in \mathcal{Y}_{a,f,R}$ for some $a, f, R \in \mathbb{N}$, and $f > 2R + 1$. There exist $k \in \mathbb{N}$ and sofic subshifts $Y_1, \dots, Y_k \subseteq Y$ such that*

(1) *For $i = 1, \dots, k$, each shift $Y_i \in \mathcal{Y}_{a,f,R}$;*
(2) *For $i = 1, \dots, k$, each shift Y_i is forward transitive;*
(3) *For $i = 1, \dots, k$, we have $h_{\text{top}}(Y_i) = h_{\text{top}}(Y)$;*
(4) *Every ergodic measure of maximal entropy supported on Y is supported on $\bigcup_{i=1}^{k} Y_i$;*

(5) *For any $\alpha \in \mathrm{Aut}(Y)$, we have $\alpha(\bigcup_{i=1}^k Y_i) = \bigcup_{i=1}^k Y_i$.*

Proof Let $X \in \mathcal{X}_{a,f}$ and $\varphi \colon X \to Y$ a factor map that can be implemented by a range R block code. By Lemma 4.7, for any ergodic measure of maximal entropy μ on Y, there exists a forward transitive shift $X_\mu \subseteq X$ such that $X_\mu \in \mathcal{X}_{a,f}$ and μ is supported on the forward transitive sofic shift $Y_\mu := \varphi(X_\mu) \in \mathcal{Y}_{a,f,R}$ which satisfies $h_{\mathrm{top}}(Y_\mu) = h_{\mathrm{top}}(Y)$. Since a forward transitive sofic shift is entropy minimal, it follows that Y_μ is actually equal to the support of μ. Since $\mathcal{X}_{a,f}$ is finite, there are only finitely many distinct shifts X_μ and only finitely many distinct shifts Y_μ. Enumerate the collection of all shifts that arise as Y_μ, for some ergodic measure of maximal entropy μ, as $Y_1, \ldots, Y_k$. If $\alpha \in \mathrm{Aut}(X)$, then α permutes the ergodic measures of maximal entropy supported on Y, and therefore permutes the associated supports. Therefore α permutes $Y_1, \ldots, Y_k$ and so preserves $\bigcup_{i=1}^k Y_i$. □

Theorem 4.9 *For any $a, f, R \in \mathbb{N}$ and $Y \in \mathcal{Y}_{a,f,R}$, the subshift Y has a characteristic measure of maximal entropy that is a weak* limit of the sequence of measures $\{\nu_n\}_{n=1}^\infty$ where*

$$\nu_n := \frac{1}{|\mathrm{Per}_{\leq n}(Y)|} \sum_{z \in \mathrm{Per}_{\leq n}(Y)} \delta_z.$$

Moreover, if $\mathcal{S} \subseteq \mathbb{N}$ is infinite, there is a characteristic measure of maximal entropy that is obtained as the weak* limit of ν_n along a subsequence of elements of $\mathcal{S}$ that is uniform for any $Y \in \mathcal{Y}_{a,f,R}$.

It is easy to check that a sofic shift always has a characteristic measure of maximal entropy, noting for example that every forward transitive component is intrinsically ergodic. Theorem 4.9 shows the finer result that every subsequence of $\{\nu_n\}$ has a weak* limit point that is a characteristic measure. In particular, some accumulation point of the set of periodic measures $\{\nu_n\}_{n=1}^\infty$ is characteristic and is the measure of maximal entropy. We do not know if this method gives rise to all characteristic measures of maximal entropy.

Proof Let $Y \in \mathcal{Y}_{a,f,R}$ and let $Y_1, \ldots, Y_k \subseteq Y$ be as in Lemma 4.8. Suppose $\mathcal{S} \subseteq \mathbb{N}$ is given. Recall that any forward transitive sofic shift admits a unique measure of maximal entropy (see [8, Theorem 3.2]). For each $i = 1, \ldots, k$ let μ_i be the unique measure of maximal entropy supported on Y_i. We first show any measure of the form

$$\mu := \sum_{i=1}^k \tilde{c}_i \mu_i$$

is a characteristic measure on Y, provided $\tilde{c}_i \in [0, 1]$ for all $i = 1, \ldots, k$, $\sum c_i = 1$, and whenever Y_i and Y_j are topologically conjugate we have $\tilde{c}_i = \tilde{c}_j$.

Any $\alpha \in \mathrm{Aut}(Y)$ permutes $Y_1, \ldots, Y_k$ and so permutes the measures $\mu_1, \ldots, \mu_k$, as each μ_i is the unique measure of maximal entropy on Y_i for all $i = 1, \ldots, k$. Note that if this permutation sends μ_i to μ_j then Y_i is topologically conjugate to Y_j and so $\tilde{c}_i = \tilde{c}_j$. Therefore $\alpha_* \mu = \mu$. Since $\alpha \in \mathrm{Aut}(Y)$ is arbitrary, μ is an $\mathrm{Aut}(Y)$-characteristic measure.

Next, find a sequence $\{n_t\}_{t=1}^\infty$ of elements of $\mathcal{S}$ along which

$$c_i := \lim_{t \to \infty} \frac{|\mathrm{Per}_{\leq n_t}(Y_i)|}{\left|\bigcup_{j=1}^k \mathrm{Per}_{\leq n_t}(Y_j)\right|}$$

exists for all $i = 1, \ldots, k$. Note that $c_i \in [0, 1]$ for all $i = 1, \ldots, k$. Entropy minimality of Y_i implies that for any $i \neq j$, we have that $h_{\mathrm{top}}(Y_i \cap Y_j) < h_{\mathrm{top}}(Y_i)$.

Recall that for any subshift W, the entropy is an upper bound for the exponential growth rate of $|\text{Per}_{\leq n}(W)|$, and the two quantities are equal whenever W is a transitive sofic shift. It follows that for any sofic shift Y_i

$$\lim_{n \to \infty} \frac{|\text{Per}_{\leq n}(Y_i \cap Y_j)|}{|\text{Per}_{\leq n}(Y_i)|} = 0.$$

Defining $P_n(Y_i) := \text{Per}_{\leq n}(Y_i) \setminus \left(\bigcup_{j \neq i} \text{Per}_{\leq n}(Y_j) \right)$, we have that

$$\lim_{n \to \infty} \frac{|P_n(Y_i)|}{|\text{Per}_{\leq n}(Y_i)|} = 1 \tag{7}$$

and so

$$c_i = \lim_{t \to \infty} \frac{|P_{n_t}(Y_i)|}{\left| \bigcup_{j=1}^k P_{n_t}(Y_j) \right|}. \tag{8}$$

Therefore the coefficients c_i satisfy $\sum_i c_i = 1$. Finally note that if Y_i is topologically conjugate to Y_j then $|\text{Per}_{\leq n_t}(Y_i)| = |\text{Per}_{\leq n_t}(Y_j)|$ for all t, so $c_i = c_j$. It follows that

$$\mu_Y := \sum_{i=1}^k c_i \mu_i$$

is an $\text{Aut}(Y)$-characteristic measure.

We next show that μ_Y is the weak* limit of the sequence $\{\nu_{n_t}\}_{t=1}^\infty$. Indeed, for each index i, the set of points of period at most n_t forms an (n_t, ε_0)-separated set in the subshift Y_i for some constant $\varepsilon_0 > 0$ determined by the expansivity constant of the shift. Furthermore, we have that $\limsup_t 1/n_t \log(|\text{Per}_{\leq n_t}(Y_i)|)$ is the topological entropy $h(Y_i)$ of the transitive sofic shift Y_i (see for instance [18, Theorem 4.3.6]). Each of the periodic measures $1/|\text{Per}_{\leq n_t}(Y_i)| \sum_{z \in \text{Per}_{\leq n_t}(Y_i)} \delta_z$ is shift invariant, and any weak* limit of these measures taken along a subsequence is a measure of maximal entropy (this follows, for example, from the proof of the variational principle in [22, theorem 8.6] and is given more explicitly in [8, Proposition 5.4]). Since for a forward transitive sofic shift, the measure of maximal entropy is unique [8, Theorem 3.2], for each $i = 1, \dots, k$ we have the convergence

$$\frac{1}{|\text{Per}_{\leq n_t}(Y_i)|} \sum_{z \in \text{Per}_{\leq n_t}(Y_i)} \delta_z \xrightarrow[t \to \infty]{} \mu_i.$$

Combining this with Eq. (7) we have that

$$\frac{1}{|P_{n_t}(Y_i)|} \sum_{z \in P_{n_t}(Y_i)} \delta_z \xrightarrow[t \to \infty]{} \mu_i.$$

Therefore, μ_Y is the weak* limit

$$\lim_{t \to \infty} \sum_{i=1}^k \frac{|P_{n_t}(Y_i)|}{\left| \bigcup_{j=1}^k P_{n_t}(Y_j) \right|} \cdot \frac{1}{|P_{n_t}(Y_i)|} \sum_{z \in P_{n_t}(Y_i)} \delta_z = \lim_{t \to \infty} \frac{1}{\left| \bigcup_{j=1}^k P_{n_t}(Y_j) \right|} \sum_{z \in \bigcup_{j=1}^k P_{n_t}(Y_j)} \delta_z.$$

Therefore, again using from (7), it follows that μ_Y is the weak-* limit of the sequence

$$\frac{1}{\left| \text{Per}_{\leq n_t} \left(\bigcup_{i=1}^k Y_i \right) \right|} \sum_{z \in \text{Per}_{\leq n_t} \left(\bigcup_{i=1}^k Y_i \right)} \delta_z.$$

Finally, let $X \in \mathcal{X}_{a,f}$ and let $\varphi \colon X \to Y$ be a factor map that can be implemented by a range R block code. For each periodic point $y \in Y$, there is a periodic point $x \in X$ such that $y = \varphi(x)$. Let $X' \subseteq X$ be the shift of finite type obtained by forbidding all words of length f that do not appear in the point x. Note that $X' \in \mathcal{X}_{a,f}$, X' is forward transitive, and $x \in X'$. Therefore $y \in \varphi(X')$. It follows that every periodic point in Y lies in $\varphi(X')$ for some $X' \subseteq X$ with $X' \in \mathcal{X}_{a,f}$. Since there are only finitely many such subshifts and all subshifts with entropy $h_{\text{top}}(Y)$ already appear in the enumeration $Y_1, \ldots, Y_k$ (see Lemma 4.8, Part 3), it follows that all periodic points in $Y \setminus \left(\bigcup_{i=1}^{k} Y_i \right)$ are in the union of a finite number of shifts of the form $\varphi(Z)$, where $Z \in \mathcal{X}_{a,f}$ and $h_{\text{top}}(\varphi(Z)) < h_{\text{top}}(Y)$. Using that for any shift W, its topological entropy $h_{\text{top}}(W)$ is greater than or equal to $\limsup_n \log(|\text{Per}_{\leq n}(W)|)/n$, with equality when W is a transitive sofic shift, we have that

$$\lim_{n \to \infty} \frac{\left| \text{Per}_{\leq n} \left(\bigcup_{i=1}^{k} Y_i \right) \right|}{|\text{Per}_{\leq n}(Y)|} = 1$$

and so μ_Y is the weak* limit of the sequence

$$\frac{1}{|\text{Per}_{\leq n_t}(Y)|} \sum_{z \in \text{Per}_{\leq n_t}(Y)} \delta_z.$$

$\square$

The next lemma strengthens Theorem 4.9 by providing for any sofic shift, a uniform sequence of periods that equidistributes to a measure of maximal entropy, and this is used in the proof of Theorem 1.2.

Lemma 4.10 *There exists an infinite subset $\mathcal{S} \subseteq \mathbb{N}$ such that for any sofic shift Y, the weak* limit*

$$\mu_Y := \lim_{s \in \mathcal{S}, s \to \infty} \frac{1}{|\text{Per}_{\leq s}(Y)|} \sum_{z \in \text{Per}_{\leq s}(Y)} \delta_z \tag{9}$$

exists and is a characteristic measure of maximal entropy on Y.

Proof Note that

$$\bigcup_{a=1}^{\infty} \bigcup_{f=1}^{\infty} \bigcup_{R=1}^{\infty} \mathcal{Y}_{a,f,R} = \bigcup_{n=1}^{\infty} \mathcal{Y}_{n,n,n}$$

and so it suffices to ensure sure that the limit in (9) exists for all $Y \in \bigcup_{n=1}^{\infty} \mathcal{Y}_{n,n,n}$. We construct $\mathcal{S}$ iteratively using a diagonalization argument. Enumerate the elements (of the finite set) $\mathcal{Y}_{1,1,1}$ as $Y_1, \ldots, Y_r$. Begin by setting $\mathcal{S}_0 := \mathbb{N}$. By Theorem 4.9, there exists a sequence $\{n_t\}_{t=1}^{\infty}$ along which

$$\lim_{t \to \infty} \frac{1}{|\text{Per}_{\leq n_t}(Y_1)|} \sum_{z \in \text{Per}_{\leq n_t}(Y_1)} \delta_z$$

exists and is a characteristic measure of maximal entropy on Y_1. Set $\mathcal{S}_1 := \{1\} \cup \{n_t : t \in \mathbb{N}\}$. Suppose we have constructed nested infinite sets

$$\mathcal{S}_0 \supseteq \mathcal{S}_1 \supseteq \cdots \supseteq \mathcal{S}_u$$

such that $\mathcal{S}_u$ contains the $v + 1$ smallest elements of $\mathcal{S}_v$ for all $v < u$ and such that

$$\lim_{s \in \mathcal{S}_v, s \to \infty} \frac{1}{|\mathrm{Per}_{\leq s}(Y_v)|} \sum_{z \in \mathrm{Per}_{\leq s}(Y_v)} \delta_z$$

exists and is a characteristic measure of maximal entropy on Y_v for all $0 < v \leq u$. If $u = r$ this part of the construction ends and we move onto the next step. Otherwise, use Theorem 4.9 with the infinite set $\mathcal{S}_u$ to find a subsequence $\{n_t\}_{t=1}^{\infty}$ of elements of $\mathcal{S}_u$ for which

$$\lim_{t \to \infty} \frac{1}{|\mathrm{Per}_{\leq n_t}(Y_{u+1})|} \sum_{z \in \mathrm{Per}_{\leq n_t}(Y_{u+1})} \delta_z$$

exists and is a characteristic measure on Y_{u+1}. Define the set $\mathcal{S}_{u+1}$ to be the union of $\{n_t : t \in \mathbb{N}\}$ with the u smallest elements of $\mathcal{S}_u$. This part of the construction terminates when we construct $\mathcal{S}_r$. For convenience in the next step, we define $\mathcal{S}^{(1)} := \mathcal{S}_r$.

Suppose we have defined a nested sequence of infinite sets

$$\mathcal{S}^{(1)} \supseteq \mathcal{S}^{(2)} \supseteq \cdots \supseteq \mathcal{S}^{(j)}$$

such that for all $Y \in \mathcal{Y}_{j,j,j}$ the weak* limit

$$\lim_{s \in \mathcal{S}^{(j)}, s \to \infty} \frac{1}{|\mathrm{Per}_{\leq s}(Y)|} \sum_{z \in \mathrm{Per}_{\leq s}(Y)} \delta_z$$

exists and is a characteristic measure of maximal entropy on Y. Suppose further that $\mathcal{S}^{(j)}$ contains the i smallest elements of $\mathcal{S}^{(i)}$ for all $i < j$. Proceed as in the construction of $\mathcal{S}^{(1)}$, but using $\mathcal{S}^{(j)}$ in place of $\mathcal{S}_0$ at the start, to build an infinite set $\mathcal{S}^{(j+1)} \subseteq \mathcal{S}^{(j)}$ that contains the j smallest elements of $\mathcal{S}^{(j)}$ and for which

$$\lim_{s \in \mathcal{S}^{(j+1)}, s \to \infty} \frac{1}{|\mathrm{Per}_{\leq s}(Y)|} \sum_{z \in \mathrm{Per}_{\leq s}(Y)} \delta_z$$

exists and is a characteristic measure of maximal entropy on Y for all $Y \in \mathcal{Y}_{j+1,j+1,j+1}$.

Continuing inductively, we construct an infinite sequence of nested, infinite sets

$$\mathcal{S}^{(1)} \supseteq \mathcal{S}^{(2)} \supseteq \cdots \supseteq \mathcal{S}^{(j)} \supseteq \cdots$$

such that for any $i < j$ the set $\mathcal{S}^{(j)}$ contains the i smallest elements of $\mathcal{S}^{(i)}$ and such that for any $Y \in \mathcal{Y}_{j,j,j}$ the weak* limit

$$\lim_{s \in \mathcal{S}^{(j)}, s \to \infty} \frac{1}{|\mathrm{Per}_{\leq s}(Y)|} \sum_{z \in \mathrm{Per}_{\leq s}(Y)} \delta_z$$

exists and is a characteristic measure of maximal entropy on Y. We define the (nonempty by construction) set $\mathcal{S} := \bigcap_{j=1}^{\infty} \mathcal{S}^{(j)}$. □

We use this to motivate the relevant class of shifts and approximations.

Definition 4.11 Let $\mathcal{S} \subseteq \mathbb{N}$ be the set constructed in Lemma 4.10. For $a, f, R \in \mathbb{N}$ and $Y \in \mathcal{Y}_{a,f,R}$, let

$$\mu_Y := \lim_{s \in \mathcal{S}, s \to \infty} \frac{1}{|\mathrm{Per}_{\leq s}(Y)|} \sum_{z \in \mathrm{Per}_{\leq s}(Y)} \delta_z.$$

For $k, m \in \mathbb{N}$, define $A(a, f, R, k, m) \in \mathbb{N}$ to be the least element of $\mathcal{S}$ such that for all $Y \in \mathcal{Y}_{a,f,R}$ and all $w \in \bigcup_{i=1}^{m} \mathcal{L}_i(Y)$ we have

$$\left| \mu_Y([w]) - \frac{1}{|\mathrm{Per}_{\leq s}(Y)|} \sum_{z \in \mathrm{Per}_{\leq s}(Y)} \delta_z([w]) \right| < \frac{1}{k}$$

for all $s \in \mathcal{S}$ with $s \geq A(a, f, R, k, m)$.

For $n \in \mathbb{N}$, let $m = m(n)$ be

$$m := \inf\{\mathcal{S} \setminus \{1, 2, \ldots, n\}\}.$$

Define $\omega \colon \mathbb{N} \to \mathbb{N}$ by

$$\omega(n) := -n + 2m + A(m, m, m, m, m) \cdot m^{2m+1} \tag{10}$$

We are now ready to formulate a version of Theorem 4.6 that relies on a stronger assumption on how well-approximable the shift is to obtain the stronger conclusion that the resulting characteristic measure is a measure of maximal entropy, giving us the statement of Theorem 1.2.

Proposition 4.12 *Let X be a shift and $(X_n)_{n \in \mathbb{N}}$ its SFT cover. If there are infinitely many $n \in \mathbb{N}$ such that $X_n = X_{n+\omega(n)}$ for the function $\omega(n)$ defined in (10), then every subshift factor of X has a characteristic measure which is a measure of maximal entropy.*

We note that Example 2.2 shows that the class of shifts verifying this condition is larger than that of shifts of finite type.

Proof Let

$$\tau(n) := A(n, n, n, n, n) \cdot n^{2n+1} + n$$

and note that, by definition of ω, for any $n \in \mathbb{N}$ such that $X_n = X_{n+\omega(n)}$, there exists $m \geq n$ with $m \in \mathcal{S}$ and such that $X_m = X_{m+\tau(m)}$. Therefore our assumption that there are infinitely many n such that $X_n = X_{n+\omega(n)}$ implies that there exist infinitely many $m \in \mathcal{S}$ such that $X_m = X_{m+\tau(m)}$. To avoid unnecessary confusion, we make a small abuse of notation and simply state that

$$\text{there are infinitely many } n \in \mathcal{S} \text{ such that } X_n = X_{n+\tau(n)}. \tag{11}$$

Let Y be a subshift factor of X and let $\varphi \colon X \to Y$ be a factor map with range R_φ. Setting $a := \max\{|\mathcal{L}_1(X)|, |\mathcal{L}_1(Y)|\}$, without loss of generality (renaming the letters of the alphabets if necessary) we can assume that $X, Y \subseteq \{0, 1, \ldots, a-1\}^{\mathbb{Z}}$. Since X_n is defined by forbidding only words of length n, we have that $X_n \in \mathcal{X}_{a,n}$ for all $n \in \mathbb{N}$.

We begin as in the proof of Theorem 4.6. For each $n \geq 2R_\varphi + 1$, define $Y_n := \varphi(X_n)$ and observe that $\mathcal{L}_k(Y_n) = \mathcal{L}_k(Y)$ for all $k \leq n - 2R_\varphi$. For all such n, we have that $Y_n \in \mathcal{Y}_{a,n,R_\varphi}$. Let $\beta \in \mathrm{Aut}(Y)$ and let R_β be a symmetric range for β. For any $n \geq 2R_\varphi + 2R_\beta + 1$ and any $k \leq n - 2R_\beta$, we have that $\mathcal{L}_k(\beta(Y_n)) = \mathcal{L}_k(Y)$. Provided that $n \geq 2R_\beta + \max\{2R_\varphi, 2R_\beta\} + 1$, we have that β is a topological conjugacy between Y_n and $\beta(Y_n)$ and that

$$\mathcal{L}_k(Y) = \mathcal{L}_k(Y_n) = \mathcal{L}_k(\beta(Y_n)) \quad \text{for all } k \leq n - 2R_\beta.$$

It follows immediately that $\beta(Y_n) \in \mathcal{Y}_{a,n,R_\varphi+R_\beta}$ for all such n and that $Y_n \in \mathcal{Y}_{a,n,R_\varphi} \subseteq \mathcal{Y}_{a,n,R_\varphi+R_\beta}$. For any n such that $X_n = X_{n+T}$ (meaning we also have that $Y_n = Y_{n+T}$), we have that

$$\mathcal{L}_k(Y) = \mathcal{L}_k(Y_n) = \mathcal{L}_k(\beta(Y_n)) \quad \text{for all } k \leq n + T - 2R_\beta.$$

For such n, using the parameters a, $f := n$, $R := R_\varphi + R_\beta$, and $N := n + T - 2R_\beta$ in Lemma 4.4, it follows that $\text{Per}_{\leq p}(Y_n) = \text{Per}_{\leq p}(\beta(Y_n))$ so long as $T > p \cdot a^{\max\{n, 2R_\varphi + 2R_\beta + 1\}} + 2R_\beta - n$. In particular, when

$$T > A(n, n, n, n, n) \cdot a^{\max\{n, 2R_\varphi + 2R_\beta + 1\}} + 2R_\beta - n,$$

it follows that

$$\text{Per}_{\leq A(n,n,n,n,n)}(Y_n) = \text{Per}_{\leq A(n,n,n,n,n)}(\beta(Y_n)). \tag{12}$$

For fixed φ and β, by Eq. (11), there are infinitely many $n \in \mathcal{S}$ for which $X_n = X_{n+\tau(n)}$ where

$$\tau(n) = A(n, n, n, n, n) \cdot n^{2n+1} + n > A(n, n, n, n, n) \cdot a^{\max\{n, 2R_\varphi + 2R_\beta + 1\}} + 2R_\beta - n$$

(since $n > \max\{R_\varphi + R_\beta, a\}$ for all but finitely many n). For any such n, recall that we have $Y_n, \beta(Y_n) \in \mathcal{Y}_{a,n,R_\varphi + R_\beta} \subseteq \mathcal{Y}_{n,n,n}$. Therefore, by definition of $A(n, n, n, n, n)$ and Theorem 4.9, for any $w \in \mathcal{L}_n(Y_n)$ we have (for convenience, we write $A(n) = A(n, n, n, n, n)$ in this calculation)

$$|\mu_{Y_n}([w]) - \mu_{\beta(Y_n)}([w])| \leq \left|\mu_{Y_n}([w]) - \frac{1}{|\text{Per}_{\leq A(n)}(Y_n)|} \sum_{z \in \text{Per}_{\leq A(n)}(Y_n)} \delta_z([w])\right|$$

$$+ \left|\mu_{\beta(Y_n)}([w]) - \frac{1}{|\text{Per}_{\leq A(n)}(Y_n)|} \sum_{z \in \text{Per}_{\leq A(n)}(Y_n)} \delta_z([w])\right|$$

$$= \left|\mu_{Y_n}([w]) - \frac{1}{|\text{Per}_{\leq A(n)}(Y_n)|} \sum_{z \in \text{Per}_{\leq A(n)}(Y_n)} \delta_z([w])\right|$$

$$+ \left|\mu_{\beta(Y_n)}([w]) - \frac{1}{|\text{Per}_{\leq A(n)}(\beta(Y_n))|} \sum_{z \in \text{Per}_{\leq A(n)}(\beta(Y_n))} \delta_z([w])\right|,$$

where we use Eq. (12) to make the change from Y_n to $\beta(Y_n)$. By definition of $A(n, n, n, n, n)$ and the fact that $Y_n, \beta(Y_n) \in \mathcal{Y}_{n,n,n}$, it follows that for any $w \in \mathcal{L}_n(Y_n) = \mathcal{L}_n(\beta(Y_n))$ we have that

$$|\mu_{Y_n}([w]) - \mu_{\beta(Y_n)}([w])| \leq \frac{2}{n}. \tag{13}$$

By Theorem 4.9, the measure μ_{Y_n} is the average of all ergodic measures of maximal entropy on Y_n and $\mu_{\beta(Y_n)}$ is the average of all ergodic measures of maximal entropy on $\beta(Y_n)$. Since β is a topological conjugacy between Y_n and $\beta(Y_n)$, it follows that $\beta_* \mu_{Y_n} = \mu_{\beta(Y_n)}$. Combining this with Eq. (13), if μ is any weak* limit of the sequence $\{\mu_{Y_n}\}_{n=1}^{\infty}$, then $\beta_* \mu = \mu$. Since $\beta \in \text{Aut}(Y)$ is arbitrary, any such weak* limit is an $\text{Aut}(Y)$-characteristic measure supported on $Y = \bigcap_{n=1}^{\infty} Y_n$. Finally, note that $h_{\mu_{Y_n}}(\sigma) = h_{\text{top}}(Y_n) \geq h_{\text{top}}(Y)$ for all $n \in \mathbb{N}$ and so any weak* limit point is also a measure of maximal entropy on Y. $\square$

5 Further properties of well-approximable language stable shifts

5.1 Not every shift is a factor of a well-approximable language stable shift

Suppose $(a_n)_{n \in \mathbb{N}}$ is a non-decreasing sequence of positive integers such that $\lim_n a_n = \infty$. Define $\mathcal{Z}(a_n)$ to be the set of all shifts (X, σ) such that X has no minimal forbidden words of

length between n and $n + a_n$ for infinitely many $n \in \mathbb{N}$. Taking all choices of such sequences $(a_n)_{n\in\mathbb{N}}$, we obtain all of the language stable shifts, and we show that there is some shift that is not a factor of $\mathcal{Z}(a_n)$ for some particular choice of sequence.

Proposition 5.1 *There exists a non-decreasing sequence of positive integers $(a_n)_{n\in\mathbb{N}}$ and a subshift $B \subseteq \{0, 1\}^{\mathbb{Z}}$ such that B is not a factor of any element of $\mathcal{Z}(a_n)$.*

Proof We construct the shift B recursively, constructing a nested sequence of subshifts and then defining B to be their intersection.

Let $\mathcal{A}_0 := \{0, 1\}$. For each $t \in \mathbb{N}$, let $B_{0,t}$ denote the subshift of all possible bi-infinite concatenations of the words $0(1^{2t})$ and $0(1^{2t+1})$. Note that

$$\mathcal{L}_{4t+4}(B_{0,t}) \cap \mathcal{L}_{4t+4}(B_{0,t'}) = \emptyset \quad \text{for all } t < t', \tag{14}$$

since every element of $\mathcal{L}_{4t+4}$ contains either $01^{2t}0$ or $01^{2t+1}0$ as a subword whereas no element of $\mathcal{L}_{4t+4}(B_{0,t'})$ does. Observe that $\mathcal{L}_1(B_{0,t}) = \{0, 1\}$ for all $t \in \mathbb{N}$. Find some $t_0 > |\mathcal{Y}_{1,1,1}|$ such that

$$\mathcal{L}_{4|\mathcal{Y}_{1,1,1}|+4}(B_{0,t_0}) \neq \mathcal{L}_{4|\mathcal{Y}_{1,1,1}|+4}(Y) \quad \text{for all } Y \in \mathcal{Y}_{1,1,1}$$

which is possible by (14). Let $u_0 \in \mathcal{L}(B_{0,t_0})$ be such that u_0 is a concatenation of the words 01^{2t_0} and 01^{2t_0+1} and every element of $\mathcal{L}_{4|\mathcal{Y}_{1,1,1}|+4}(B_{0,t_0})$ occurs as a subword of u_0. We define two words $w_{0,0}$ and $w_{1,0}$ which play the role of the alphabet in the next stage of our construction, meaning that every element of the shift we construct at this stage is a bi-infinite concatenation of these words. Set $w_{0,0} := u_0(01^{2t_0})^{(2t_0+2)|u_0|(4|\mathcal{Y}_{1,1,1}|+4)}$ and $w_{1,0} = u_0(01^{2t_0+1})^{(2t_0+1)|u_0|(4|\mathcal{Y}_{1,1,1}|+4)}$. The words $w_{0,0}$ and $w_{1,0}$ have the following properties:

(1) $|w_{0,0}| = |w_{1,0}|$;
(2) every element of $\mathcal{L}_{4|\mathcal{Y}_{1,1,1}|+4}(B_{0,t_0})$ occurs as a subword of $w_{0,0}$ and $w_{1,0}$;
(3) any $\mathbb{Z}$-coloring that can be written as a bi-infinite concatenation of $w_{0,0}$ and $w_{1,0}$ can be written uniquely in this way (this is done by identifying where the word u_0 occurs);
(4) for any $a, b \in \{0, 1\}$, we have $w_{a,0}w_{b,0} \in \mathcal{L}(B_{0,t_0})$.

Inductively, suppose we have constructed words $w_{0,r}$ and $w_{1,r}$ which have the same lengths and the property that any $\mathbb{Z}$-coloring that can be written as a bi-infinite concatenation of $w_{0,r}$ and $w_{1,r}$ can be written uniquely in this way. Further suppose we have constructed an integer t_r and a subshift B_{r,t_r} such that

$$\mathcal{L}_{4|\mathcal{Y}_{r,r,r}|+4}(B_{r,t_r}) \neq \mathcal{L}_{4|\mathcal{Y}_{1,1,1}|+4}(Y) \quad \text{for all } Y \in \mathcal{Y}_{r,r,r}$$

and that every element of $\mathcal{L}_{4|\mathcal{Y}_{r,r,r}|+4}(B_{r,t_r})$ occurs as a subword of both $w_{0,r}$ and $w_{1,r}$. Also suppose $|w_{0,r}|, |w_{1,r}| > 4|\mathcal{Y}_{r,r,r}| + 4$. Finally suppose that for any $a, b \in \{0, 1\}$ we have $w_{a,r}w_{b,r} \in \mathcal{L}(B_{r,t_r})$. We now mimic the construction of B_{0,t_0} with $w_{0,r}$ and $w_{1,r}$ playing the role of an alphabet. For each $t_r \in \mathbb{N}$, let B_{r+1,t_r} be the subshift of all possible bi-infinite concatenations of the words $w_{0,r}(w_{1,r}^{2t_r})$ and $w_{0,r}(w_{1,r}^{2t_r+1})$. Note that $\mathcal{L}_{4|\mathcal{Y}_{r,r,r}|+4}(B_{r+1,t_r}) = \mathcal{L}_{4|\mathcal{Y}_{r,r,r}|+4}(B_{r,t_r})$ because every element of $\mathcal{L}_{4|\mathcal{Y}_{r,r,r}|+4}(B_{r,t_r})$ occurs as a subword of both $w_{0,r}$ and $w_{1,r}$. Moreover, we have that

$$\mathcal{L}_{(4t_r+4)|w_{0,r}|}(B_{r+1,t_r}) \cap \mathcal{L}_{(4t_r+4)|w_{0,r}|}(B_{r+1,t'_r}) = \emptyset \quad \text{for all } t_r < t'_r \tag{15}$$

because every element of $\mathcal{L}_{(4t_r+4)|w_{0,r}|}(B_{r+1,t_r})$ contains either the word $w_{0,r}w_{1,r}^{2t_r}w_{0,r}$ or the word $w_{0,r}w_{1,r}^{2t_r+1}w_{0,r}$ as a subword, whereas no element of $\mathcal{L}_{(4t_r+4)|w_{0,r}|}(B_{r+1,t'_r})$ does. Choose some $t_{r+1} > \max\{|\mathcal{Y}_{r+1,r+1,r+1}|, t_r\}$ such that

$$\mathcal{L}_{(4t_{r+1}+4)|w_{0,r}|}(B_{r+1,t_{r+1}}) \neq \mathcal{L}_{(4t_{r+1}+4)|w_{0,r}|}(Y) \quad \text{for any } Y \in \mathcal{Y}_{r+1,r+1,r+1},$$

which is possible by (15) and the fact that $\mathcal{Y}_{r+1,r+1,r+1}$ is finite. Let $u_{r+1} \in \mathcal{L}(B_{r+1,t_{r+1}})$ be such that u_{r+1} is a concatenation of the words $w_{0,r}(w_{1,r}^{2t_{r+1}})$ and $w_{0,r}(w_{1,r}^{2t_{r+1}+1})$ and every element of $\mathcal{L}_{(4|\mathcal{Y}_{r+1,r+1,r+1}|+4)|w_{0,r}|}(B_{r+1,t_{r+1}})$ occurs as a subword of u_{r+1}. Define words

$$w_{0,r+1} := u_{r+1}(w_{0,r}w_{1,r}^{2t_{r+1}})^{(2t_{r+1}+2)|u_{r+1}|(4|\mathcal{Y}_{r+1,r+1,r+1}|+4)}$$

$$w_{1,r+1} := u_{r+1}(w_{0,r}w_{1,r}^{1+2t_{r+1}})^{(2t_{r+1}+1)|u_{r+1}|(4|\mathcal{Y}_{r+1,r+1,r+1}|+4)}.$$

These words have the following features:

(1) $|w_{0,r+1}| = |w_{1,r+1}|$;
(2) every element of $\mathcal{L}_{(4|\mathcal{Y}_{r+1,r+1,r+1}|+4)|w_{0,r}|}(B_{r+1,t_{r+1}})$ occurs as a subword of $w_{0,r+1}$ and $w_{1,r+1}$;
(3) any $\mathbb{Z}$-coloring that can be written as a bi-infinite concatenation of $w_{0,r+1}$ and $w_{1,r+1}$ can be written uniquely in this way, by identifying where the word u_{r+1} occurs;
(4) for any $a, b \in \{0, 1\}$, we have $w_{a,r+1}w_{b,r+1} \in \mathcal{L}(B_{r+1,t_{r+1}})$.

Therefore this procedure inductively defines the shift B_{r,t_r} for all $r \in \mathbb{N}$. By construction, for any $r < r'$ we have tgat

$$\mathcal{L}_{(4|\mathcal{Y}_{r,r,r}|+4)|w_{0,r}|}(B_{r,t_r}) = \mathcal{L}_{(4|\mathcal{Y}_{r,r,r}|+4)|w_{0,r}|}(B_{r',t_{r'}}).$$

Furthermore, $B_{r+1,t_{r+1}} \subseteq B_{r,t_r}$ for all $r \in \mathbb{N}$. Define the shift

$$B := \bigcap_{r=1}^{\infty} B_{r,t_r}.$$

Then for any $r \in \mathbb{N}$, we have that

$$\mathcal{L}_{(4|\mathcal{Y}_{r,r,r}|+4)|w_{0,r}|}(B) \neq \mathcal{L}_{(4|\mathcal{Y}_{r,r,r}|+4)|w_{0,r}|}(Y) \quad \text{for any } Y \in \mathcal{Y}_{r,r,r}. \tag{16}$$

We next show that B is not a factor of any element of $\mathcal{Z}(a_n)$, where the sequence (a_n) is defined by

$$a_n := 2 \cdot (4|\mathcal{Y}_{n,n,n}| + 4)|w_{0,n}|.$$

For contradiction, suppose B is a factor of some shift $X \in \mathcal{Z}(a_n)$. Let $\varphi \colon X \to B$ be a factor map and let R be its (symmetric) range. Find some $n > \max\{R, |\mathcal{L}_1(X)|\}$ such that X has no minimal forbidden words of lengths between n and $n + 2(4|\mathcal{Y}_{n,n,n}| + 4)|w_{0,n}|$. If X_n is the nth term in the SFT cover of X, then we have that

$$\mathcal{L}_{(4|\mathcal{Y}_{n,n,n}|+4)|w_{0,n}|}(B) = \mathcal{L}_{(4|\mathcal{Y}_{n,n,n}|+4)|w_{0,n}|}(\varphi(X_n)).$$

But $\varphi(X_n) \in \mathcal{Y}_{n,n,n}$ and so $\mathcal{L}_{(4|\mathcal{Y}_{n,n,n}|+4)|w_{0,n}|}(\varphi(X_n))$ is the language of an element of $\mathcal{Y}_{n,n,n}$, meaning that $\mathcal{L}_{(4|\mathcal{Y}_{n,n,n}|+4)|w_{0,n}|}(B)$ is also the language of an element of $\mathcal{Y}_{n,n,n}$. However, this contradicts Eq. (16), and so no such shift X or factor map φ can exist. $\square$

We note that the sequence $(a_n)_{n \in \mathbb{N}}$ in Proposition 5.1 is computable. For instance, from the proof it follows that one may define this sequence as

$$a_n := 2 \cdot (4|\mathcal{Y}_{n,n,n}| + 4)|w_{0,n}|,$$

where $|\mathcal{Y}_{n,n,n}|$ is the number of sofic shifts on an n letter alphabet that can be written as range n block codes of a shift of finite type, also on an n letter alphabet (and thus is less than $2^{n^{2n+1}}$) and defined with minimal forbidden words of length at most n. The number $|w_{0,n}|$ is recursively defined and depends on $|\mathcal{Y}_{r,r,r}|$ for $r \leq n$.

We conclude this section by showing that $\mathcal{Z}(a_n)$ is a large set, in a sense made precise in the next proposition. The result of the proposition, with $\mathcal{Z}(a_n)$ replaced with the set of all language stable shifts, appeared in [10, Corollary 5.2] with essentially the same proof. As the proof is short and needs a small amount of adapting to apply to $\mathcal{Z}(a_n)$, we include it here for completeness.

Proposition 5.2 *Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of positive integers and fix $a \in \mathbb{N}$. For any $0 \leq h \leq \log(a)$, the set*

$$T_h := \{Y \in \mathcal{Z}(a_n) : \mathcal{L}_1(Y) \subseteq \{0, 1, \ldots, a-1\} \text{ and } h_{top}(Y) \geq h\}$$

is a dense G_δ subset, with respect to the Hausdorff metric, of the space of all subshifts of $\{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ that have entropy at least h.

Proof Fix $0 \leq h \leq \log(a)$. Let $S \subseteq \{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ be a shift with entropy at least h. Let $(S_k)_{k \in \mathbb{N}}$ be the SFT cover of S. Note that $S_k \in \mathcal{Z}(a_n)$ for all k since S_k has only finitely many minimal forbidden words. Moreover, $h_{top}(S_k) \geq h_{top}(S)$ for all k. Finally, with respect to the Hausdorff metric, $d(S, S_k) \leq 2^{-k}$ because $\mathcal{L}_k(S) = \mathcal{L}_k(S_k)$. Therefore S is a limit point of the set of all subshifts of $\{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ that have entropy at least h. It follows that $\overline{T_h}$ is the set of all subshifts of $\{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ that have entropy at least h.

We are left with checking that T_h is a G_δ set. Note that if T_0 is a G_δ subset of the space of all subshifts of $\{0, 1, \ldots, a-1\}^{\mathbb{Z}}$, then T_h is a G_δ subset of the space of all subshifts of $\{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ that have entropy at least h. Thus it suffices to prove the claim for T_0. To do so, recall that $\mathcal{X}_{a,n}$ denotes the set of shifts of finite type on the alphabet $\{0, 1, \ldots, a-1\}^{\mathbb{Z}}$ that can be defined using forbidden words of length at most n. For each $X \in \mathcal{X}_{a,n}$, let

$$\mathcal{U}(X) := \{Y \subseteq \{0, 1, \ldots, a-1\}^{\mathbb{Z}} : \mathcal{L}_{n+a_n}(Y) = \mathcal{L}_{n+a_n}(X)\},$$

which is open with respect to the Hausdorff metric. Then we have that

$$T_0 = \bigcap_{n=1}^{\infty} \left(\bigcup_{m=n}^{\infty} \bigcup_{X \in \mathcal{X}_{a,m}} \mathcal{U}(X) \right)$$

and so T_0 is a G_δ set. □

5.2 Questions about well approximable shifts

It is natural to expect that there are some effective bounds for which language stable shifts are well approximable.

Question 5.3 Are there effective bounds on the function $A(n, n, n, n, n)$ from Definition 4.11?

It is shown in [10] that every language stable shift has a characteristic measure that is a measure of maximal entropy, and in Theorem 4.6 we showed that the same holds for any symbolic factor of a well-approximable language stable shifts. However, we do not know how general this result is, and so we ask the following question.

Question 5.4 Does every subshift factor of a language stable shift have a characteristic measure of maximal entropy?

More generally, we can ask if the same holds for any shift, but we still do not know if every shift even has a characteristic measure. This brings us to a related question about construction of characteristic measures.

Question 5.5 Assume that X is a mixing subshift and let $\varphi \in \mathrm{Aut}(X)$ be an automorphism. Let $(X_n)_{n \in \mathbb{N}}$ be the SFT cover of X and let μ_n be the unique measure of maximal entropy on X_n. If μ is a weak* limit point of $(\mu_n)_{n=1}^{\infty}$, is $\varphi_* \mu = \mu$?

Note that this question is particularly relevant for shifts that are not necessarily language stable.

6 Largeness of the class of language stable shifts

6.1 Language stable shifts are closed under bounded speedups

While language stable shifts are not closed under passage to factors, they are closed under other operations. In particular, we show that language stability is preserved by speeding up the transformation.

The classic speedup is given by taking a power of the transformation, for example (X, σ^n) is a bounded speedup of (X, σ) for any $n \in \mathbb{N}$. We consider the more general setting and for a shift (X, σ), define a *bounded speedup* is a self-homeomorphism S of X of the form

$$S(x) = \sigma^{\rho(x)}(x)$$

for some bounded function $\rho \colon X \to \mathbb{N}^*$.

The power map ρ in a bounded speed in an aperiodic system must be continuous (see [1, Proposition 2.2]), and it is known that bounded speedups preserve many properties of the initial system. For example, a bounded speedup of a subshift is expansive, and hence is conjugate to a subshift, and similarly the bounded speedup of a substitution subshift is also conjugate to a substitution shift (see [1] for these results and more background on such systems). We give a result in this spirit in Theorem 6.2: a bounded speedup of a language stable shift, and more generally of an induced map on a clopen set of a minimal language stable shift, is itself conjugate to a language stable shift.

We start with an elementary lemma used to bound the lengths of minimal forbidden words.

Lemma 6.1 *Let X be a shift. If u, v, s are words such that $uv, vs \in \mathcal{L}(X)$ but $uvs \in \mathcal{F}(X)$, then the word uvs contains a minimal forbidden word of length at least $|v|$.*

Proof If f is a subword of uvs that lies in $\mathcal{F}(X)$, contains v as a subword and is of minimal length, then f is a minimal forbidden word. □

Theorem 6.2 *Let $X \subset A^{\mathbb{Z}}$ be a shift and let $\rho \colon U \to \mathbb{N}^*$ be a continuous function such that the map $S(\cdot) = \sigma^{\rho(\cdot)}(\cdot)$ is a homeomorphism on a clopen set $U \subset X$. Then (U, S) is conjugate to a language stable shift if and only if (X, σ) is.*

It follows that for a minimal language stable shift, any induced map $\tau(\cdot) = \sigma^{\inf\{n>0 \colon \sigma^n(\cdot) \in U\}}(\cdot)$ on a clopen set U is conjugate to a language stable. Furthermore, using an argument similar to that in [1, Proposition 2.2], we can relax the hypothesis of continuity on ρ to only require that it is bounded on an aperiodic subshift.

Proof To simplify notation, we make a small abuse of notation and let $\rho \colon \mathbb{Z} \times U \to \mathbb{Z}$ denote the cocycle associated to ρ, meaning that

$$
\rho(n, x) = \begin{cases} \sum_{i=0}^{n-1} \rho(S^i x) & \text{for } n > 0 \\ 0 & \text{for } n = 0 \\ -\sum_{i=0}^{-n-1} \rho(S^{i+n} x) & \text{for } n < 0. \end{cases}
$$

Hence for any integer n and $x \in X$, we have that $S^n x = \sigma^{\rho(n,x)}(x)$.

Since ρ is a continuous function on $U \subset X$, it is locally constant. Choose an integer $N > \sup_{x \in U} |\rho(x)|$ such that ρ is constant on any cylinder defined by a word of length N. To simplify notation, for the remainder of this proof we let $J(N)$ denote the interval $[-N, N]$ and for $q \in \mathbb{Z}$ we write $q + J(N)$ for the interval $[-N + q, N + q]$, and use analogous notation for translates of intervals I, J. We claim that for sufficiently large N, the map from $\phi \colon U \to (A^{2N+1})^{\mathbb{Z}}$ given by

$$
\phi\big((x_n)_{n \in \mathbb{Z}}\big) = (x_{\rho(n,x)+J(N)})_{n \in \mathbb{Z}} \tag{17}
$$

defines a conjugacy between the induced system (U, S) and the shift (X, σ). To check this, note that since the sequence $(\rho(n, x))_{n \in \mathbb{Z}}$ is increasing and the distance between two consecutive terms is uniformly bounded, the map is a homeomorphism for any sufficiently large N. Furthermore, a direct computation shows that the map ϕ intertwines the action of $S(\cdot) = \sigma^{\rho(\cdot)}(\cdot)$ and the shift map. Let (Y, σ) denote the image of the shift (X, σ) under ϕ.

Assume first that (X, σ) is language stable. Let $\mathbf{f} = \mathbf{f}_0 \ldots \mathbf{f}_{\ell-1}$ be a minimal forbidden word for Y of length ℓ, where ℓ satisfies $\inf_{x \in U} \rho(\ell, x) > 2N$. This means that there exist $x, y \in U$ such that

$$
\begin{array}{cc}
x_{J(N)} = \mathbf{f}_0 & y_{J(N)} = \mathbf{f}_1 \\
x_{\rho(x)+J(N)} = \mathbf{f}_1 & y_{\rho(y)+J(N)} = \mathbf{f}_2 \\
\vdots & \vdots \\
x_{\rho(\ell-2,x)+J(N)} = \mathbf{f}_{\ell-2} & y_{\rho(\ell-2,y)+J(N)} = \mathbf{f}_{\ell-1} \\
x_{\rho(\ell-1,x)+J(N)} \neq \mathbf{f}_{\ell-1} & y_{\rho(-1,y)+J(N)} \neq \mathbf{f}_0.
\end{array}
$$

For $i = 0, 1$ and $j = \ell - 1, \ell - 2, \ell - 3$, let $I_{i,j}(x)$ denote the interval $I_{i,j}(x) = \bigcup_{k=i}^{j} \rho(k, x) + J(N)$, and define the interval $I_{i,j}(y)$ analogously. Note that $\rho(i, x) = \rho(i - 1, y) + \rho(x)$ for each $0 < i < \ell - 1$ and $x_{I_{1,\ell-2}(x)} = y_{I_{0,\ell-3}(y)}$. Moreover the interval $J(N)$ is not a subset of $I_{1,\ell-2}(x)$, because ρ takes only positive values. Similarly the interval $\rho(\ell - 1, x) + J(N)$ is not a subset of $I_{1,\ell-2}(x)$. Let I denote the interval $J(N) \setminus I_{1,\ell-2}(x)$, and let J denote the interval $(\rho(\ell - 1, x) + J(N)) \setminus I_{1,\ell-2}(x)$. Since ℓ is sufficiently large, the intervals I and J are disjoint.

Choose w to be the word such that

$$
w_I = x_I,
$$
$$
w_{I_{1,\ell-2}(x)} = x_{I_{1,\ell-2}(x)},
$$
$$
w_J = y_{J+\rho(x)}.
$$

By construction, we have that $w_{\rho(i,x)+[-N,N]} = \mathbf{f}_i$ for each $0 \leq i < \ell$. Since $\mathbf{f}$ is forbidden in Y, the word w is forbidden in X. It follows from Lemma 6.1 that w contains a minimal forbidden word for X of length $f(\ell)$ between $|I_{1,\ell-2}(x)| = 2N + \rho(\ell - 2, x) - \rho(x) + 1$ and $|I_{0,\ell-1}(x)| = 2N + \rho(\ell - 1, x) + 1$. Thus for two distinct lengths $\ell_1 < \ell_2$ of minimal

forbidden words, we have that

$$f(\ell_2) - f(\ell_1) \leq \rho(\ell_2 - 1, x) - (\rho(\ell_1 - 2, x) - \rho(x))$$

$$= \sum_{i=\ell_1-2}^{\ell_2-2} \rho(S^i x) + \rho(x) \leq (\ell_2 - \ell_1 + 2) \max_{x \in U} \rho(x).$$

It follows that when there are arbitrary large gaps between consecutive lengths of minimal forbidden words in X, the same holds for Y. In particular, if (X, σ) is language stable then so is (Y, σ).

Conversely, assume that (U, S) is conjugate to a language stable subshift. Since this property is invariant under conjugacy, the system (Y, σ) is also language stable, where again (Y, σ) is the system defined to be the image of the (X, σ) under the map ϕ defined in (17).

Let w be a minimal forbidden word for X of length ℓ, with this length to be determined. For $x \in U$ and integers $i < j$, let $I_{i,j}(x)$ denote the interval $\bigcup_{k=i}^{j} \rho(k, x) + J(N)$. Define n_ℓ to be the integer

$$n_\ell = \sup\{n \geq 0 : \text{there exists } x \in U \text{ such that } x_{I_{0,n}(x)} \text{ is a subword of } w\}.$$

By definition, for any $\ell > 2N + 1$, the set in this definition is nonempty. Let $x^* \in U$ be some point associated to n_ℓ. As any subword of x^* has shorter length, it follows that

$$2N + 1 + n_\ell \min_{x \in U} \rho(x) \leq |I_{0,n_\ell}(x^*)| \leq \ell - 1. \tag{18}$$

Since n_ℓ is maximal, it also follows that

$$\ell < (n_\ell + 1) \max_{x \in U} \rho(x) + 2N + 1. \tag{19}$$

Since n_ℓ is maximal, neither $x^*_{I_{-1,n_\ell}(x^*)}$ or $x^*_{I_{0,n_\ell+1}(x^*)}$ can be a subword of w. It follows from inequalities (18) and (19) that for any sufficiently large ℓ, the intervals $I_{-1,n_\ell}(x^*) \backslash I_{0,n_\ell}(x^*)$ and $I_{0,n_\ell+1}(x^*) \backslash I_{0,n_\ell}(x^*)$ are disjoint.

Let $u, s \in A^*$ be two words such that $w = u x^*_{I_{0,n_\ell}(x^*)} s$. The length of $I_{0,n_\ell+1}(x^*)$ is greater than $|x^*_{I_{0,n_\ell}(x^*)} s|$, as otherwise we can find $x \in U$ such that $x_{I_{0,n_\ell+1}(x)}$ is a subword of w, a contradiction of the maximality of n_ℓ. Similarly the length of $I_{-1,n_\ell}(x^*)$ is greater than $|u x^*_{I_{0,n_\ell}(x^*)}|$.

Since w is forbidden, the words s and u can not both be empty. Since the words $u x^*_{I_{0,n_\ell}(x^*)}$ and $x^*_{I_{0,n_\ell}(x^*)} s$ are allowable words in the language of X, there exist words u^- and s^+ in A^* such that the words $u^- u x^*_{I_{0,n_\ell}(x^*)}$ and $x^*_{I_{0,n_\ell}(x^*)} s s^+$ are allowed in X and $|u^- w s^+| = |I_{-1,n_\ell+1}(x^*)|$. Finally set w_s (respectively, w_u) to be the suffix (respectively, prefix) of length $2N + 1$ of the word $u^- w s^+$. Let $f = f_{-1} \cdots f_{n_\ell+1}$ be the word in $(A^{2N+1})^{n_\ell+3}$ defined by

$$f_i = x^*_{\rho(i,x^*)+J(N)} \quad \text{for } 0 \leq i \leq n_\ell,$$

$$f_{n_\ell+1} = w_s, \quad \text{and } f_{-1} = w_u.$$

This defines a forbidden word for Y, as otherwise the word $u^- w s^+$ occurs X. By construction, the word $f_0 \ldots f_{n_\ell}$ is allowed in Y. Because the words $u^- u x^*_{I_{0,n_\ell}(x^*)}$ and $x^*_{I_{0,n_\ell}(x^*)} s s^+$ are allowed in X, it is also straightforward to check that the words $f_{-1} \cdots f_{n_\ell}$ and $f_0 \cdots f_{n_\ell+1}$ are allowed in U. This means that f is a minimal forbidden word of Y with length $g(\ell) = n_\ell + 3$.

By Inequalities (18) and (19), as in the first part of the proof, there are arbitrary gaps between consecutive lengths of minimal forbidden words of X when this holds for Y, and language stability follows. $\square$

6.2 Many β-shifts are language stable

We describe how the class of language subshifts interacts with a well-known family parameterized by the reals, the β-shifts. We start with a brief summary of the properties of β-shifts, and refer to the survey [4] for further background.

Let $\beta > 1$ be real number. We use the greedy expansion of 1 in base β, defined by

$$d(1, \beta) = (\lfloor \beta T_\beta^n(1) \rfloor)_{n \geq 0}, \quad \text{where } T_\beta(x) = \beta x - \lfloor \beta x \rfloor.$$

When the expansion $d(1, \beta)$ is *finite*, meaning that $d(1, \beta) = d_0 \cdots d_k 00 \cdots$ for some finitely many nonzero integers $d_0, \ldots, d_k$, set

$$d^*(1, \beta) = d_0 \cdots d_{k-1}(d_k - 1)d_0 \cdots d_{k-1}(d_k - 1) \cdots$$

to be the periodic expansion. For the remainder of this section, given any finite expansion we consider the associated infinite periodic one $d^*(1, \beta)$ instead of $d(1, \beta)$. Note that $d(1, \beta)$ starts with $\lfloor \beta \rfloor$, the greatest letter of the alphabet.

Letting $\leq_{\text{lex}}$ be the lexicographic ordering, the expansions $d(1, \beta)$ are characterized in the following way:

Proposition 6.3 (see, for example, [4]) *A sequence $d \in \{0, \ldots, \lfloor \beta \rfloor\}^{\mathbb{N}}$ is the expansion of some number $\beta > 1$ if and only if*

$$\sigma^s d \leq_{lex} d, \quad \text{for all } s \geq 1.$$

The associated subshift S_β, called a β-*shift*, is defined by

$$S_\beta = \{s \in \{0, \ldots, \lfloor \beta \rfloor\}^{\mathbb{Z}} : \sigma^n(s) \leq_{\text{lex}} d(1, \beta) \quad \text{for all } n \in \mathbb{Z}\}.$$

Another equivalent way to define this subshift is by taking bi-infinite edge paths in the graph whose edges are labeled by the integers $0, 1, 2, \ldots$, with an edge from the vertex i to $i + 1$ labelled by d_i and when $d_i > 0$, there are edges between the vertices i to 0 labelled by $d_i - 1$, ..., 0.

The arithmetic properties of the real β and the dynamical properties of the shift S_β are connected. For example, the class of β such that S_β is sofic contains all Pisot numbers and is a subset of the Perron numbers. Surprisingly, S_β is a shift of finite type for $(1 + \sqrt{5})/2$ but not for $(3 + \sqrt{5})/2$. Furthermore, the shifts S_β depend continuously on the real parameter β in the Hausdorff topology. It follows that for a generic set of parameters β, the subshift S_β is language stable. We give a sufficient condition demonstrating that a large class of β-shifts are language stable. To do so, we consider the combinatorial properties of the associated expansion sequences, making use of a characterization of the associated language.

Proposition 6.4 (see, for example, [4]) *The language of a β-shift S_β is defined by concatenations of all words in the set*

$$Y = \{wb : w \text{ is a prefix of } d(1, \beta), b < d_{|w|}\}.$$

We make use of this to describe the minimal forbidden words of a β-shift.

Lemma 6.5 *For the β-shift S_β, the set of the minimal forbidden words $\mathcal{M}(S_\beta)$ is the set*

$$\{wb : w \text{ is a prefix of } d(1, \beta), b > d_{|w|}, \text{ for any strict suffix } w' \text{ of } w, w'b \leq d(1, \beta)\}.$$

Proof By definition of the subshift S_β, any word of the form wb with w a prefix of $d(1, \beta)$, $b > d_{|w|}$ is forbidden. If $w'b \leq d(1, \beta)$ for any strict suffix of w, then wb is minimal forbidden.

Conversely, assume that $u \in \mathcal{L}(S_\beta)$ is a word such that $ub \notin \mathcal{L}(S_\beta)$ is a minimal forbidden word for some letter b. By Proposition 6.4, it follows that u can be written as a concatenation of the form

$$u = w_0 b_0 w_1 b_1 \cdots w_\ell b_\ell$$

where w_i are prefixes of $d(1, \beta)$, $b_{n_i} < d_{|w_i|}$, and $b_\ell \leq d_{|w_\ell|}$. Since ub is not an allowable word and $d_0 = \lfloor \beta \rfloor \geq b$, these imply that $b_\ell = d_{|w_\ell|}$ and $b > d_{|w_\ell|+1}$. It follows that $w_\ell b_\ell b \notin \mathcal{L}(S_\beta)$. Moreover since any strict suffix of ub is allowable, we obtain the desired inclusion. □

We give a sufficient condition to ensure the existence of a minimal forbidden word with certain bounds on the length.

Lemma 6.6 *Assume that any $i \in \{i_1 + 1, i_1 + 2, \ldots, i_2\}$, satisfies $d_i < d_0$. Then for any $i_1 < n < i_2$, there exists a minimal forbidden word of length between $n - i_1$ and $n + 2$.*

Proof Consider the word $d_0 \cdots d_n \in \mathcal{L}(S_\beta)$. Since $d_{n+1} < d_0$, the word $d_0 \cdots d_n d_0$ is forbidden. However it follows from Proposition 6.4 that $d_{i_1+1} \cdots d_n d_0 \in \mathcal{L}(S_\beta)$, as each $d_j < d_0$ for $i_1 < j \leq n$. Thus it follows from Proposition 6.1 that the word $d_0 \cdots d_n d_0$ contains a minimal forbidden word of length at least $n - i_1$. □

Corollary 6.7 *For a β-shift, if the set $\{i \in \mathbb{N} : d_i = d_0\}$ is finite, then the set of lengths of minimal forbidden words is relatively dense.*

In particular the subshift S_β is not language stable.

Proof Assume that for any index i greater than D, $d_i < d_0$. Lemma 6.6 ensures there exists a length of a minimal forbidden word at distance at most $\max(D, 2)$ of any integer $n > D$. □

It follows from Corollary 6.7 that we can restrict our focus to expansion sequences of β-shifts where the letter d_0 occurs infinitely many times.

Lemma 6.8 *For a β-shift, if some nonempty word w is the prefix and suffix of some word $d_0 \cdots d_n$, then there is no minimal forbidden word of length $n + 1$.*

Proof The word w is of the form $w = d_0 \cdots d_\ell$. By contradiction, assume there is a minimal forbidden word of length $n + 1$. By Lemma 6.5, this word has the form $d_0 d_1 \cdots d_{n-1} b$ for some $b > d_n = d_\ell$. Since any strict suffix of this word is allowable, the word $d_0 d_1 \cdots d_{\ell-1} b$ is also allowable. But this is a contradiction, since $b > d_\ell$. □

When the word w is a prefix and a suffix of $d_0 \cdots d_{n-1}$, the word w is said to *occur* in $d(1, \beta)$ and the index $n - |w|$ is called an *occurence* of w. In this case, any prefix of w also occurs in $d(1, \beta)$, and so we can apply Lemma 6.8 for each length of the prefix w. This means that there is no minimal forbidden word of length $n - (i - 1)$ for any $0 \leq i < |w|$, providing gaps in the complement of lengths of minimal forbidden words. Summarizing this, we have the following corollary:

Corollary 6.9 *For a β-shift, the complement of the set of lengths of minimal forbidden words contains the set*

$$\bigcup_{w \text{ prefix of } d(1,\beta)} \{n + |w| : n \text{ is an occurrence of } w \text{ in } d(1, \beta)\}.$$

In particular, if every prefix of $d(1, \beta)$ occurs at least twice in $d(1, \beta)$, then S_β is language stable.

Schmeling [20] shows that the sequence $d(1, \beta)$ is generic for the measure of maximal entropy in S_β for Lebesgue almost every $\beta \in (1, +\infty)$. In particular, since the measure of maximal entropy has full support, any prefix of $d(1, \beta)$ occurs in $d(1, \beta)$ on a set of indices of positive density. Combining this result with Corollary 6.9, we conclude that the set of β-shifts is large.

Corollary 6.10 *The set*

$$\{\beta > 1 : S_\beta \text{ is language stable }\}$$

contains a dense G_δ in $(1, +\infty)$ of full Lebesgue measure.

Hofbauer [15] and Walters [21] show that any β-shift admits a unique measure of maximal entropy, and hence has a characteristic measure. Moreover, Climenhaga and Thompson [7] show that the same holds for any factor of a β-shift, and so any factor of a β-shift also has a characteristic measure.

We conclude with a particular example that goes beyond shifts of finite type (see [7] for the precise definitions).

Example 6.11 (*A specified/synchronized β-shift that is language stable*) We define an expansion sequence by induction using 2 sequences of words. Consider the alphabet $\{0, 1, 2\}$ and set $w_0 = 222$ and $u_0 = 111$. Define $w_{n+1} = w_n u_n w_n$ and fix some word $u_{n+1} \in \{0, 1\}^{|u_n w_n|}$ that is smaller (in lexicographic order) than $u_n w_n$. In particular, the length of this word is smaller than that of $u_n w_n$. Taking the limit of the words w_n, we obtain an infinite word w in $\{0, 1, 2\}^{\mathbb{N}}$

By construction, the sequence w satisfies $\sigma^s w \leq_{\text{lex}} w$ for any $s \geq 1$. Hence, by Proposition 6.3, the sequence w is the expansion of some number $\beta > 1$. Moreover, by construction, any prefix of w occurs infinitely many times. It follows from Corollary 6.9, the associated subshift S_β is language stable.

Additionally, at each step n, we can choose u_n to be a power of 1's. Then no 0 occurs in the sequence and so the subshift S_β is specified (see [3] or [4]). In particular, since the sequence of powers of 1 is increasing, then the sequence w is not ultimately periodic and the subshift S_β is not sofic.

Similarly, if u_n is taken to be an increasing sequence of strings of 0's, the word 111 never occurs in w but is allowable in the subshift. Then the subshift S_β is synchronizing but not specified (again, see [3] or [4]).

7 Aperiodic linear complexity shifts are language stable

We start with a general result on sufficiently low complexity shifts that is useful for checking language stability.

Theorem 7.1 *Assume that (X, σ) is an aperiodic shift. If the complexity p_X of (X, σ) satisfies*

$$\liminf_{n \to \infty} \left(p_X(n + 1) - p_X(n) \right) < \infty,$$

then the lower density of the set of lengths of bispecial words is zero.

In particular, if the complexity satisfies $p_X(n) = O(n)$, then the upper uniform density of the set of lengths of the bispecial words is zero.

Our proof uses the combinatorial result of Fine and Wilf Theorem [12]. For clarity of exposition, we first treat the linear complexity case, and then explain how to modify the argument to cover the general statement.

For clarity, we also include a standard method for removing the transitivity assumption in a combinatorial result of Cassaigne [5] on shifts with linear complexity.

Lemma 7.2 *Let $X \subset \mathcal{A}^{\mathbb{Z}}$ be an aperiodic subshift whose complexity satisfies*

$$p_X(n) \leq Ln + p_X(1) \quad \text{for all } n \in \mathbb{N}$$

for some $L \in \mathbb{N}$. Then there is a constant $K > 0$ such that $\sup_n p_X(n+1) - p_X(n) \leq K$.

Proof By [9, Lemma 3.4], X is the union of (at most) $L - 1$ transitive shifts $X_1, \ldots, X_\ell$. For each transitive shift, a result of Cassaigne [5] implies that $p_{X_i}(n + 1) - p_{X_i}(n) \leq 8L^3 + 8L^2 + 2L$ for any $n \in \mathbb{N}$. Hence there are at most $8L^3 + 8L^2 + 2L$ (left-) special word of length n in $\mathcal{L}(X_i)$. It follows there are at most $(L - 1)(8L^3 + 8L^2 + 2L)$ (left-) special word of length n in $\mathcal{L}(X)$. Since $p_X(n + 1) - p_X(n)$ is bounded by the product of $|\mathcal{A}|$ and the number of (left-) special words of length n in X, the statement follows. □

Proof of Theorem 7.1 We first consider the case when the shift (X, σ) has linear complexity, and so by Lemma 7.2 there exists some constant $K \in \mathbb{N}$ such that $p_X(n + 1) - p_X(n) \leq K$ for all $n \in \mathbb{N}$. It follows that for each $n \in \mathbb{N}$, there are at most K right special (respectively, left special) words of length n.

We proceed by contradiction. Assume that the set F of lengths of the bispecial words of X has upper uniform density α for some $\alpha > 0$. Choose $k \in \mathbb{N}$ such that $k\alpha/2 > K^2$. Since X is aperiodic, there is some power $p > 0$ such that for any word u of length at most k, the word u^p is forbidden. By assumption, we can choose sufficiently large $n \in \mathbb{N}$, in particular greater than $(p + 1)k$, such that the interval $[n + 1, n + k]$ contains at least $k\alpha/2$ different lengths of bispecial words. Let $\mathtt{f}_1, \ldots, \mathtt{f}_{\lfloor k\alpha/2 \rfloor}$ denote these bispecial words.

By construction, each prefix and each suffix of each $\mathtt{f}_i$ is a special word. Thus by the Pigeonhole Principle, there are two indices $0 < i < j \leq K^2 + 1$ such that $\mathtt{f}_i$ and $\mathtt{f}_j$ share the same prefix $\mathtt{p}$ and share the same suffix $\mathtt{s}$ of length n. It follows that the suffixes of $\mathtt{p}$ of lengths $2n - |\mathtt{f}_i|$ and $2n - |\mathtt{f}_i|$ are both prefixes of $\mathtt{s}$. Thus, by the Fine-Wilf Theorem [12], the largest of these prefixes, of length at least $n - k$, is of the form $\mathtt{u}^\ell \mathtt{u}'$ where $\mathtt{u}$ and $\mathtt{u}'$ are words such that $|\mathtt{u}| = ||\mathtt{f}_j| - |\mathtt{f}_i||(\leq k)$, $|\mathtt{u}'| < ||\mathtt{f}_j| - |\mathtt{f}_i||$, and

$$\ell \geq \lfloor (n - k)/||\mathtt{f}_j| - |\mathtt{f}_i|| \rfloor \geq (n - k)/k.$$

However, this contradicts the choice of n.

We now turn to the general case: assume the set $A = \{n : p_X(n + 1) - p_X(n) \leq K\}$ is infinite for some $K \in \mathbb{N}$. Hence there are at most K right special words of length $n \in A$, and the same bound holds for the number of left special words. Again, we proceed by contradiction, assuming that the set F of lengths of the bispecial words of X has lower density α for some $\alpha > 0$. We choose $k \in \mathbb{N}$ as in the linear complexity case and note that again there exist arbitrarily large integers $n \in A$, and in particular larger than $(p + 1)k$, such that the interval $[n, n + k]$ contains at least $k\alpha/2$ different lengths of bispecial words. We then continue exactly as previously. □

Corollary 7.3 *Any aperiodic shift with complexity satisfying*

$$\liminf_{n \to \infty} (p_X(n + 1) - p_X(n)) < \infty,$$

is language stable.

It follows immediately that any aperiodic shift with linear complexity is language stable.

Proof A minimal forbidden word has the form awb for some letters a, b in the alphabet and some bispecial word w in the language. By Theorem 7.1, the set of lengths of bispecial words has lower density zero, and language stability follows. □

We note that the assumption of aperiodicity in Corollary 7.3 is necessary, as can be seen by considering the closure under the shift of the sequence with a single 1 and all other entries 0 (note that this system contains the fixed point of all 0s). The complexity of this system satisfies $p(n) = n + 1$ for all $n \in \mathbb{N}$, while the words $10^m 1$ are minimal forbidden words for all $m \in \mathbb{N}$ and so the system is not language stable.

Example 7.4 *(An aperiodic non-language stable shift with subquadratic complexity)* Let τ be the substitution defined by $0 \mapsto 010$ and $1 \mapsto 11$, and let (X, σ) be the associated shift defined by $x \in X$ if every word in x is a subword of $\tau^n(0)$ for some $n \geq 0$. As we consider the fixed point of 0, this shift is aperiodic, and the complexity of the shift is approximately $n \log_2 \log_2 n$. Cassaigne [6, Sect. 4.4] shows that for any ℓ with $\ell \geq 3$ not a power of 2, this shift contains a minimal forbidden word of length ℓ (precisely a non-strict bispecial word) and so this shift is not language stable.

We note that the shift defined in this example is not minimal. This leads to a natural question, namely if every minimal shift is language stable. However, we believe that this is unlikely to hold, and so we ask:

Question 7.5 Assuming that not every minimal shift is language stable, what is the complexity threshold for the existence of a minimal shift that is not language stable?

References

1. Alvin, L., Ash, D., Ormes, N.: Bounded topological speedups. Dyn. Syst. **33**(2), 303–331 (2018)
2. Béal, M.P., Mignosi, F., Restivo, A.: Minimal forbidden words and symbolic dynamics. In: STACS 96 (Grenoble, 1996). Lecture Notes in Comput. Sci., vol. 1046, pp. 555–566. Springer, Berlin (1996)
3. Bertrand-Mathis, A.: Développement en base θ; répartition modulo un de la suite $(x\theta_n)_{n \geq 0}$; langages codés et θ-shift. Bull. Soc. Math. France **114**(3), 271–323 (1986)
4. Blanchard, F.: β-expansions and symbolic dynamics. Theor. Comput. Sci. **65**(2), 131–141 (1989)
5. Cassaigne, J.: Special factors of sequences with linear subword complexity. In: Developments in Language Theorem, II (Magdeburg, 1995), pp. 25–34. World Sci. Publ, River Edge, NJ (1996)
6. Cassaigne, J.: Complexité et facteurs spéciaux. Journées Montoises (Mons, 1994). Bull. Belg. Math. Soc. Simon Stevin **4**(1), 67–88 (1997)
7. Climenhaga, V., Thompson, D.J.: Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors. Israel J. Math. **192**(2), 785–817 (2012)
8. Climenhaga, V., Thompson, D.J.: Beyond Bowen's specification property. In: Thermodynamic Formalism: CIRM Jean-Morlet Chair, Fall 2019, pp. 3–82. Springer International Publishing (2021)
9. Cyr, V., Kra, B.: The automorphism group of a shift of linear growth: beyond transitivity. Forum Math. Sigma **3**, 27 (2015)
10. Cyr, V., Kra, B.: Characteristic measures for language stable subshifts. Monatsh. Math. **201**, 659–701 (2023)
11. Coven, E., Smital, M.: Entropy minimality. Acta Math. Univ. Comenianea Vol. LXII **1**, 117–121 (1993)
12. Fine, N., Wilf, H.: Uniqueness theorem for periodic functions. Proc. Am. Math. Soc. **16**, 109–114 (1965)
13. Frisch, J., Tamuz, O.: Characteristic measures of symbolic dynamical systems. Ergodic Theory Dyn. Syst. **42**(5), 1655–1661 (2022)
14. Hedlund, G.A.: Endomorphisms and automorphisms of the shift dynamical system. Math. Syst. Theory **3**, 320–375 (1969)

15. Hofbauer, F.: On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. Israel J. Math. **34**, 213–237 (1979)
16. Kitchens, B.P.: Symbolic Dynamics. One-Sided, Two-sided and Countable State Markov Shifts. Universitext. Springer-Verlag, Berlin (1998)
17. Kryloff, N., Bogoliouboff, N.: La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. Ann. Math. (2) **38**(1), 65–113 (1937)
18. Lind, D., Marcus, B.: An Introduction to Symbolic Dynamics and Coding. Cambridge University Press, Cambridge (1995)
19. Parry, W.: Intrinsic Markov chains. Trans. Am. Math. Soc. **112**, 55–66 (1964)
20. Schmeling, J.: Symbolic dynamics for β-shifts and self-normal numbers. Ergodic Theory Dyn. Syst. **17**(3), 675–694 (1997)
21. Walters, P.: Equilibrium states for β-transformations and related transformations. Math. Z. **159**(1), 65–88 (1978)
22. Walters, P.: An Introduction to Ergodic Theory. Graduate Texts in Mathematics, vol. 79. Springer-Verlag, New York-Berlin (1982)
23. Weiss, B.: Subshifts of finite type and sofic systems. Mont. Math. **77**, 462–474 (1973)