Interplay between combinatorics and dynamical properties in minimal symbolic systems.

Paulina Cecchi Bernales.
(joint work with V. Berthé and B. Espinoza)

Departamento de Matemáticas-Facultad de Ciencias-Universidad de Chile.

One World Combinatorics on Words Seminar

March 31, 2026.

Framework: Symbolic dynamics

- A (topological) dynamical system is a pair (X, T), consisting of a compact metric space X together with a homeomorphism T acting on X.
- (X, T) is called *minimal* if $\forall F \subseteq X$ which is closed and T-invariant ($T^{-1}F = F$), either $F = X$ or $F = \emptyset$.
- Equivalently, $\forall x \in X$, the *orbit* $\mathcal{O}(x) := \{T^n(x) \mid n \in \mathbb{Z}\}$ is a dense subset.
- Let $\mathcal{A}$ be a finite alphabet and $X \subseteq \mathcal{A}^{\mathbb{Z}}$ a set of infinite words on $\mathcal{A}$. Endow $\mathcal{A}^{\mathbb{Z}}$ with the product topology of the discrete topology on $\mathcal{A}$, and consider the *shift* transformation on $\mathcal{A}$,

$$S((x_k)_{k \in \mathbb{Z}}) = x_{k+1}.$$

- If X is closed and shift invariant, then (X, S) defines a topological dynamical system, called a *subshift* or a *symbolic system*.

Languages

- Let $x = (x_k)_{k\in\mathbb{Z}}$ be an infinite word in $\mathcal{A}$. The *language* $\mathcal{L}(x)$ of x is the set of finite words occurring in x, also called *factors* of x.
- The *language* of a subshift (X, S) is $\mathcal{L}(X) = \bigcup_{x\in X} \mathcal{L}(x)$.
- Languages and subshifts are the same: given $\overline{x} \in \mathcal{A}^{\mathbb{Z}}$, define

$$X_{\overline{x}} := \{x = (x_k)_{k\in\mathbb{Z}} \mid \mathcal{L}(x) \subseteq \mathcal{L}(\overline{x})\}.$$

- One can also define $\mathcal{L}(x)$ for one-sided sequences, and then the two-sided associated subshift.
- Consider, for instance, a one-sided infinite sequence $x = (x_k)_{k\in\mathbb{N}}$,

$$x_F = abaababaabaab...$$

and then define the associated subshift

$$X_{x_F} := \{x = (x_k)_{k\in\mathbb{Z}} \mid \mathcal{L}(x) \subseteq \mathcal{L}(x_F)\}.$$

Minimality, recurrence

- Consider a minimal subshift (X, S). For $w \in \mathcal{L}(X)$, denote $|w|$ its *length*: number of letters.

- Since all shift-orbits are dense, one has that $\forall w \in \mathcal{L}(X)$,
  1. $w \in \mathcal{L}(x)$ for any $x \in X$,
  2. $\forall x \in X$, $\exists N \in \mathbb{N}$ s.t. $v \in \mathcal{L}(x)$ and $|v| \geq N \implies w$ occurs on v.

- Thus, in a minimal subshift all languages are equal and all sequences are *uniformly recurrent*.

- For $n \in \mathbb{N}$, let $\mathcal{L}_n(X)$ denote the set $\mathcal{L}(X) \cap \mathcal{A}^n$.

Example: substitutive subshifts

- Let $\mathcal{A}$ be an alphabet and $\mathcal{A}^*$ the free monoid of finite words on $\mathcal{A}$. A morphism $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ which is *non-erasing* and *letter-onto* is called a *substitution*.

- If $\min_{a\in\mathcal{A}} |\sigma^n(a)| \nearrow \infty$ as $n \to \infty$, we say that the substitution σ is *everywhere growing*. Everywhere growing substitutions produce infinite sequences, thus subshifts.

- Ex: $\sigma_F : \{a,b\}^* \to \{a,b\}^*$, $\sigma_F(a) = ab, \sigma_F(b) = a$ and extended by concatenation.

- Given a substitution σ on $\mathcal{A}$, define the substitutive subshift

$$X_\sigma := \{x = (x_k)_{k\in\mathbb{Z}} \mid \forall w \sqsubseteq x, \exists n \in \mathbb{N}, a \in \mathcal{A} \text{ s.t. } w \sqsubseteq \sigma^n(a)\}.$$

- A subtitution σ is *primitive* if there exists $N \in \mathbb{N}$ such that $\forall a, b \in \mathcal{A}$, a appears in $\sigma^N(b)$. In this case, (X_σ, S) is minimal.

Interplay between combinatorics and dynamics

- We already know: (X, S) minimal $\iff \forall x \in X$, x uniformly recurrent.
- Let (X, T) be a dynamical system. A Borel probability measure μ on X is called an *invariant measure* if $\forall A \in \mathcal{B}(X)$, $\mu(T^{-1}A) = \mu(A)$.
- An invariant measure μ is said to be *ergodic* if whenever $T^{-1}A = A$, either $\mu(A) = 0$ or $\mu(A) = 1$.
- Denote by $\mathcal{M}(X, T)$ the set of invariant measures of (X, T). It is a non-empty set. If $|\mathcal{M}(X, T)| = 1$, (X, T) is called a *uniquely ergodic* system.
- What is the combinatorial property which reflects unique ergodicity in subshifts?

Frequencies and unique ergodicity

- For any sequence $x \in X$, w a factor of x and $k \leq \ell \in \mathbb{Z}$, let $|x_k \dots x_\ell|_w$ denote the number of times that w appears in $x_k \dots x_\ell$.

- The *frequency* of w in x is the following limit, if it exists,

$$f_w(x) = \lim_{n\to\infty} \frac{|x_{-n} \dots x_0 \dots x_n|_w}{2n+1}.$$

- By the Ergodic Theorem, for any ergodic measure $\mu \in \mathcal{M}(X,S)$, $f_w(x)$ equals $\mu([w])$ for μ-a.e. $x \in X$, where

$$[w] = \{x \in X \mid x_0 \dots x_{|w|-1} = w\}.$$

- The subshift (X,S) is uniquely ergodic if and only if every factor w has *uniform frequencies*, in which case $f_w(x) = \mu([w])$ for all $x \in X$.

Factor complexity and invariant measures

- $\mathcal{M}(X,T)$ is a convex space whose extreme points are the ergodic measures of the system.

- How many ergodic measures does a dynamical system admit?

- Consider a subshift X on the alphabet $\mathcal{A}$. The *factor complexity* of (X,S) is the function $p_X\colon \mathbb{N} \to \mathbb{N}$ given by

$$p_X(n) = |\mathcal{L}_n(X)|.$$

- **Theorem [Boshernitzan '84]** For a minimal subshift (X,S), if $\liminf_{n\to\infty} \frac{p_X(n)}{n} < \alpha < +\infty$, then (X,S) admits at most $\max\{\lfloor \alpha \rfloor, 1\}$ ergodic measures.

Continuous eigenvalues

- A complex number λ is a *continuous eigenvalue* of (X, T) if $\exists$ non-zero continuous function $f \colon X \to \mathbb{C}$ such that $f \circ T = \lambda f$.
- We consider *additive* eigenvalues: $\alpha \in \mathbb{R}$ such that $\exists$ non-zero continuous function $f \colon X \to \mathbb{R}/\mathbb{Z}$ such that $f \circ T = f + \alpha \pmod{\mathbb{Z}}$.
- Let $E(X, T) \subseteq \mathbb{R}$ be the (additive) subgroup of additive continuous eigenvalues of (X, S).
- $E(X, T)$ is preserved under conjugacy of dynamical systems.
- We are interested deciding whether a given λ/α is an eigenvalue.
- We are also interested in knowing the rank of $E(X, T)$ over $\mathbb{Q}$, denoted $\dim_{\mathbb{Q}}(E(X, T))$.

Coboundaries I: description of continuous eigenvalues

- Let (X, S) be a minimal subshift. A word $w \in \mathcal{L}(X)$ is called a *return word* to the letter $a \in \mathcal{A}$ if
  - w starts with a,
  - $wa \in \mathcal{L}(X)$,
  - there are no other occurrences of a in $w_1 \dots w_{|w|-1}$.
- A letter coboundary of a minimal subshift (X, S) is a morphism $c\colon \mathcal{A}^\star \to \mathbb{R}/\mathbb{Z}$ such that for every return word $w \in \mathcal{L}(X)$, $c(w) = 0$.
- A coboundary c is *trivial* is $c(a) = 0$ for each letter $a \in \mathcal{A}$.
- For instance, consider the substitution $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ on $\mathcal{A} = \{a, b, c\}$ given by

$$\sigma : \begin{cases} a & \mapsto aba \\ b & \mapsto cb \\ c & \mapsto cba. \end{cases}$$

  Set $h(a) = 1$, $h(b) = -1$, $h(c) = 0$. The word bac is a return word to b, $h(bac) = -1 + 1 + 0 = 0$.

Coboundaries I: description of continuous eigenvalues

- **Theorem [Host '86]** Let σ be a primitive substitution which is injective on the letters and *recognizable*. Then, $\lambda \in \mathbb{C}$ is an eigenvalue of (X_σ, S) if and only if $\exists p \in \mathbb{N}^*$ such that, for all $a \in \mathcal{A}$,

$$h(a) := \lim_{n \to \infty} \lambda^{|\sigma^{np}(a)|}$$

exists and define a letter coboundary.[1]

- *Recongizable* means essentially that any sequence $x \in X_\sigma$ can be uniquely decomposed into elementary words of the form $\sigma^k(a)$, $k \geq 1$, $a \in \mathcal{A}$.

- When is it possible to say that any letter coboundary of a given (X, S) is trivial?

[1] This is written in multiplicative form.

Extension graphs

- Let (X, S) be a subshift on the alphabet $\mathcal{A}$. We consider two copies $\mathcal{A}_L$ and $\mathcal{A}_R$ of $\mathcal{A}$, with canonical bijections $a \mapsto a_L$ and $a \mapsto a_R$ from $\mathcal{A}$ to $\mathcal{A}_L$ and $\mathcal{A}_R$, respectively.

- Let $w \in \mathcal{L}(X)$, let

$$E_L(w) = \{a_L \in \mathcal{A}_L : aw \in \mathcal{L}(X)\},$$

$$E_R(w) = \{b_R \in \mathcal{A}_R : wb \in \mathcal{L}(X)\}.$$

- The *extension graph* $\Gamma_X(w)$ of w in X is the undirected bipartite graph whose vertices are $E_L(w) \cup E_R(w)$ and whose edges are

$$B(w) = \big\{(a_L, b_R) \in \mathcal{A}_L \times \mathcal{A}_R : awb \in \mathcal{L}(X)\big\}.$$

Extension graphs

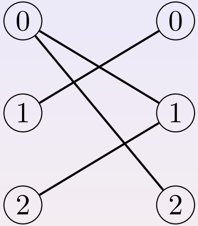
- Example: Consider the substitution $\sigma\colon \mathcal{A}^* \to \mathcal{A}^*$ on $\mathcal{A} = \{0,1,2\}$ given by

$$\sigma : \begin{cases} 0 & \mapsto 010 \\ 1 & \mapsto 21 \\ 2 & \mapsto 210 \end{cases}$$

- Let X_σ be the substitutive subshift associated to σ.
- Observe that $\{01, 10, 21, 02\}$ is the set of words of length 2 in the language of X_σ
- How does $\Gamma_{X_\sigma}(\varepsilon)$ look like?

Extension graphs

$\Gamma_{X_\sigma}(\varepsilon)$



- A *connected component* in $\Gamma_X(w)$ is a subset $K \subseteq \mathcal{A}_L \cup \mathcal{A}_R$ which is maximal (for the inclusion) such that every pair of vertices in K are connected by a path in $\Gamma_X(w)$.
- We will see that coboundaries are closely related to $\Gamma_X(\varepsilon)$.
- Let $\mathcal{C}_X$ denote the set of letter coboundaries of (X, S), let r denote the number of connected components of $\Gamma_X(\varepsilon)$.

Extension graphs and coboundaries

Theorem (Berthé-CB.-Espinoza '25)

Let (X, S) be a transitive subshift (at least one dense orbit). Then, as an $\mathbb{R}$-vector space, $\mathcal{C}_X$ has dimension $r-1$. In particular, $\Gamma_X(\varepsilon)$ is connected if and only if every coboundary of (X, S) is trivial.

- Let (X, S) be a subshift. For $n \in \mathbb{N}$ and $\mu \in \mathcal{M}(X, S)$, let

$$d(n, \mu) = \dim_{\mathbb{Q}} \langle \{\mu([w]) \mid w \in \mathcal{L}_n(X)\} \rangle.$$

- Let $\mathcal{K}_n$ denote the set of connected components in all the extension graphs of words $w \in \mathcal{L}_n(X)$

Lemma (Berthé-CB.-Espinoza '25)

Let (X, S) be a transitive subshift, let $\mu \in \mathcal{M}(X, S)$. Then, for each $n \geq 1$,

$$d(\mu, n) \leq p_X(n) - |\mathcal{K}_{n-1}| + 1.$$

Extension graphs and coboundaries

- The previous lemma allows us to give a sufficient condition for having trivial coboundaries.

Proposition

Let (X,S) be a minimal subshift on the alphabet $\mathcal{A}$, such that there exists $\mu \in \mathcal{M}(X,S)$ for which the vector $(\mu([a]))_{a\in\mathcal{A}}$ has $\mathbb{Q}$-independent coordinates. Then, any coboundary of (X,S) is trivial.

- This follows from

$$d(\mu,n) \leq p_X(n) - |\mathcal{K}_{n-1}| + 1,$$

for $n = 1$,

$$d(\mu,1) = |\mathcal{A}| \leq |\mathcal{A}| - |\mathcal{K}_0| + 1 \implies |\mathcal{K}_0| \leq 1.$$

Coboundaries II: the size of $E(X,S)$

Proposition

Let (X,S) be a minimal subshift. Then,

$$\dim_{\mathbb{Q}} E(X,S) \leq \liminf_{n\to\infty}(p_X(n) - |\mathcal{K}_{n-1}|) + 1.$$

- This follows from the general fact that

$$E(X,S) \subseteq I(X,S) = \bigcap_{\mu\in\mathcal{M}(X,S)} \langle\{\mu([w]) \mid w \in \mathcal{L}(X)\}\rangle.$$

Coboundaries II: the size of $E(X,S)$

- In some cases, $I(X,S)$ is generated by letter frequencies,

$$I(X,S) = \bigcap_{\mu \in \mathcal{M}(X,S)} \langle \{\mu([a]) \mid a \in \mathcal{A}\} \rangle.$$

- In this case,

$$\dim_{\mathbb{Q}} E(X,S) \leq \dim_{\mathbb{Q}}(I(X,S)) \leq d(\mu,1) \leq |\mathcal{A}| - r + 1,$$

where r is the number of connected components of $\Gamma(\varepsilon)$.

- This is the case, for instance, of primitive *unimodular* substitutions: those for which the indicence matrix M_σ verifies

$$\det(M_\sigma) = \pm 1.$$

(Recall that M_σ is the $|\mathcal{A}| \times |\mathcal{A}|$ matrix defined by $M_\sigma(a,b) = |\sigma(b)|_a$).

Balance

- Let $x \in \mathcal{A}^{\mathbb{Z}}$ be a infinite word, let $w \in \mathcal{L}(x)$. The word x is said to be *balanced* on w if there exists $C_w > 0$ such that for all $v, v' \in \mathcal{L}(x)$ with $|v| = |v'|$, one has

$$||v|_w - |v'|_w| \leq C_w.$$

- A minimal subshift (X, S) is said to be *balanced on factors* if every $x \in X$ is balanced on every factor.

- It is said to be *letter-balanced* if every $x \in X$ is balanced on every letter $a \in \mathcal{A}$.

- Sturmian sequences are exactly the infinite aperiodic binary words which are 1-letter balanced ($C_0 = C_1 = 1$)

Balance in substitutive sequences

- Dynamically, for a sequence x to be balanced on the factor w means that the Birkhoff sums of $\mathbf{1}_{[w]}$ have *bounded deviation*,

$$|\sum_{k=0}^{n-1} \mathbf{1}_{[w]}(S^k x) - n\mu([w])| < C_w.$$

- In particular, balanced subshifts are uniquely ergodic and the ergodic averages tend fast to the frequency of a given word.

- If (X, S) is balanced on the factor $w \in \mathcal{L}(X)$, then $\mu([v])$ is an additive topological eigenvalue of (X, S).

- This implies that in balanced systems, $E(X, S) = I(X, S)$.

A characterization using coboundaries

- Let σ be a primitive aperiodic substitution on the alphabet $\mathcal{A}$. The incidence matrix M_σ acts on $\mathbb{R}^{\mathcal{A}}$ on the right,

$$\vec{v}M_\sigma = (\sum_{b\in\mathcal{A}} v_b M_\sigma(b,a))_{a\in\mathcal{A}}.$$

- The *stable space* of σ is the vector subspace

$$V_\sigma := \{v \in \mathbb{R}^{\mathcal{A}} \mid \vec{v}M_\sigma^n \to 0 \text{ as } n \to \infty\}.$$

- The *coboundary space* of σ is the vector subspace

$$C_\sigma = \{(c(a))_{a\in\mathcal{A}} \mid c\colon A^* \to \mathbb{R}/\mathbb{Z} \text{ is a letter coboundary}\}.$$

Theorem (Berthé-CB.-Espinoza '25)

Let σ be a primitive aperiodic substitution on the alphabet $\mathcal{A}$. Then, X_σ is letter-balanced if an only if $V_\sigma + C_\sigma$ has codimension 1 in $\mathbb{R}^{\mathcal{A}}$.

Minimal complexity

- **Theorem [Tijdeman '99]**: Let (X, S) be a transitive subshift, let $\mu \in \mathcal{M}(X, S)$ and $t = \dim_{\mathbb{Q}}\{\mu([a]) \mid a \in \mathcal{A}\}$. Then,

$$p_X(n) \geq (n-1)(t-1) + |\mathcal{A}| \quad \forall n \geq 1.$$

- Proof: Let $n \geq 1$. For $j \geq 1$,
  - $t = d(\mu, 1) \leq d(\mu, j)$,
  - $|\mathcal{K}_{j-1}| \geq p_X(j-1)$ since each extension graph of a word $w \in \mathcal{L}_{j-1}(X)$ has at least one connected component.
  - Thanks to the previous lemma, for all $j \geq 1$,

$$t \leq d(\mu, j) \leq p_X(j) - |\mathcal{K}_{j-1}| + 1 \leq p_X(j) - p_X(j-1) + 1.$$

  - Summing from $j = 2$ to $j = n$, one gets

$$(n-1)t \leq p_X(n) - p_X(1) + (n-1) = p_X(n) - |\mathcal{A}| + (n-1),$$

$$\iff p_X(n) \geq (n-1)(t-1) + |\mathcal{A}|.$$

□

Minimal complexity

Theorem (Andrieu-Cassaigne '22/'26, Berthé-CB.-Espinoza '25)

Let (X, S) *be a transitive subshift, let* $\mu \in \mathcal{M}(X, S)$. *Let* $t = \dim_{\mathbb{Q}}\{\mu([a]) \mid a \in \mathcal{A}\}$. *If*

$$p_X(n) = (n-1)(t-1) + |\mathcal{A}| \quad \forall n \geq 1$$

and $t = |\mathcal{A}|$, *then every extension graph in* (X, S) *is a tree.*

- The converse is false. There exists minimal subshifts for which $\{\mu([a]) \mid a \in \mathcal{A}\}$ are rationally dependent and all the extension graphs are trees.

Dendric subshifts

- A subshift (X, S) verifying that $\forall w \in \mathcal{L}(X)$, $\Gamma_X(w)$ is a tree, is called a *dendric shift*.

- Dendric subshifts where introduced a decade ago by Berthé *et al.* and they constitute a model for a generalization of Sturmian subshifts.

- The family of Dendric subshifts include Sturmian subshifts, Arnoux-Rauzy subshifts and codings of interval exchange transformations.

- Dendric subshifts have factor complexity

$$p_X(n) = (|\mathcal{A}| - 1)n + 1.$$

[Berthé *et al.* 2014].

Dendricity, factor complexity and independence of frequencies

- One has the following relations,
  - Dendric + $\mathbb{Q}$-independent frequencies $\implies$ minimal complexity.
  - $\mathbb{Q}$-independent frequencies + minimal complexity $\implies$ dendric.
  - Dendric + minimal complexity $\implies$ $\mathbb{Q}$-independent frequencies.

Recap

- Extension graphs and coboundaries are very interesting objects.
- They are closely related to factor complexity and frequencies.
- From this relation one recovers relevant information to decide whether a given λ is a continuous eigenvalue of a primitive substitutive system.
- They also provide bounds for the size of the $\mathbb{Q}$-vector space generated by continuous eigenvalues.
- They allow us to characterize letter-balance in primitive substitutive subshifts.
- They allow to significantly simplify the proof of Tijdeman's Theorem on minimal complexity.
- They allow us to provide sufficient conditions for being dendric.

Thank you for your attention!!