Coboundaries and balance in words

Paulina CECCHI B.
(Joint work with **Valérie Berthé**)

Institute de Recherche en Informatique Fondamentale
Université Paris Diderot - Paris 7

Departamento de Matemática y Ciencia de la Computación
Facultad de Ciencia. Universidad de Santiago de Chile

Let $\mathcal{A}$ be a finite alphabet and consider $x \in \mathcal{A}^{\mathbb{Z}}$,

$$x = \cdots x_{-2}x_{-1}x_0x_1x_2x_3\cdots$$

Let v be a factor of x. We are interested in the behavior of the function

$$B^v(n) = \max_{|w|=|w'|=n} \{||w|_v - |w'|_v|\}$$

If this quantity is bounded, x is balanced on v.

Let $\mathcal{A}$ be a finite alphabet and consider $x \in \mathcal{A}^{\mathbb{Z}}$,

$$x = \cdots x_{-2}x_{-1}x_0x_1x_2x_3\cdots$$

Let v be a factor of x. We are interested in the behavior of the function

$$B^v(n) = \max_{|w|=|w'|=n} \{||w|_v - |w'|_v|\}$$

If this quantity is bounded, x is balanced on v.

Question: given x and v, is x balanced on v?

Outline

- Balance
- Coboundaries
- Substitution systems
  - Rational Frequencies
  - Examples

Balance

Balance

- (X, T) is a minimal subshift on the finite alphabet $\mathcal{A}$.
  - $\rightsquigarrow$ Every $x \in X$ is uniformly recurrent: every factor of x occurs infinitely often with boundend gaps.
  - $\rightsquigarrow$ $\mathcal{L}(x) = \mathcal{L}(X) \quad \forall x \in X$

Balance

- (X, T) is a minimal subshift on the finite alphabet $\mathcal{A}$.
  - $\rightsquigarrow$ Every $x \in X$ is uniformly recurrent: every factor of x occurs infinitely often with boundend gaps.
  - $\rightsquigarrow$ $\mathcal{L}(x) = \mathcal{L}(X) \quad \forall x \in X$

- A word $x \in \mathcal{A}^{\mathbb{Z}}$ is balanced on the factor $v \in \mathcal{L}(x)$ if there exists a constant C_v such that for every pair of factors u, w in $\mathcal{L}(x)$ with $|u| = |w|$,

$$||u|_v - |w|_v| \leq C_v.$$

Balance

- (X, T) is a minimal subshift on the finite alphabet $\mathcal{A}$.
  - $\rightsquigarrow$ Every $x \in X$ is uniformly recurrent: every factor of x occurs infinitely often with boundend gaps.
  - $\rightsquigarrow$ $\mathcal{L}(x) = \mathcal{L}(X) \quad \forall x \in X$

- A word $x \in \mathcal{A}^{\mathbb{Z}}$ is balanced on the factor $v \in \mathcal{L}(x)$ if there exists a constant C_v such that for every pair of factors u, w in $\mathcal{L}(x)$ with $|u| = |w|$,
$$||u|_v - |w|_v| \leq C_v.$$

- Since (X, T) is minimal, balance is a property of the language.

Balance

- (X, T) is a minimal subshift on the finite alphabet $\mathcal{A}$.
  - $\rightsquigarrow$ Every $x \in X$ is uniformly recurrent: every factor of x occurs infinitely often with boundend gaps.
  - $\rightsquigarrow$ $\mathcal{L}(x) = \mathcal{L}(X) \quad \forall x \in X$

- A word $x \in \mathcal{A}^{\mathbb{Z}}$ is balanced on the factor $v \in \mathcal{L}(x)$ if there exists a constant C_v such that for every pair of factors u, w in $\mathcal{L}(x)$ with $|u| = |w|$,

$$||u|_v - |w|_v| \leq C_v.$$

- Since (X, T) is minimal, balance is a property of the language.
- Sturmian words are exactly the 1-balanced words on the letters.

Invariant measures and frequencies

- A probability measure μ on X space is T-invariant if for all Borel subset B of X,

$$\mu(T^{-1}(B)) = \mu(B)$$

Invariant measures and frequencies

- A probability measure μ on X space is T-invariant if for all Borel subset B of X,

$$\mu(T^{-1}(B)) = \mu(B)$$

$\rightsquigarrow$ μ ergodic: $T^{-1}(B) = B$ implies $\mu(B)$ is either 0 or 1.

Invariant measures and frequencies

- A probability measure μ on X space is T-invariant if for all Borel subset B of X,

$$\mu(T^{-1}(B)) = \mu(B)$$

$\rightsquigarrow$ μ ergodic: $T^{-1}(B) = B$ implies $\mu(B)$ is either 0 or 1.

$\rightsquigarrow$ (X, T) uniquely ergodic: X admits a unique T-invariant measure.

Invariant measures and frequencies

- A probability measure μ on X space is T-invariant if for all Borel subset B of X,

$$\mu(T^{-1}(B)) = \mu(B)$$

  - $\rightsquigarrow$ μ ergodic: $T^{-1}(B) = B$ implies $\mu(B)$ is either 0 or 1.
  - $\rightsquigarrow$ (X, T) uniquely ergodic: X admits a unique T-invariant measure.

- A word $x \in \mathcal{A}^{\mathbb{Z}}$ has uniform frequencies if for every $w \in \mathcal{L}(x)$, the ratio

$$\frac{|x_k \cdots x_{k+n}|_w}{n+1}$$

has a limit when n tends to ∞, uniformly in k.

Invariant measures and frequencies

Fact

A minimal symbolic system (X, T) is uniquely ergodic if and only if, for all $x \in X$, x has uniform frequencies.

Invariant measures and frequencies

Fact

A minimal symbolic system (X, T) is uniquely ergodic if and only if, for all $x \in X$, x has uniform frequencies.

$\rightsquigarrow$ This is a direct consequence of the Birkhoff ergodic Theorem.

Invariant measures and frequencies

Fact

A minimal symbolic system (X, T) is uniquely ergodic if and only if, for all $x \in X$, x has uniform frequencies.

$\rightsquigarrow$ This is a direct consequence of the Birkhoff ergodic Theorem.

- In this case, for all $x \in X$ the frequency of the factor w is equal to $\mu([w])$, where

$$[w] = \{x \in X : x_0 \cdots x_{|w|-1} = w\}$$

Balance

Proposition

The language $\mathcal{L}(X)$ is balanced in the factor v if and only if v has a frequency μ_v and there exists a constant B_v such that for any factor $w \in \mathcal{L}(X)$, we have

$$||w|_v - \mu_v|w|| \leq B_v$$

Balance

Proposition

The language $\mathcal{L}(X)$ is balanced in the factor v if and only if v has a frequency μ_v and there exists a constant B_v such that for any factor $w \in \mathcal{L}(X)$, we have

$$||w|_v - \mu_v|w|| \leq B_v$$

- Equivalently, v has a frequency μ_v and there exists B_v such that for all $x \in X$ and for all $n \geq 1$,

$$||x_{[0,n)}|_v - \mu_v n| \leq B_v$$

Balance

Proposition

The language $\mathcal{L}(X)$ is balanced in the factor v if and only if v has a frequency μ_v and there exists a constant B_v such that for any factor $w \in \mathcal{L}(X)$, we have

$$||w|_v - \mu_v|w|| \leq B_v$$

- Equivalently, v has a frequency μ_v and there exists B_v such that for all $x \in X$ and for all $n \geq 1$,

$$|\underbrace{|x_{[0,n)}|_v - \mu_v n}_{\text{Discrepancy}}| \leq B_v$$

Coboundaries

Coboundaries

- Let (X, T) a topological dynamical system.

Coboundaries

- Let (X, T) a topological dynamical system.
- Continuous functions: $C(X, \mathbb{R})$, $C(X, \mathbb{Z})$.

Coboundaries

- Let (X, T) a topological dynamical system.
- Continuous functions: $C(X, \mathbb{R})$, $C(X, \mathbb{Z})$.
- The coboundary map $\partial : C(X, \mathbb{R}) \to C(X, \mathbb{R})$ is given by

$$\partial f = f \circ T - f$$

Coboundaries

- Let (X, T) a topological dynamical system.
- Continuous functions: $C(X, \mathbb{R})$, $C(X, \mathbb{Z})$.
- The coboundary map $\partial : C(X, \mathbb{R}) \to C(X, \mathbb{R})$ is given by

$$\partial f = f \circ T - f$$

- Dimension group

$$K^0(X, T) = (C(X, \mathbb{Z})/\partial C(X, \mathbb{Z}), (C(X, \mathbb{Z})/\partial C(X, \mathbb{Z}))^+, \mathbf{1})$$

is a complete invariant of (strong) orbit equivalence.

Gotschalk-Hedlund's Theorem

Theorem (Gotschalk-Hedlund '55)

Let (X, T) *be a minimal topological dynamical system. The map* $f \in C(X, \mathbb{R})$ *is a coboundary if and only if there exists* $x_0 \in X$ *such that the sequence* $(f^{(n)}(x_0))_{n \geq 1}$ *is bounded, where*

$$f^{(n)}(x) = f(x) + f \circ T(x) + \cdots + f \circ T^{n-1}(x).$$

Gotschalk-Hedlund's Theorem

Theorem (Gotschalk-Hedlund '55)

Let (X, T) be a minimal topological dynamical system. The map $f \in C(X, \mathbb{R})$ is a coboundary if and only if there exists $x_0 \in X$ such that the sequence $(f^{(n)}(x_0))_{n\geq 1}$ is bounded, where

$$f^{(n)}(x) = f(x) + f \circ T(x) + \cdots + f \circ T^{n-1}(x).$$

The sequence $(f^{(n)})_{n\geq 1}$ is called the cocycle of f.

Gotschalk-Hedlund's Theorem

Theorem (Gotschalk-Hedlund '55)

Let (X, T) be a minimal topological dynamical system. The map $f \in C(X, \mathbb{R})$ is a coboundary if and only if there exists $x_0 \in X$ such that the sequence $(f^{(n)}(x_0))_{n \geq 1}$ is bounded, where

$$f^{(n)}(x) = f(x) + f \circ T(x) + \cdots + f \circ T^{n-1}(x).$$

The sequence $(f^{(n)})_{n \geq 1}$ is called the cocycle of f.

Remark. If $f = g \circ T - g$, $(f^{(n)}(x))_{n \in \mathbb{N}}$ is bounded for every $x \in X$.

Gotschalk-Hedlund's Theorem

Theorem (Gotschalk-Hedlund '55)

Let (X, T) be a minimal topological dynamical system. The map $f \in C(X, \mathbb{R})$ is a coboundary if and only if there exists $x_0 \in X$ such that the sequence $(f^{(n)}(x_0))_{n \geq 1}$ is bounded, where

$$f^{(n)}(x) = f(x) + f \circ T(x) + \cdots + f \circ T^{n-1}(x).$$

The sequence $(f^{(n)})_{n \geq 1}$ is called the cocycle of f.

Remark. If $f = g \circ T - g$, $(f^{(n)}(x))_{n \in \mathbb{N}}$ is bounded for every $x \in X$.

$\rightsquigarrow$ f coboundary $\Leftrightarrow$ $(f^{(n)})_{n \in \mathbb{N}}$ bounded.

Coboundaries and balance

Coboundaries and balance

Remark

Let (X, T) *a subshift on* $\mathcal{A}$*. Suppose* (X, T) *is minimal and uniquely ergodic with measure* μ*. Given a factor* $v \in \mathcal{L}_(X)$*, define*

$$f_v = \chi_{[v]} - \mu_v \in C(X, \mathbb{R})$$

Then, (X, T) *is balanced in the factor* v *if and only if* f_v *is a coboundary.*

Coboundaries and balance

Remark

Let (X, T) a subshift on $\mathcal{A}$. Suppose (X, T) is minimal and uniquely ergodic with measure μ. Given a factor $v \in \mathcal{L}_(X)$, define

$$f_v = \chi_{[v]} - \mu_v \in C(X, \mathbb{R})$$

Then, (X, T) is balanced in the factor v if and only if f_v is a coboundary.

Proof.

Note that for all $x \in X$, for all $n \geq 1$,

$$\begin{aligned} f_v^{(n)}(x) &= \chi_{[v]}(x) - \mu_v + \cdots + \chi_{[v]}(T^{n-1}x) - \mu_v \\ &= |x_{[0,n+|v|)}|_v - n\mu_v \end{aligned}$$

□

Substitution systems

Substitution systems

Consider a primitive substitution $\sigma : \mathcal{A} \to \mathcal{A}^*$.

The symbolic system generated by σ is (X_σ, T), where

$$X_\sigma = \{x \in \mathcal{A}^{\mathbb{Z}} : \forall w \prec x, \exists a \in \mathcal{A}, \exists n \in \mathbb{N} : w \prec \sigma^n(a)\}.$$

Substitution systems

Consider a primitive substitution $\sigma : \mathcal{A} \to \mathcal{A}^*$.

The symbolic system generated by σ is (X_σ, T), where

$$X_\sigma = \{x \in \mathcal{A}^{\mathbb{Z}} : \forall w \prec x, \exists a \in \mathcal{A}, \exists n \in \mathbb{N} : w \prec \sigma^n(a)\}.$$

Theorem (Queffélec '87)

(X_σ, T) *is minimal and uniquely ergodic.*

Partitions in towers

Partitions in towers

A partition in towers of the symbolic system (X, T) is

$$\mathcal{P} = \{T^j B_i : 1 \leq i \leq m, 0 \leq j < h_i\}$$

where the B_i's are clopen and nonempty, $m, h_i \in \mathbb{N}$.

Partitions in towers

A partition in towers of the symbolic system (X, T) is

$$\mathcal{P} = \{ T^j B_i : 1 \leq i \leq m, 0 \leq j < h_i \}$$

where the B_i's are clopen and nonempty, $m, h_i \in \mathbb{N}$.

Proposition

Let (X_σ, T) be the symbolic system generated by σ. Let $\mathcal{L}_2(X_\sigma) = \mathcal{L}(X_\sigma) \cap \mathcal{A}^2$. For all $n \in \mathbb{N}$, define

$$\mathcal{P}_n = \{ T^j \sigma^n([ab]) : ab \in \mathcal{L}_2(X_\sigma), 0 \leq j < |\sigma^n(a)| \}$$

The sequence $(\mathcal{P}_n)_{n\in\mathbb{N}}$ is a sequence of partitions in towers of (X, T), such that for all $n \in \mathbb{N}$, $\mathcal{P}_{n+1}$ is finer than $\mathcal{P}_n$.

Partitions in towers

Let $\mathcal{L}_n(X_\sigma) = \mathcal{L}(X_\sigma) \cap \mathcal{A}^n$.

Partitions in towers

Let $\mathcal{L}_n(X_\sigma) = \mathcal{L}(X_\sigma) \cap \mathcal{A}^n$.

- For all $k \geq 1$, the sequence

  $$\mathcal{P}_n^k = \{ T^j \sigma^n([a_0 \cdots a_{k-1}]) : a_0 \cdots a_{k-1} \in \mathcal{L}_k(X_\sigma), 0 \leq j < |\sigma^n(a_0)| \}$$

  is also a nested sequence of partitions in towers.

Partitions in towers

Let $\mathcal{L}_n(X_\sigma) = \mathcal{L}(X_\sigma) \cap \mathcal{A}^n$.

- For all $k \geq 1$, the sequence

  $$\mathcal{P}_n^k = \{ T^j \sigma^n([a_0 \cdots a_{k-1}]) : a_0 \cdots a_{k-1} \in \mathcal{L}_k(X_\sigma), 0 \leq j < |\sigma^n(a_0)|\}$$

  is also a nested sequence of partitions in towers.
- **Question:** Why not to use

  $$\mathcal{Q}_n = \{ T^j \sigma^n([a]) : a \in \mathcal{A}, 0 \leq j < |\sigma^n(a)|\} \quad ??$$

Partitions in towers

Let $\mathcal{L}_n(X_\sigma) = \mathcal{L}(X_\sigma) \cap \mathcal{A}^n$.

- For all $k \geq 1$, the sequence

$$\mathcal{P}_n^k = \{ T^j \sigma^n([a_0 \cdots a_{k-1}]) : a_0 \cdots a_{k-1} \in \mathcal{L}_k(X_\sigma), 0 \leq j < |\sigma^n(a_0)| \}$$

  is also a nested sequence of partitions in towers.
- **Question:** Why not to use

$$\mathcal{Q}_n = \{ T^j \sigma^n([a]) : a \in \mathcal{A}, 0 \leq j < |\sigma^n(a)| \} \quad ??$$

- We need the atoms of the partition to determine as many starting letters as we want!

Partitions in towers

Let $\mathcal{L}_n(X_\sigma) = \mathcal{L}(X_\sigma) \cap \mathcal{A}^n$.

- For all $k \geq 1$, the sequence

  $$\mathcal{P}_n^k = \{ T^j \sigma^n([a_0 \cdots a_{k-1}]) : a_0 \cdots a_{k-1} \in \mathcal{L}_k(X_\sigma), 0 \leq j < |\sigma^n(a_0)|\}$$

  is also a nested sequence of partitions in towers.
- **Question:** Why not to use

  $$\mathcal{Q}_n = \{ T^j \sigma^n([a]) : a \in \mathcal{A}, 0 \leq j < |\sigma^n(a)|\} \quad ??$$

- We need the atoms of the partition to determine as many starting letters as we want! $\rightsquigarrow$ cocycle of f_v.

Coboundaries and partitions

Define $R_n(X) = \{\phi : \mathcal{L}_n(X) \to \mathbb{R}\}$ and

$$\beta : \underbrace{R_1(X)}_{\mathbb{R}^{\mathcal{A}}} \to \underbrace{R_2(X)}_{\mathbb{R}^{\mathcal{L}_2(X)}} ; \varphi \longmapsto (\beta\varphi)(ab) = \varphi(b) - \varphi(a) \quad \forall ab \in \mathcal{L}_2(X)$$

Coboundaries and partitions

Define $R_n(X) = \{\phi : \mathcal{L}_n(X) \to \mathbb{R}\}$ and

$$\beta : \underbrace{R_1(X)}_{\mathbb{R}^{\mathcal{A}}} \to \underbrace{R_2(X)}_{\mathbb{R}^{\mathcal{L}_2(X)}}; \varphi \longmapsto (\beta\varphi)(ab) = \varphi(b) - \varphi(a) \quad \forall ab \in \mathcal{L}_2(X)$$

Proposition ($\star$)

Let $f \in C(X_\sigma, \mathbb{Z})$ such that there exists $k \in \mathbb{N}$ for which f is constant in the atoms of $\mathcal{P}_k$. For all $n \geq k$, define $\phi_n \in \mathbb{R}^{\mathcal{L}_2(X_\sigma)}$ by

$$\phi_n(ab) = \sum_{j=0}^{|\sigma^n(a)|-1} f\mid_{T^j\sigma^n([ab])} \quad \forall ab \in \mathcal{L}_2(X_\sigma).$$

If f is a coboundary, then $\phi_n \in \beta(R_1(X_\sigma))$ for all n large enough.

Remarks

Remarks

- The subspace $\beta(R_1(X_\sigma))$ is easy to handle.

Remarks

- The subspace $\beta(R_1(X_\sigma))$ is easy to handle.
- Examples:

Remarks

- The subspace $\beta(R_1(X_\sigma))$ is easy to handle.
- Examples:
  - Thue-Morse substitution on $\{0,1\}$ $\sigma : 0 \mapsto 01, 1 \mapsto 10$

$$\mathcal{L}_2(X_\sigma) = \{00, 01, 10, 11\}$$

$$\beta(R_1(X_\sigma)) = \left\langle \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \right\rangle \leqslant \mathbb{R}^4$$

  - Sturmians,

$$\mathcal{L}_2(X_\sigma) = \{00, 01, 10\}$$

$$\beta(R_1(X_\sigma)) = \left\langle \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\rangle \leqslant \mathbb{R}^3$$

  - We need that $f \in C(X_\sigma, \mathbb{Z})$.

Rational frequencies

Rational frequencies

- For a factor $v \in \mathcal{L}(X_\sigma)$, we want the map

$$f_v = \chi_{[v]} - \mu_v$$

to *be* integer valued.

Rational frequencies

- For a factor $v \in \mathcal{L}(X_\sigma)$, we want the map

$$f_v = \chi_{[v]} - \mu_v$$

  to *be* integer valued.
- Rational frequencies ☺

Rational frequencies

- For a factor $v \in \mathcal{L}(X_\sigma)$, we want the map

$$f_v = \chi_{[v]} - \mu_v$$

to *be* integer valued.

- Rational frequencies ☺
- Irrational frequencies ☹

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ *and suppose that* $\mu_v \in \mathbb{Q}$. *Then, there exists* $k \geq 1$ *such that* f_v *is constant in the atoms of* $\mathcal{P}_k$ *and if* (X_σ, T) *is balanced on* v, ϕ_n *(defined for* f_v*) belongs to* $\beta(R_1(X_\sigma))$ *for all* n *large enough.*

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ and suppose that $\mu_v \in \mathbb{Q}$. Then, there exists $k \geq 1$ such that f_v is constant in the atoms of $\mathcal{P}_k$ and if (X_σ, T) is balanced on v, ϕ_n (defined for f_v) belongs to $\beta(R_1(X_\sigma))$ for all n large enough.

Proof.

- There exists $m \in \mathbb{N}$ such that $mf_v \in C(X, \mathbb{Z})$.

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ and suppose that $\mu_v \in \mathbb{Q}$. Then, there exists $k \geq 1$ such that f_v is constant in the atoms of $\mathcal{P}_k$ and if (X_σ, T) is balanced on v, ϕ_n (defined for f_v) belongs to $\beta(R_1(X_\sigma))$ for all n large enough.

Proof.

- There exists $m \in \mathbb{N}$ such that $mf_v \in C(X, \mathbb{Z})$.
- For all n, the elements in an atom of $\mathcal{P}_n$ share at least their $L_n + 1$ letters, where $L_n = \min\{|\sigma^n(a)| : a \in \mathcal{A}\} \Rightarrow f_v$ constant.

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ and suppose that $\mu_v \in \mathbb{Q}$. Then, there exists $k \geq 1$ such that f_v is constant in the atoms of $\mathcal{P}_k$ and if (X_σ, T) is balanced on v, ϕ_n (defined for f_v) belongs to $\beta(R_1(X_\sigma))$ for all n large enough.

Proof.

- There exists $m \in \mathbb{N}$ such that $mf_v \in C(X, \mathbb{Z})$.
- For all n, the elements in an atom of $\mathcal{P}_n$ share at least their $L_n + 1$ letters, where $L_n = \min\{|\sigma^n(a)| : a \in \mathcal{A}\} \Rightarrow f_v$ constant.
- f_v constant in the atoms of $\mathcal{P}_k \Leftrightarrow mf_v$ constant in the atoms of $\mathcal{P}_k$.

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ and suppose that $\mu_v \in \mathbb{Q}$. Then, there exists $k \geq 1$ such that f_v is constant in the atoms of $\mathcal{P}_k$ and if (X_σ, T) is balanced on v, ϕ_n (defined for f_v) belongs to $\beta(R_1(X_\sigma))$ for all n large enough.

Proof.

- There exists $m \in \mathbb{N}$ such that $mf_v \in C(X, \mathbb{Z})$.
- For all n, the elements in an atom of $\mathcal{P}_n$ share at least their $L_n + 1$ letters, where $L_n = \min\{|\sigma^n(a)| : a \in \mathcal{A}\} \Rightarrow f_v$ constant.
- f_v constant in the atoms of $\mathcal{P}_k \Leftrightarrow mf_v$ constant in the atoms of $\mathcal{P}_k$.
- If (X_σ, T) is balanced on v, f_v is a coboundary.

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ and suppose that $\mu_v \in \mathbb{Q}$. Then, there exists $k \geq 1$ such that f_v is constant in the atoms of $\mathcal{P}_k$ and if (X_σ, T) is balanced on v, ϕ_n (defined for f_v) belongs to $\beta(R_1(X_\sigma))$ for all n large enough.

Proof.

- There exists $m \in \mathbb{N}$ such that $mf_v \in C(X, \mathbb{Z})$.
- For all n, the elements in an atom of $\mathcal{P}_n$ share at least their $L_n + 1$ letters, where $L_n = \min\{|\sigma^n(a)| : a \in \mathcal{A}\} \Rightarrow f_v$ constant.
- f_v constant in the atoms of $\mathcal{P}_k \Leftrightarrow mf_v$ constant in the atoms of $\mathcal{P}_k$.
- If (X_σ, T) is balanced on v, f_v is a coboundary.
- f_v coboundary $\Leftrightarrow mf_v$ coboundary.

Rational frequencies

Proposition

Let $v \in \mathcal{L}(X_\sigma)$ and suppose that $\mu_v \in \mathbb{Q}$. Then, there exists $k \geq 1$ such that f_v is constant in the atoms of $\mathcal{P}_k$ and if (X_σ, T) is balanced on v, ϕ_n (defined for f_v) belongs to $\beta(R_1(X_\sigma))$ for all n large enough.

Proof.

- There exists $m \in \mathbb{N}$ such that $mf_v \in C(X, \mathbb{Z})$.
- For all n, the elements in an atom of $\mathcal{P}_n$ share at least their $L_n + 1$ letters, where $L_n = \min\{|\sigma^n(a)| : a \in \mathcal{A}\} \Rightarrow f_v$ constant.
- f_v constant in the atoms of $\mathcal{P}_k \Leftrightarrow mf_v$ constant in the atoms of $\mathcal{P}_k$.
- If (X_σ, T) is balanced on v, f_v is a coboundary.
- f_v coboundary $\Leftrightarrow mf_v$ coboundary.
- By Proposition $(\star)$, ϕ_n belongs to $\beta(R_1(X_\sigma))$ for all n large enough. □

Example: the Thue-Morse substitution

Example: the Thue-Morse substitution

Corollary

Let σ be the Thue-Morse substitution and let (X, T) the symbolic system generated by σ. For every $\ell \geq 2$, (X, T) is not balanced in the factors of length ℓ.

Example: the Thue-Morse substitution

Corollary

Let σ be the Thue-Morse substitution and let (X, T) the symbolic system generated by σ. For every $\ell \geq 2$, (X, T) is not balanced in the factors of length ℓ.

Proof.

Example: the Thue-Morse substitution

Corollary

Let σ be the Thue-Morse substitution and let (X, T) the symbolic system generated by σ. For every $\ell \geq 2$, (X, T) is not balanced in the factors of length ℓ.

Proof.

- Let v a factor with $|v| = \ell$. The frequency of v verifies (Dekking '92)

$$\mu([v]) = \frac{1}{6}2^{-m} \quad \textit{or} \quad \mu([v]) = \frac{1}{3}2^{-m}$$

where m is such that $2^m < \ell \leq 2^{m+1}$.

Example: the Thue-Morse substitution

Corollary

Let σ be the Thue-Morse substitution and let (X, T) the symbolic system generated by σ. For every $\ell \geq 2$, (X, T) is not balanced in the factors of length ℓ.

Proof.

- Let v a factor with $|v| = \ell$. The frequency of v verifies (Dekking '92)

$$\mu([v]) = \frac{1}{6} 2^{-m} \quad \textit{or} \quad \mu([v]) = \frac{1}{3} 2^{-m}$$

  where m is such that $2^m < \ell \leq 2^{m+1}$.
- Frequencies are then rational, and $p_v = 1$, $q_v \in \{3 \cdot 2^{m+1}, 3 \cdot 2^m\}$.

Example: the Thue-Morse substitution

Corollary

Let σ be the Thue-Morse substitution and let (X, T) the symbolic system generated by σ. For every $\ell \geq 2$, (X, T) is not balanced in the factors of length ℓ.

Proof.

- Let v a factor with $|v| = \ell$. The frequency of v verifies (Dekking '92)

$$\mu([v]) = \frac{1}{6}2^{-m} \quad or \quad \mu([v]) = \frac{1}{3}2^{-m}$$

  where m is such that $2^m < \ell \leq 2^{m+1}$.
- Frequencies are then rational, and $p_v = 1$, $q_v \in \{3 \cdot 2^{m+1}, 3 \cdot 2^m\}$.

Example: the Thue-Morse substitution

- For all $ab \in \mathcal{L}_2(X)$ and for n large enough,

$$\phi_n(ab) = \sum_{j=0}^{|\sigma^n(a)|-1} f_v \mid_{T^j \sigma^n([ab])} .$$

Example: the Thue-Morse substitution

- For all $ab \in \mathcal{L}_2(X)$ and for n large enough,

$$\phi_n(ab) = \sum_{j=0}^{|\sigma^n(a)|-1} f_v \mid_{T^j\sigma^n([ab])} .$$

- $f_v = \chi_{[v]} - \mu_v$, then

$$\phi_n(ab) = \alpha_{ab}\left(1 - \frac{1}{q_v}\right) + (|\sigma^n(a)| - \alpha_{ab}) \cdot -\frac{1}{q_v}.$$

$$\alpha_{ab} = \#\{0 \leq j < |\sigma^n(a)| : T^j\sigma^n([ab]) \subseteq [v]\}.$$

Example: the Thue-Morse substitution

- For all $ab \in \mathcal{L}_2(X)$ and for n large enough,

$$\phi_n(ab) = \sum_{j=0}^{|\sigma^n(a)|-1} f_v \mid_{T^j\sigma^n([ab])} .$$

- $f_v = \chi_{[v]} - \mu_v$, then

$$\phi_n(ab) = \alpha_{ab}\left(1 - \frac{1}{q_v}\right) + (|\sigma^n(a)| - \alpha_{ab}) \cdot -\frac{1}{q_v}.$$

$$\alpha_{ab} = \#\{0 \leq j < |\sigma^n(a)| : T^j\sigma^n([ab]) \subseteq [v]\}.$$

- $\phi_n \in \beta(R_1(X)) = \langle(0,1,-1,0)\rangle \Rightarrow \phi_n(00) = 0$

Example: the Thue-Morse substitution

- For all $ab \in \mathcal{L}_2(X)$ and for n large enough,

$$\phi_n(ab) = \sum_{j=0}^{|\sigma^n(a)|-1} f_v \mid_{T^j\sigma^n([ab])} .$$

- $f_v = \chi_{[v]} - \mu_v$, then

$$\phi_n(ab) = \alpha_{ab}\left(1 - \frac{1}{q_v}\right) + (|\sigma^n(a)| - \alpha_{ab}) \cdot -\frac{1}{q_v}.$$

$$\alpha_{ab} = \#\{0 \leq j < |\sigma^n(a)| : T^j\sigma^n([ab]) \subseteq [v]\}.$$

- $\phi_n \in \beta(R_1(X)) = \langle(0,1,-1,0)\rangle \Rightarrow \phi_n(00) = 0$
- We obtain

$$q_v\alpha_{00} = |\sigma^n(0)| = 2^n \quad \rightarrow\leftarrow$$

Return words

Return words

- Let σ be a primitive substitution on $\mathcal{A}$, $a \in \mathcal{A}$. A return word to a in $\mathcal{L}(X_\sigma)$ is a finite word $w \in \mathcal{L}(X_\sigma)$ such that $w_0 = a$ and $wa \in \mathcal{L}(X_\sigma)$.

Return words

- Let σ be a primitive substitution on $\mathcal{A}$, $a \in \mathcal{A}$. A return word to a in $\mathcal{L}(X_\sigma)$ is a finite word $w \in \mathcal{L}(X_\sigma)$ such that $w_0 = a$ and $wa \in \mathcal{L}(X_\sigma)$.

$$\cdots \overbrace{abcbbc}^{w}\underbrace{\,abc}_{}\,a \cdots$$

$$\cdots \underbrace{abcbbc\,abc}_{w'}\,a \cdots$$

Return words

- Let σ be a primitive substitution on $\mathcal{A}$, $a \in \mathcal{A}$. A return word to a in $\mathcal{L}(X_\sigma)$ is a finite word $w \in \mathcal{L}(X_\sigma)$ such that $w_0 = a$ and $wa \in \mathcal{L}(X_\sigma)$.

$$\cdots \underbrace{\overbrace{abcbbc}^{w}\,abc}_{w'}\,a \cdots$$

- Since (X, T) is minimal, return words always exists for every letter.

Return words

- Let σ be a primitive substitution on $\mathcal{A}$, $a \in \mathcal{A}$. A return word to a in $\mathcal{L}(X_\sigma)$ is a finite word $w \in \mathcal{L}(X_\sigma)$ such that $w_0 = a$ and $wa \in \mathcal{L}(X_\sigma)$.

$$\cdots \underbrace{\overbrace{abcbbc}^{w}\, abc}_{w'}\, a \cdots$$

- Since (X, T) is minimal, return words always exists for every letter.
- If $w = w_0 \cdots w_{|w|-1}$ is a return word to a,

$$w_0 w_1, w_1 w_2, \cdots, w_{|w|-2} w_{|w|-1}, w_{|w|-1} a \in \mathcal{L}_2(X_\sigma),$$

and if $\phi \in \beta(R_1(X))$,

$$\phi(w_{|w|-1} a) + \sum_{i=1}^{|w|-1} \phi(w_{i-1} w_i) = 0$$

Return words

- Let σ be a primitive substitution on $\mathcal{A}$, $a \in \mathcal{A}$. A return word to a in $\mathcal{L}(X_\sigma)$ is a finite word $w \in \mathcal{L}(X_\sigma)$ such that $w_0 = a$ and $wa \in \mathcal{L}(X_\sigma)$.

$$\cdots \underbrace{\overbrace{abcbbc}^{w}\,abc}_{w'}\,a \cdots$$

- Since (X, T) is minimal, return words always exists for every letter.
- If $w = w_0 \cdots w_{|w|-1}$ is a return word to a,

$$w_0 w_1, w_1 w_2, \cdots, w_{|w|-2} w_{|w|-1}, w_{|w|-1} a \in \mathcal{L}_2(X_\sigma),$$

and if $\phi \in \beta(R_1(X))$,

$$\underbrace{\phi(w_{|w|-1} a)}_{\varphi(w_{|w|-1}) - \varphi(a)} + \sum_{i=1}^{|w|-1} \underbrace{\phi(w_{i-1} w_i)}_{\varphi(w_i) - \varphi(w_{i-1})} = 0$$

Theorem

Let $v \in \mathcal{L}(X_\sigma)$ *such that* $\mu_v \in \mathbb{Q}$. *Write* $\mu_v = p_v/q_v$ *with* $(p_v, q_v) = 1$. *There exists* $N \geq 1$ *such that for all* $a \in \mathcal{A}$ *and for all return word* w *to* a, q_v *divides* $|\sigma^n(w)|$ *for all* $n \geq N$. *In particular, if* $aa \in \mathcal{L}_2(X)$, *then* q_v *divides* $|\sigma^n(a)|$ *for all* $n \geq N$.

Theorem

Let $v \in \mathcal{L}(X_\sigma)$ such that $\mu_v \in \mathbb{Q}$. Write $\mu_v = p_v/q_v$ with $(p_v, q_v) = 1$. There exists $N \geq 1$ such that for all $a \in \mathcal{A}$ and for all return word w to a, q_v divides $|\sigma^n(w)|$ for all $n \geq N$. In particular, if $aa \in \mathcal{L}_2(X)$, then q_v divides $|\sigma^n(a)|$ for all $n \geq N$.

Proof.

Theorem

Let $v \in \mathcal{L}(X_\sigma)$ such that $\mu_v \in \mathbb{Q}$. Write $\mu_v = p_v/q_v$ with $(p_v, q_v) = 1$. There exists $N \geq 1$ such that for all $a \in \mathcal{A}$ and for all return word w to a, q_v divides $|\sigma^n(w)|$ for all $n \geq N$. In particular, if $aa \in \mathcal{L}_2(X)$, then q_v divides $|\sigma^n(a)|$ for all $n \geq N$.

Proof.

- $\phi_n \in \beta(R_1(X))$ for n large enough.

Theorem

Let $v \in \mathcal{L}(X_\sigma)$ such that $\mu_v \in \mathbb{Q}$. Write $\mu_v = p_v/q_v$ with $(p_v, q_v) = 1$. There exists $N \geq 1$ such that for all $a \in \mathcal{A}$ and for all return word w to a, q_v divides $|\sigma^n(w)|$ for all $n \geq N$. In particular, if $aa \in \mathcal{L}_2(X)$, then q_v divides $|\sigma^n(a)|$ for all $n \geq N$.

Proof.

- $\phi_n \in \beta(R_1(X))$ for n large enough.
- $\phi_n(w_{|w|-1}a) + \sum_{i=1}^{|w|-1} \phi_n(w_{i-1}w_i) = 0$

Theorem

Let $v \in \mathcal{L}(X_\sigma)$ such that $\mu_v \in \mathbb{Q}$. Write $\mu_v = p_v/q_v$ with $(p_v, q_v) = 1$. There exists $N \geq 1$ such that for all $a \in \mathcal{A}$ and for all return word w to a, q_v divides $|\sigma^n(w)|$ for all $n \geq N$. In particular, if $aa \in \mathcal{L}_2(X)$, then q_v divides $|\sigma^n(a)|$ for all $n \geq N$.

Proof.

- $\phi_n \in \beta(R_1(X))$ for n large enough.
- $\phi_n(w_{|w|-1}a) + \sum_{i=1}^{|w|-1} \phi_n(w_{i-1}w_i) = 0$
- For all $ab \in \mathcal{L}_2(X)$ and for n large enough,

$$\phi_n(ab) = \alpha_{ab}\left(1 - \frac{p_v}{q_v}\right) + (|\sigma^n(a)| - \alpha_{ab}) \cdot -\frac{p_v}{q_v},$$

where $\alpha_{ab} = \#\{0 \leq j < |\sigma^n(a)| : T^j\sigma^n([ab]) \subseteq [v]\}$.

- We obtain

$$0 = \alpha_{w_{|w|-1}a}(q_v - p_v) - (|\sigma^n(w_{|w|-1})| - \alpha_{w_{|w|-1}a}) \cdot p_v + \sum_{i=1}^{|w|-1} \alpha_{w_{i-1}w_i}(q_v - p_v) - (|\sigma^n(w_{i-1})| - \alpha_{w_{i-1}w_i}) \cdot p_v$$

- which implies

$$\begin{aligned} q_v \left(\alpha_{w_{|w|-1}a} + \sum_{i=1}^{|w|-1} \alpha_{w_{i-1}w_i} \right) &= p_v \left(|\sigma^n(w_{|w|-1})| + \sum_{i=1}^{|w|-1} |\sigma^n(w_{i-1})| \right) \\ &= p_v |\sigma^n(w)| \end{aligned}$$

□

Example: the Thue-Morse substitution again

- Suppose (X, T) is balanced on v and $q_v = 3 \cdot 2^{m+1}$.

Example: the Thue-Morse substitution again

- Suppose (X, T) is balanced on v and $q_v = 3 \cdot 2^{m+1}$.
- For all $a \in \mathcal{A} = \{0, 1\}, aa \in \mathcal{L}_2(X)$.

Example: the Thue-Morse substitution again

- Suppose (X, T) is balanced on v and $q_v = 3 \cdot 2^{m+1}$.
- For all $a \in \mathcal{A} = \{0, 1\}, aa \in \mathcal{L}_2(X)$.
- Then, q_v divides $|\sigma^n(a)|$ for every n large enough.

Example: the Thue-Morse substitution again

- Suppose (X, T) is balanced on v and $q_v = 3 \cdot 2^{m+1}$.
- For all $a \in \mathcal{A} = \{0, 1\}, aa \in \mathcal{L}_2(X)$.
- Then, q_v divides $|\sigma^n(a)|$ for every n large enough.
- But $q_v = 3 \cdot 2^{m+1}$ and $|\sigma^n(a)| = 2^n, \rightarrow\leftarrow$.

Another examples

- We consider constant length substitutions $\rightsquigarrow$ rational frequencies.

Another examples

- We consider constant length substitutions $\rightsquigarrow$ rational frequencies.
- We focus on letter frequencies:

Incidence matrix $\rightsquigarrow$ Perron-Frobenuis right eigenvalue
$\rightsquigarrow$ normalized right eigenvector $\rightsquigarrow$ frequencies.

Another examples

- We consider constant length substitutions $\rightsquigarrow$ rational frequencies.
- We focus on letter frequencies:

  Incidence matrix $\rightsquigarrow$ Perron-Frobenuis right eigenvalue
  $\rightsquigarrow$ normalized right eigenvector $\rightsquigarrow$ frequencies.

- Example: $\sigma : 0 \to 001, 1 \to 101$.

$$M_\sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad E_\sigma = \{1,3\} \quad f_0 = f_1 = 1/2 \quad q_0 = q_1 = 2,$$

$00 \in \mathcal{L}_2(X_\sigma) \Rightarrow 2$ divides 3^n for n large enough $\to\leftarrow$.

Another examples

- $\sigma : \{0,1,2,3\} \to \{0,1,2,3\}^*$ given by

$$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 23, 3 \mapsto 20$$

$$M_\sigma = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad E_\sigma = \{0, -1, 1, 2\}$$

$$f_0 = f_2 = 1/3 \quad f_1 = f_3 = 1/6 \quad q_i \in \{3, 6\}$$

Another examples

- $\sigma : \{0, 1, 2, 3\} \to \{0, 1, 2, 3\}^*$ given by

$$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 23, 3 \mapsto 20$$

$$M_\sigma = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad E_\sigma = \{0, -1, 1, 2\}$$

$$f_0 = f_2 = 1/3 \quad f_1 = f_3 = 1/6 \quad q_i \in \{3, 6\}$$

- There are no *aa*'s.

Another examples

- $\sigma : \{0,1,2,3\} \to \{0,1,2,3\}^*$ given by

$$0 \mapsto 01, 1 \mapsto 02, 2 \mapsto 23, 3 \mapsto 20$$

$$M_\sigma = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad E_\sigma = \{0, -1, 1, 2\}$$

$$f_0 = f_2 = 1/3 \quad f_1 = f_3 = 1/6 \quad q_i \in \{3, 6\}$$

- There are no *aa*'s.
- But 01 is a return word to 0. If (X, T) is balanced on i,

$$\Rightarrow q_i \quad \textit{divides} \quad |\sigma^n(01)| = 2^{n+1} \quad n \ggg$$

$$\to\leftarrow$$

.

Another examples

- In $\sigma : 0 \to 001, 1 \to 101$, M_σ is symmetric and the length of the substitution is coprime with the size of the alphabet.

Another examples

- In $\sigma : 0 \to 001, 1 \to 101$, M_σ is symmetric and the length of the substitution is coprime with the size of the alphabet.
- We can generalize this:

Corollary

If a primitive substitution of length ℓ on an alphabet of size d is such that $(\ell, d) = 1$ and its incidence matrix is symmetric, then (X, T) is not balanced on the letters.

Another examples

- In $\sigma : 0 \to 001, 1 \to 101$, M_σ is symmetric and the length of the substitution is coprime with the size of the alphabet.
- We can generalize this:

Corollary

If a primitive substitution of length ℓ on an alphabet of size d is such that $(\ell, d) = 1$ and its incidence matrix is symmetric, then (X, T) is not balanced on the letters.

Proof.

Another examples

- In $\sigma : 0 \to 001, 1 \to 101$, M_σ is symmetric and the length of the substitution is coprime with the size of the alphabet.
- We can generalize this:

Corollary

If a primitive substitution of length ℓ on an alphabet of size d is such that $(\ell, d) = 1$ and its incidence matrix is symmetric, then (X, T) is not balanced on the letters.

Proof.

- Since the sum of every column of M_σ equals ℓ, $(1, \cdots, 1) \in \mathbb{R}^d$ is a left eigenvector associated to the eigenvalue ℓ.

Another examples

- In $\sigma : 0 \to 001, 1 \to 101$, M_σ is symmetric and the length of the substitution is coprime with the size of the alphabet.
- We can generalize this:

Corollary

If a primitive substitution of length ℓ on an alphabet of size d is such that $(\ell, d) = 1$ and its incidence matrix is symmetric, then (X, T) is not balanced on the letters.

Proof.

- Since the sum of every column of M_σ equals ℓ, $(1, \cdots, 1) \in \mathbb{R}^d$ is a left eigenvector associated to the eigenvalue ℓ.
- Since M_σ is symmetric,

$$M_\sigma \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \ell \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \Rightarrow f_i = 1/d \quad \square$$