An automaton to explore imbalances in S-adic systems

Mélodie Andrieu

[PhD thesis version, March 2021]

Abstract

We describe a semi-algorithm consisting of an automaton, or rather, an ever-building family of automata, whose states contain all the possible imbalances of S-adic words, where S is a finite set of substitutions. The implementation of this "infinite automaton" led to the construction of words with infinite imbalance for the Cassaigne-Selmer multidimensional continued fraction algorithm, as well as Arnoux-Rauzy words whose Rauzy fractal is unbounded in all directions of the plane.

Contents

1 Motivations

A S-adic word is an infinite word which can be written as the limit of the iterated images of a letter by an infinite composition of substitutions (not necessarily the same). S-adic systems are thus a large generalization of substitutive systems. They emerged from the study of the symbolic dynamics of continued fraction algorithms (see for instance [Fog02], [Ber11] or [BD14]).

The imbalance is a combinatorial quantity which measures inequities in the distribution of letters in an infinite word. This quantity is linked to the quality of the convergence of the frequency of letters in growing prefixes of the word, so, in the context of multidimensional continued fraction algorithms, to the quality of the convergence of the approximations proposed by the algorithm.

To our knowledge, two distinct behaviors have been observed.

On one hand, the regular continued fraction algorithm is associated with the S-adic system $X_{S_{st}}$ of Example 5. Every infinite word in X_{st} has its imbalance bounded above by 1.

On the other hand, the S-adic systems associated with Arnoux-Rauzy and Brun (2-dimensional) continued fraction algorithms contain words with infinite imbalance (see respectively [CFZ00] and [DHS13]). The existence of such words was unexpected; in fact, they have measure zero ([DHS13]). The construction of these exceptional words relied on thorough observations and an outstanding intuition.

The following questions have not been answered yet.

1. Does there exist a S-adic system that satisfies the 'intermediate situation', i.e. such that (i) all S-adic words have a finite imbalance; (ii) there is no upper bound for these imbalances.

2. Does there exist a multidimensional continued fraction algorithm which generate words with bounded, or at least finite, imbalances?

This document introduces a semi-algorithm that would help in developing intuitions for a wide range of S-adic systems. Roughly speaking, this semi-algorithm consists in tracking backwards, in an economical way, the common desubstitution history of pairs of factors. This tool is already behind the construction of a C-adic word with infinite imbalance [And18] and behind the construction of a Rauzy fractal unbounded in all directions of the plane [And21]. Most probably, it will be useful to address other questions, such that Questions 1 and 2 above.

2 Preliminaries

2.1 General definitions

An *alphabet* A is a finite set, whose elements are called *letters*. We will work with $(A^* := \cup_{n\in\mathbb{N}} A^n, \cdot)$ the free monoid over A for the concatenation operation. Its elements are called *finite words*; its neutral element ϵ is *the empty word*. A *language* L is a subset of A^*. An *infinite word* over A is an element of $A^{\mathbb{N}}$. The *length* of a finite word u, denoted by $|u|$, is the total number of letters it is written with. A finite word $u = u[0]u[1]...u[n-1]$, where $u[k]$ denotes the $(k+1)$-th letter of u, is a *factor* of a (finite or infinite) word w if there exists a nonnegative integer i such that for all $k \in \{0, ..., n-1\}$, $w[i+k] = u[k]$. In the particular case $i = 0$, we say that u is the *prefix of length* n of w, and denote it by $u = p_n(w)$. Symmetrically, a (finite or infinite) word w' is a *suffix* of w if there exists $l \in \mathbb{N}$ such that $w = p_l(w) \cdot w'$. We immediately see that w' is finite if and only if w is finite; in that case, if n denotes the length of w', we say that w' is the suffix of length n of w, that we denote by $w' = s_n(w)$. At last, we denote by $\mathcal{F}_n(w)$ the set of factors of w of length n and by $\mathcal{F}(w)$ the *language of* w, which is the set of all its factors.

The *abelianized vector* (sometimes called *Parikh vector*) of a finite word $u \in A^*$ is the line vector $\mathrm{ab}(u) := (|u|_a)_{a\in A}$ which counts the number of occurrences of each letter in u. We immediately see that the sum of its coordinates is equal to $|u|$, the length of the word u. If two words u and v have the same length, we call *imbalance vector of* u *and* v the difference of their abelianized vectors:

$\mathrm{imb}(u,v) = \mathrm{ab}(u) - ab(v)$. The sum of its coordinates is of course equal to zero. The *imbalance* of a (finite or infinite) word w is the quantity, possibly infinite:

$$\begin{aligned}\mathrm{imb}(w) &= \sup_{n\in\mathbb{N}} \sup_{u,v\in\mathcal{F}_n(w)} ||\mathrm{imb}(u,v)||_\infty \\ &= \sup_{n\in\mathbb{N}} \sup_{u,v\in\mathcal{F}_n(w)} \max_{a\in A} ||u|_a - |v|_a|.\end{aligned}$$

The imbalance measures inequities in the distribution of letters in w.

A *substitution* is an application mapping letters to finite words: $A \mapsto A^*$, that we extend into a morphism on the free monoid A^* on one hand, and on the set of infinite words $A^{\mathbb{N}}$ on the other hand. For instance, consider the Thue-Morse substitution:

$$\begin{aligned}\sigma_{TM}: \ & \{a,b\} \to \{a,b\}^* \\ & a \mapsto ab \\ & b \mapsto ba.\end{aligned}$$

A substitution is *nonerasing* if no image of letter is the empty word. At last, the *incidence matrix* of a substitution σ defined over A is:

$$M_\sigma := (|\sigma(i)|_j)_{i,j\in A}.$$

Incidence matrices and abelianized vectors are made to satisfy $\mathrm{ab}(\sigma(u)) = \mathrm{ab}(u)M_\sigma$, for any finite word u.

In this document, we are interested in the imbalance of words associated with a set of substitutions.

2.2 S-adic systems

First of all, we endow the set of infinite words $A^{\mathbb{N}}$ with the distance δ, which makes it compact: for all $w, w' \in A^{\mathbb{N}}$, $\delta(w,w') = 2^{-n_0}$, where $n_0 = \min\{n \in \mathbb{N} | w[n] \neq w'[n]\}$ if $w \neq w'$, and $\delta(w,w') = 0$ otherwise. We say that a sequence of finite words $(u_n)_{n\in\mathbb{N}} \in (A^*)^{\mathbb{N}}$ *converges* to an infinite word $w \in A^{\mathbb{N}}$ if for any sequence of infinite words $(v_n)_{n\in\mathbb{N}} \in (A^{\mathbb{N}})^{\mathbb{N}}$, the sequence of infinite words $(u_n \cdot v_n)_{n\in\mathbb{N}} \in (A^{\mathbb{N}})^{\mathbb{N}}$ converges to w.

Let S be a finite set of substitutions, defined over a common alphabet A. An infinite word $w \in A^{\mathbb{N}}$ is *S-adic* if there exist a *directive sequence* $(d_n) \in S^{\mathbb{N}}$, together with a *seed* $a \in A$ such that the sequence of finite words $(d_0 \circ ... \circ d_{n-1}(a))_{n\in\mathbb{N}}$ converges to w.

Example 1. *The set $S = \{\sigma\}$ where σ is the substitution defined over $\{1,2\}$ by $\sigma(1) = 2$, $\sigma(2) = 1$, generates no S-adic word.*

Example 2. *The set $S = \{\sigma_{TM}\}$, where σ_{TM} is the Thue-Morse substitution defined in Subsection 2.1, generates two S-adic words:*

$$w_a = \lim_n (\sigma_{TM})^n(a) = abbabaabbaababbabaababbaabbabaabbaababbaabbabaababbab...$$

$$w_b = \lim_n (\sigma_{TM})^n(b) = baababbaabbabaababbabaabbaababbaabbabaabbaababbabaaba...$$

Example 3. *The set $S = \{\sigma_{fib}\}$, where σ_{fib} is the substitution defined by $\sigma_{fib}(1) = 12$ and $\sigma_{fib}(2) = 1$ generates a unique S-adic word (called* Fibonacci word*):*

$$w_{fib} = \lim_n (\sigma_{fib})^n(1) = \lim_n (\sigma_{fib})^n(2) = 1211212112112121121211211212112112...$$

When the set S contains a unique substitution, S-adic words are said *substitutive*. Substitutive words have been intensively studied (see for instance [Fog02]).

Example - definition 4. *Let $C := \{c_1, c_2\}$ the set of Cassaigne-Selmer substitutions, defined over $\{1,2,3\}$ by:*

$$c_1: \begin{array}{ll} 1 & \mapsto 1 \\ 2 & \mapsto 13 \\ 3 & \mapsto 2, \end{array} \qquad \text{and} \qquad c_2: \begin{array}{ll} 1 & \mapsto 2 \\ 2 & \mapsto 13 \\ 3 & \mapsto 3. \end{array}$$

By definition, C-adic words are exactly S-adic words for the specific set of substitutions C.

Example - definition 5. *Let $d \in \mathbb{N}$ such that $d \geq 2$. We consider the set of d substitutions $S_d = \{\sigma_{d,1}, ..., \sigma_{d,d}\}$ defined over the alphabet $A_d := \{1, ..., d\}$ by:*

$$\sigma_{d,i}: \begin{array}{lll} i & \mapsto & i \\ j & \mapsto & ij \quad \text{if } j \neq i, \end{array}$$

for all $i \in \{1, ..., d\}$.

An infinite word over A_d is episturmian *if it has the same language than a S-adic word, for $S = S_d$. It is furthermore* strictly *episturmian if it has the same langage than a S-adic word whose directive sequence contains infinitely many occurrences of each substitutions in S_d.* Sturmian *words are exactly strict episturmian words for $d = 2$;* Arnoux-Rauzy *words are exactly strict episturmian words for $d = 3$. The Fibonacci word (see Example 3) is Sturmian: indeed, it is the limit of the sequence $((\sigma_{2,1} \circ \sigma_{2,2})^n(1))_{n\in\mathbb{N}}$.*

From these five examples, we deduce that:

1. Not all sets of substitutions generate S-adic words.
2. Given a directive sequence $(d_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$ and a seed $a \in A$, the sequence of finite words $(d_0 \circ ... \circ d_{n-1}(a))_{n\in\mathbb{N}}$ may not even have an adherence value in $A^{\mathbb{N}}$.
3. In general, S-adic words are not substitutive. Indeed, there are only a countable number of substitutive words, whereas the set of Sturmian words, which is a subset of the class of all S-adic words, is already uncountable (the slope function maps them onto the irrational numbers).
4. Given a S-adic word, the directive sequence and the seed are not *a priori* unique. Nonetheless, for some sets of substitutions (e.g. standard episturmian words or C-adic words), it is possible to find sufficient conditions on $(d_n) \in S^{\mathbb{N}}$ and $a \in A$ to ensure both the convergence of the sequence $(d_0 \circ ... \circ d_{n-1}(a))_n$ to an infinite word w, and the uniqueness of its directive sequence (see for instance [AR91], [AS13] for Arnoux-words, [GLR09] for episturmian words in general, and [CLL17] for C-adic words).

Let T denote the *shift map*, which erases the first letter of infinite words: if $w = w[0]w[1]w[2]... \in A^{\mathbb{N}}$, then $T(w) = w[1]w[2]w[3]....$ The *S-adic system* associated with a set of substitutions S is the minimal close set containing all S-adic words and stable under the action of T:

$$X_S = \overline{\{T^n(w) \mid \text{ for } n \in \mathbb{N} \text{ and } w \, S\text{-adic word}\}}.$$

Lemma 6 (immediate)**.** *Let S be a set of substitutions. The language of X_S (i.e. the set of factors of words in X_S) is exactly the language of the set of all S-adic words.*

3 Main result

3.1 Teaser

Given a set S of substitutions, we want to answer the questions:

1. Are the imbalances of words in X_S bounded?
2. If they are, give an upper bound.
3. If they are not, for an arbitrary $d \in \mathbb{N}$, exhibit a word in X_S whose imbalance is greater than d.

Theorem A. *Let S denote a finite set of nonerasing substitutions over a common alphabet A, and assume that all letters in A belong to the language of X_S. If D_S denotes the quantity (possibly infinite):*

$$D_S = \sup_{w \in X_S} \operatorname{imb}(w),$$

then a breadth first search in the automaton of imbalances, from its initial states, outputs, for any $d \leq D_S$, a finite sequence of substitutions $(\sigma_i)_{i \in \{1,\dots,n\}}$ in S such that the imbalance of any word in X_S, whose directive sequence starts with $(\sigma_i)_{i \in \{1,\dots,n\}}$, is larger than d.

Remark 7. *We will see (Proposition 13) that the automaton of imbalances has:*

- *an infinite number of states;*
- *a bounded number of transitions from any state;*
- *a finite number of initial states, which only depends on the alphabet A;*

so its breadth first traversal is possible (but would, of course, never end).

Corollary 8 (immediate)**.** *Let $d \in \mathbb{N}$. The question "does there exists a word in X_S with imbalance greater than d?" is semi-decidable.*

3.2 Description of the automaton of imbalances

This subsection is devoted to the harsh (but complete) description of the automaton of imbalances. It is written for readers willing to program it. To understand its construction, it may be preferable to read Section 4 or to consider the Thue-Morse example, in Subsection 3.4.

Definition 9. *Let S be a finite set of nonerasing substitutions defined over a common alphabet A. Let $(l_1, l_2, l_3, l_4) \in (A \cup \{\epsilon\})^4$ and $(x_a)_{a \in A} \in \mathbb{Z}^A$; $(\delta_1, \delta_2, \delta_3, \delta_4) \in \mathbb{N}^4$ and $\sigma \in S$. We say that ${}_{\delta_3}^{\delta_1}\sigma_{\delta_4}^{\delta_2}$ is a* substitute and cut operation *allowed on $X := \left(\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix}, (x_a)_{a \in A}\right)$ if it satisfies the two conditions:*

1. $\begin{cases} (\delta_1, \delta_2) = (0,0) & \text{if } (l_1, l_2) = (\epsilon, \epsilon), \\ \delta_1 < |\sigma(l_1)|,\ \delta_2 < |\sigma(l_1)| \text{ and } \delta_1 + \delta_2 < |\sigma(l_1)| & \text{if } (l_1, l_2) \in A \times \{\epsilon\}, \\ \delta_1 < |\sigma(l_1)| \text{ and } \delta_2 < |\sigma(l_2)| & \text{if } (l_1, l_2) \in A^2; \end{cases}$
2. $\begin{cases} (\delta_3, \delta_4) = (0,0) & \text{if } (l_3, l_4) = (\epsilon, \epsilon), \\ \delta_3 < |\sigma(l_3)|,\ \delta_4 < |\sigma(l_3)| \text{ and } \delta_3 + \delta_4 < |\sigma(l_3)| & \text{if } (l_3, l_4) \in A \times \{\epsilon\}, \\ \delta_3 < |\sigma(l_3)| \text{ and } \delta_4 < |\sigma(l_4)| & \text{if } (l_3, l_4) \in A^2. \end{cases}$

In this case, the image of X by the substitute and cut operation ${}^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}$ is well defined and equals:

$$\left(\begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}, (y_a)_{a\in A}\right),$$

where m_1, m_2, m_3, m_4 and $(y_a)_{a\in A}$ are given by:

- $m_1 = \sigma(l_1)[\delta_1]$;
- $\begin{cases} m_2 = \sigma(l_1)[-\delta_2 - 1] & \textit{if } l_2 = \epsilon, \\ m_2 = \sigma(l_2)[-\delta_2 - 1] & \textit{otherwise;} \end{cases}$
- $m_3 = \sigma(l_3)[\delta_3]$;
- $\begin{cases} m_4 = \sigma(l_3)[-\delta_4 - 1] & \textit{if } l_4 = \epsilon, \\ m_4 = \sigma(l_4)[-\delta_4 - 1] & \textit{otherwise;} \end{cases}$

where, following `Python`*, $u[k]$ denotes the $(k+1)$-th letter of u and $u[-k]$ its k-th letter, reading backwards;*

- $\begin{cases} (y_a) = (x_a)M_\sigma - \mathrm{ab}(p_{\delta_1}(\sigma(l_1))) - \mathrm{ab}(s_{\delta_2}(\sigma(l_1))) + \mathrm{ab}(p_{\delta_3}(\sigma(l_3))) + \mathrm{ab}(s_{\delta_4}(\sigma(l_3))) & \textit{if } (l_2, l_4) = (\epsilon, \epsilon) \\ (y_a) = (x_a)M_\sigma - \mathrm{ab}(p_{\delta_1}(\sigma(l_1))) - \mathrm{ab}(s_{\delta_2}(\sigma(l_1))) + \mathrm{ab}(p_{\delta_3}(\sigma(l_3))) + \mathrm{ab}(s_{\delta_4}(\sigma(l_4))) & \textit{if } (l_2, l_4) \in \{\epsilon\} \times A \\ (y_a) = (x_a)M_\sigma - \mathrm{ab}(p_{\delta_1}(\sigma(l_1))) - \mathrm{ab}(s_{\delta_2}(\sigma(l_2))) + \mathrm{ab}(p_{\delta_3}(\sigma(l_3))) + \mathrm{ab}(s_{\delta_4}(\sigma(l_3))) & \textit{if } (l_2, l_4) \in A \times \{\epsilon\} \\ (y_a) = (x_a)M_\sigma - \mathrm{ab}(p_{\delta_1}(\sigma(l_1))) - \mathrm{ab}(s_{\delta_2}(\sigma(l_2))) + \mathrm{ab}(p_{\delta_3}(\sigma(l_3))) + \mathrm{ab}(s_{\delta_4}(\sigma(l_4))) & \textit{otherwise.} \end{cases}$

In the last point, (x_a) and (y_a) are line vectors indexed by A, and M_σ denotes the incidence matrix of the substitution σ, indexed by $A \times A$.

Notation 10. *Let α and ω be the functions defined by:*

$$\begin{array}{rccc} \alpha : & A^* & \to & A \cup \{\epsilon\} \\ & u & \mapsto & \begin{cases} u[0] \textit{ if } |u| \geq 1 \\ \epsilon \textit{ otherwise;} \end{cases} \end{array} \qquad \begin{array}{rccc} \omega : & A^* & \to & A \cup \{\epsilon\} \\ & u & \mapsto & \begin{cases} u[-1] \textit{ if } |u| \geq 2 \\ \epsilon \textit{ otherwise.} \end{cases} \end{array}$$

They are constructed so as to output the extremal letters of a word (when this makes sense), but never point to the same letter. For instance, $\alpha(examples) = e$, $\omega(examples) = s$, $\alpha(a) = a$, $\omega(a) = \epsilon$, $\alpha(\epsilon) = \omega(\epsilon) = \epsilon$.

Definition 11. *Let S be a finite set of nonerasing substitutions defined over a common alphabet A. Assume that each letter in A appears in a S-adic word (not necessarily the same for all letters). The* automaton of imbalances *of the S-adic system X_S is the infinite oriented graph $\mathtt{G}$ whose vertices are:*

$$\mathtt{V} = \bigcup_{w \in X_S} \left\{ \left(\begin{pmatrix} \alpha(u) & \omega(u) \\ \alpha(v) & \omega(v) \end{pmatrix}, \mathrm{ab}(u) - \mathrm{ab}(v) \right) \mid u, v \in \mathcal{F}(w) \right\},$$

and whose edges map vertices to each of their images by all allowed substitute and cut operations ${}^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}$, with $\sigma \in S$.

A vertex $X = \left(\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix}, (x_a)_{a\in A}\right)$ is a final state *if $\sum_{a \in A} x_a = 0$. The set of final states of $\mathtt{G}$ is denoted by $\mathtt{F}$.*

The set of initial states *of $\mathtt{G}$ is:*

$$\mathtt{I} := \mathtt{V} \cap \left\{ \{((\begin{smallmatrix} \epsilon & \epsilon \\ \epsilon & \epsilon \end{smallmatrix}), (0)_{a\in A})\} \cup \bigcup_{a \in A} \{((\begin{smallmatrix} a & \epsilon \\ \epsilon & \epsilon \end{smallmatrix}), \mathrm{ab}(a)); ((\begin{smallmatrix} \epsilon & \epsilon \\ a & \epsilon \end{smallmatrix}), -\mathrm{ab}(a)); ((\begin{smallmatrix} a & \epsilon \\ a & \epsilon \end{smallmatrix}), (0)_{a\in A})\} \right\}.$$

Remark 12. *Contrary to what its name suggests, the automaton of imbalances is not an automaton. Indeed, it has a infinite number of vertices (see Proposition 13 hereafter). Nonetheless, since we are interested in the labeling of the paths between a particular finite class of vertices (the "initial states") and another subclass -possibly infinite- (the "final states"), we choose to keep the name "automaton".*

3.3 Properties and consequences

Proposition 13. *Let S be a finite set of nonerasing substitutions over a common alphabet A. Denote by $\mathtt{G}$ the automaton of imbalances of X_S. If X_S is empty, so is $\mathtt{G}$. Otherwise:*

1. *there are infinitely many vertices in $\mathtt{G}$;*
2. *there exists $N \in \mathbb{N}$ such that all vertices have less than N outgoing edges;*
3. *there are exactly $3\mathrm{card}(A)+1$ initial states.*

Proof. (1) Assume that X_S is nonempty and pick $w \in X_S$. Then for all integers $n \geq 2$,

$$X_n := \left(\begin{pmatrix} w[0] & w[n-1] \\ \epsilon & \epsilon \end{pmatrix}, \mathrm{ab}(p_n(w)) \right),$$

is a vertex of $\mathtt{G}$, corresponding to the pair of factors of w: $(p_n(w), \epsilon)$. Furthermore, since the sum of the coordinates of the vector $\mathrm{ab}(p_n(w))$ is equal to n (the length of $p_n(w)$), the vertices X_n, for n in $\mathbb{N}$, are pairwise distinct. Therefore, $\mathtt{G}$ contains an infinite number of vertices.
(2) Let $l = \max\{|\sigma(a)| \text{ for } a \in A, \sigma \in S\}$. Then for all $i \in \{1,2,3,4\}$, $\delta_i < l$; so the number of outgoing edges from any vertex is bounded above by $\mathrm{card}(S) \times l^4$.
(3) Immediate. □

Proposition 14. *Let $(l_1,l_2,l_3,l_4) \in (A \cup \{\epsilon\})^4$ and $(x_a)_{a\in A} \in \mathbb{Z}^A$; $(\delta_1,\delta_2,\delta_3,\delta_4) \in \mathbb{N}^4$ and $\sigma \in S$ such that the substitute and cut operation ${}^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}$ is allowed on $X = \left(\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix}, (x_a)_{a\in A} \right)$ and has image $Y = \left(\begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}, (y_a)_{a\in A} \right)$.*

Then for all $\tilde{u}, \tilde{v} \in A^*$ such that:

1. $(\alpha(\tilde{u}), \omega(\tilde{u}), \alpha(\tilde{v}), \omega(\tilde{v})) = (l_1,l_2,l_3,l_4)$,
2. $\mathrm{ab}(\tilde{u}) - \mathrm{ab}(\tilde{v}) = (x_a)_{a\in A}$,

the truncations u and v of $\sigma(\tilde{u})$ and $\sigma(\tilde{v})$ defined by:

$$\begin{cases} p_{\delta_1}(\sigma(\tilde{u})) \cdot u \cdot s_{\delta_2}(\sigma(\tilde{u})) = \sigma(\tilde{u}), \\ p_{\delta_3}(\sigma(\tilde{v})) \cdot v \cdot s_{\delta_4}(\sigma(\tilde{v})) = \sigma(\tilde{v}); \end{cases}$$

satisfy:

1. $(\alpha(u), \omega(u), \alpha(v), \omega(v)) = (m_1,m_2,m_3,m_4)$,
2. $\mathrm{ab}(u) - \mathrm{ab}(v) = (y_a)_{a\in A}$.

Proposition 14 rules the construction of the automaton of imbalances, as detailed in Section 4.

3.4 Example: the imbalance automaton of the Thue-Morse subshift

Consider X_S the S-adic system for $S = \{\sigma\}$, where σ is the Thue-Morse substitution:

$$\begin{array}{rccc} \sigma : & a & \mapsto & ab \\ & b & \mapsto & ba. \end{array}$$

Observe that:

i) the substitution σ is nonerasing;

ii) the set S generates two S-adic words (described in Example 2);

iii) the letters a and b both appear in these words.

In what follows, we are going to breadth first traverse the automaton of imbalances of X_S, from its initial states. In order to reduce the growth of the tree (strictly, it is a forest with one main tree), we will:

– consider vertices up to some symmetries (see Lemma 16);

– trim the branches that will never lead to a final state (see Lemma 15).

We will obtain a finite graph (see Figure 1), on which we will read the maximal imbalance of words in X_w: 2, thus retrieving a well-known result.

Denote by $\mathtt{G} = (\mathtt{V}, \mathtt{I}, \mathtt{F})$ the automaton of imbalances of X_S. It has 7 initial states, namely:

$$\begin{pmatrix} \epsilon & \epsilon \\ \epsilon & \epsilon \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} a & \epsilon \\ \epsilon & \epsilon \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \epsilon & \epsilon \\ a & \epsilon \end{pmatrix}\begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} a & \epsilon \\ a & \epsilon \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} b & \epsilon \\ \epsilon & \epsilon \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \epsilon & \epsilon \\ b & \epsilon \end{pmatrix}\begin{pmatrix} 0 \\ -1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} b & \epsilon \\ b & \epsilon \end{pmatrix}\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

(In this subsection, it will be more convenient to represent the second vector as a column instead of a line.) Each vertex in $\mathtt{V}$ admits at most 16 outgoing edges, among:

$$\left\{ {}^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}, \ \text{for } (\delta_1, \delta_2, \delta_3, \delta_4) \in \{0,1\}^4 \right\}.$$

We recall that a vertex Y is *accessible* from a vertex X if there exists a sequence of edges in $\mathtt{G}$ which leads from X to Y - or, in other words, if Y is the image of X by a finite composition of substitute and cut operations.

Lemma 15. *No vertex in* $\mathtt{F}$ *is accessible from a vertex of the form*

$$\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix}\begin{pmatrix} x_a \\ x_b \end{pmatrix}$$

with $|x_a + x_b| \geq 2$.

Therefore, during the breadth first traversal of $\mathtt{G}$, we ignore all vertices such that $|x_a + x_b| \geq 2$: Lemma 15 guarantees that we miss no finite state.

Proof of Lemma 15. Let $X = \left(\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix}, \begin{pmatrix} x_a \\ x_b \end{pmatrix} \right)$ be a vertex in $\mathtt{V}$, ${}^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}$ a substitute and cut operation allowed from X, and denote by $Y = \left(\begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}, \begin{pmatrix} y_a \\ y_b \end{pmatrix} \right)$ its image. We are going to prove that $|y_a + y_b| \geq 2$. This first implies that $Y \notin \mathtt{F}$ (remind that a state is final if an only if it satisfies $y_a + y_b = 0$), and, iteratively, that no final state is the image of X by a finite composition of substitute and cut operations.

Let $\tilde{u}, \tilde{v} \in \{a,b\}^*$ such that $(\alpha(\tilde{u}), \omega(\tilde{u}), \alpha(\tilde{v}), \omega(\tilde{v})) = (l_1, l_2, l_3, l_4)$ and $\mathrm{ab}(\tilde{u}) - \mathrm{ab}(\tilde{v}) = (x_a, x_b)$. We have $||\tilde{u}| - |\tilde{v}|| = |x_a + x_b| \geq 2$; without loss of generality, assume that $|\tilde{u}| - |\tilde{v}| \geq 2$. Then we have $|\sigma(\tilde{u})| - |\sigma(\tilde{v})| \geq 4$. Denote by u and v the truncations of $\sigma(\tilde{u})$ and $\sigma(\tilde{v})$ given by:

$$\begin{cases} p_{\delta_1}(\sigma(\tilde{u})) \cdot u \cdot s_{\delta_2}(\sigma(\tilde{u})) = \sigma(\tilde{u}), \\ p_{\delta_3}(\sigma(\tilde{v})) \cdot v \cdot s_{\delta_4}(\sigma(\tilde{v})) = \sigma(\tilde{v}). \end{cases}$$

Then, by Proposition 14, we have $\mathrm{ab}(u) - \mathrm{ab}(v) = (y_a, y_b)$, from which we deduce that $|y_a + y_b| = |u| - |v|$. Finally, observing that $(\delta_1, \delta_2, \delta_3, \delta_4) \in \{0,1\}^4$ for the Thue-Morse substitution, we deduce that: $|u| - |v| \geq 2$; this concludes the proof. □

Lemma 16 (Symmetries)**.** *Let* $\left(\begin{pmatrix} l_1 & l_2 \\ l_3 & l_4 \end{pmatrix}, \begin{pmatrix} x_a \\ x_b \end{pmatrix}\right) \in \mathtt{V}$. *Then the vertices*

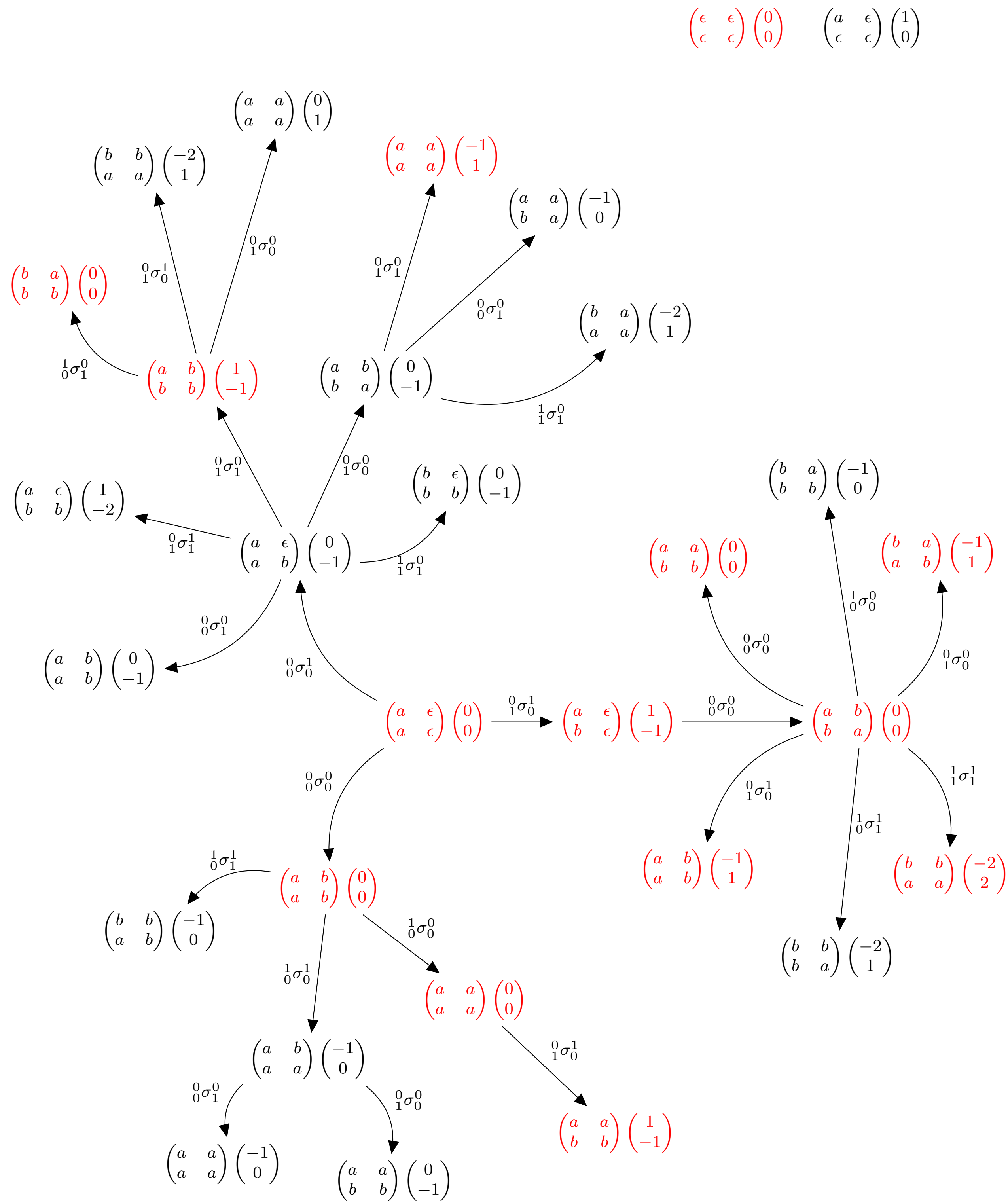
$$\begin{pmatrix} l_3 & l_4 \\ l_1 & l_2 \end{pmatrix}\begin{pmatrix} -x_a \\ -x_b \end{pmatrix}, \quad \begin{pmatrix} \neg l_1 & \neg l_2 \\ \neg l_3 & \neg l_4 \end{pmatrix}\begin{pmatrix} x_b \\ x_a \end{pmatrix} \quad \text{and} \quad \begin{cases} \begin{pmatrix} l_2 & l_1 \\ l_4 & l_3 \end{pmatrix}\begin{pmatrix} x_a \\ x_b \end{pmatrix} & \text{if } (l_2, l_4) \neq (\epsilon, \epsilon), \\ \begin{pmatrix} l_2 & l_1 \\ l_3 & l_4 \end{pmatrix}\begin{pmatrix} x_a \\ x_b \end{pmatrix} & \text{if } l_2 \neq \epsilon \text{ and } l_4 = \epsilon, \\ \begin{pmatrix} l_1 & l_2 \\ l_4 & l_3 \end{pmatrix}\begin{pmatrix} x_a \\ x_b \end{pmatrix} & \text{if } l_2 = \epsilon \text{ and } l_4 \neq \epsilon. \end{cases}$$

where $\neg a = b$ *and* $\neg b = a$, *also belong to* $\mathtt{V}$.

Proof. Let $w \in X_S$ and $(u,v) \in \mathcal{F}(w)^2$ such that $(\alpha(u), \omega(u), \alpha(v), \omega(v)) = (l_1, l_2, l_3, l_4)$ and $\mathrm{ab}(u) - \mathrm{ab}(v) = (x_a, x_b)$. We immediately see that the pair (v,u) also belong to $\mathcal{F}(w)^2$, which proves that the first symmetric vertex is also in $\mathtt{V}$. One easily checks from Example 2 that if $w \in X_S$, then $\neg w$ also belongs to X_S. Since the pair $(\neg u, \neg v)$ belongs to $\mathcal{F}(\neg w)$, the second symmetric vertex is also in $\mathtt{V}$. At last, the Thue-Morse subshift is such that if $u \in \mathcal{F}(w)$, then $\mathrm{rev}(u)$, where $\mathrm{rev}(u)$ denotes the word u written backwards, is a factor of w too; we deduce that the third symmetric vertices belong to $\mathtt{V}$. □

Below, we draw the forest obtained after a breadth first traversal of $\mathtt{G}$, from its initial states. This forest is simplified by:

- considering the vertices up to the symmetries of Lemma 16 (as a consequence, there remain only three initial states);
- deleting all vertices such that $|x_a + x_b| \geq 2$ (see Lemma 15).



In red: final states.

Figure 1: The forest obtained after a breadth first traversal of `G` for the Thue-Morse subshift.

The maximal imbalance found is 2 and corresponds to the vertex $\binom{b\,b}{a\,a}\binom{-2}{2}$.

4 Proof

In this Section, we describe step by step the construction of the automaton of imbalances.

4.1 General idea

Let S be a finite set of nonerasing substitutions over a common alphabet A. We consider the set of all imbalance vectors of words in X_S:

$$\mathtt{F}_3 := \bigcup_{w\in X_S} \{\mathrm{ab}(u) - \mathrm{ab}(v) \mid u, v \in \mathcal{F}(w) \text{ and } |u| = |v|\} \subset \mathbb{Z}^A.$$

Proposition 17 (immediate)**.** *The set $\mathtt{F}_3$ is finite if and only if there exists $D \in \mathbb{N}$ such that for all $w \in X_S$, $\mathrm{imb}(w) \leq D$.*

So we want to explore the set $\mathtt{F}_3$. To do so, we need to understand where do the imbalance vectors come from. A classic idea, in the substitutive framework, and we extend to S-adic systems, consists in tracking backwards the history of factors by successive desubstitutions. This is why we introduce the underlying set:

$$\mathtt{F}_2 := \bigcup_{w\in X_S} \{(u, v) \mid u, v \in \mathcal{F}(w) \text{ and } |u| = |v|\},$$

that we will explore through a (huge) graph $\mathtt{G}_2$, whose automaton of imbalances $\mathtt{G}$ is a quotient graph, obtained after partial abelianization.

4.2 Desubstitution

Let $u, \tilde{u} \in A^*$ and σ a substitution over A. We say that $\tilde{u}$ is a *predecessor* of u by σ if there exists a pair $p, s \in A^*$ such that $pus = \sigma(\tilde{u})$. We denote by $\mathcal{P}_\sigma(u)$ the set of all predecessors of u by σ, that we endow with the relation:

$$\tilde{u}_1 \leq \tilde{u}_2 \text{ if and only if } \tilde{u}_1 \in \mathcal{F}(\tilde{u}_2).$$

The set $\mathcal{P}_\sigma(u)$ is not totally ordered: for instance 1 and 22 are two incomparable predecessors of 2 for the Thue-Morse substitution σ_{TM} (defined in Example 2). However, the partial order is well-founded: all nonempty subset of $\mathcal{P}_\sigma(u)$ admit minimal elements. *Desubstituting* u by σ consists in returning one of its minimal predecessors.

Example 18. *The set of all predecessors of 2 by σ_{TM} is $\{1, 2\}^*\backslash\{\epsilon\}$; its minimal predecessors are 1 and 2. The word 222 has no predecessor by σ_{TM}, therefore, it can not be desubstituted.*

Observe that we always have $\mathcal{P}_\sigma(\epsilon) = A^*$, and that the only minimal predecessor of ϵ is itself.

Proposition 19. *Let $u \in A^*\backslash\{\epsilon\}$ and $\tilde{u} \in \mathcal{P}_\sigma(u)$. Then $\tilde{u}$ is nonempty. It is furthermore minimal if and only if all pairs $(p, s) \in A^*$ such that $pus = \sigma(\tilde{u})$ satisfy $|p| < |\sigma(\tilde{u}[0])|$ and $|s| < |\sigma(\tilde{u}[-1])|$ ($\star$). Remainder: $\tilde{u}[0]$ and $\tilde{u}[-1]$ denote the first and last (possibly same) letter of $\tilde{u}$.*

Proof. We immediately see that if u is nonempty, then $\tilde{u}$ is also nonempty. We now prove both implications by contraposition.

Assume that $\tilde{u}$ is not minimal. There exists $\tilde{v} \in \mathcal{P}_\sigma(u)$ and $a, b \in A^*$ non both empty (say, without loss of generality, $a \neq \epsilon$), such that $\tilde{u} = a\tilde{v}b$. Then, by writing $\sigma(\tilde{v}) = pus$, we obtain: $\sigma(\tilde{u}) = \sigma(a)p \cdot u \cdot s\sigma(b)$, with $|\sigma(a)p| \geq |\sigma(a)| \geq |\sigma(\tilde{u}[0])|$, since $a \neq \epsilon$.

Conversely, assume that there exists a pair $p, s \in A^*$ which does not satisfy $(\star)$ (say, without loss of generality, $|p| \geq |\sigma(\tilde{u}[0])|$). One easily checks that the suffix of length $|\tilde{u}| - 1$ of $\tilde{u}$ is still a predecessor of u, which implies that $\tilde{u}$ is not minimal. □

Counterintuitively, the existence of one pair satisfying $(\star)$ is not sufficient to guarantee the minimality of a predecessor.

Example 20. *Consider the substitution σ_{cex} defined by $\sigma_{cex}(0) = 10$ and $\sigma_{cex}(1) = 1010$. Denote $\tilde{u} = 101$, $u = 0101$, $p = 101$ and $s = 010$. Observe that:*

$$\begin{cases} \sigma_{cex}(\tilde{u}) = 1010101010 = pus \\ |p| < \sigma(\tilde{u}[0]) \\ |s| < \sigma(\tilde{u}[-1]) \end{cases}$$

and, yet, $\tilde{u}$ is not minimal since its strict factor 10 *is also a predecessor of u:*

$$\sigma_{cex}(10) = 101010 = 1u0.$$

Corollary 21. *Let σ be a substitution and $u \in \mathcal{F}(\sigma(w))\backslash\{\epsilon\}$, where w is a finite or infinite word. Then there exist $\tilde{u} \in \mathcal{F}(w)\backslash\{\epsilon\}$ together with a pair p, s of finite words satisfying $(\star)$ such that $pus = \sigma(\tilde{u})$.*

Proof. The set $\mathcal{P}_\sigma(u)$ is nonempty (indeed, it contains w if w is finite, or one of its factors otherwise), so it admits minimal elements. We conclude with Proposition 19. □

Finally, if u and v are two factors of a S-adic word $w = \lim_{n\to\infty} d_0 \circ ... \circ d_{n-1}(a)$, we can simultaneously desubstitute u and v by d_0; unfortunately, there is no guarantee that, among the pairs of minimal predecessors $(\tilde{u}, \tilde{v})$, there is one at least which satisfies $\tilde{u} = \tilde{v}$, or equivalently, which still belongs to $\mathtt{F}_2$. So we introduce the larger set:

$$\mathtt{V}_2 = \bigcup_{w \in X_S} \{(u, v) \,|\, u, v \in \mathcal{F}(w)\} \subset A^* \times A^*,$$

that will be the vertices of the (huge) graph we construct.

4.3 An inverse to desubstitution: the *substitute and cut* operation

In order to explore the set of factors in a constructive way, we choose for transition not the desubstitution, but its "inverse", which consists in substituting and cropping the image according to some rules.

In the sequel, we refer to the functions α and ω defined in Notation 10.

Definition 22. *Let u and $\tilde{u}$ in A^*. A* substitute and cut operation *from $\tilde{u}$ to u is a triplet $(\sigma, \delta_1, \delta_2)$, where σ is a substitution and δ_1,δ_2 two nonnegative integers such that:*

1. $p_{\delta_1}(\sigma(\tilde{u})) \cdot u \cdot s_{\delta_2}(\sigma(\tilde{u})) = \sigma(\tilde{u})$.

2. $\begin{cases} \delta_1 = \delta_2 = 0 & \text{if } \tilde{u} = \epsilon, \\ \delta_1 < |\sigma(\tilde{u})|,\ \delta_2 < |\sigma(\tilde{u})| \text{ and } \delta_1 + \delta_2 < |\sigma(\tilde{u})| & \text{if } |\tilde{u}| = 1, \\ \delta_1 < |\sigma(\alpha(\tilde{u}))| \text{ and } \delta_2 < |\sigma(\omega(\tilde{u}))| & \text{otherwise.} \end{cases}$

We denote it by:

$$^{\delta_1}\sigma^{\delta_2}(\tilde{u}) = u.$$

Proposition - definition 23. *Let $\tilde{u} \in A^*$, σ a substitution defined over A, and δ_1, $\delta_2 \in \mathbb{N}$. The two assertions are equivalent:*

1. *$\tilde{u}$, σ, δ_1 and δ_2 satisfy Condition (2) of Definition 22;*

2. *there exists $u \in A^*$ such that $u = {}^{\delta_1}\sigma^{\delta_2}(\tilde{u})$.*

When these assertions are satisfied, we say that $(\sigma, \delta_1, \delta_2)$ is a substitute and cut operation allowed *from $\tilde{u}$.*

Proof. Since (2) trivially implies (1), we just need to check that (1)⇒(2). Let us assume that $\tilde{u}$, σ, δ_1 and δ_2 satisfy Condition (2) of Definition 22. If $\tilde{u} = \epsilon$, it suffices to take $u = \epsilon$. Otherwise, the bounds on δ_1 and δ_2 ensure that the equation: $p_{\delta_1}(\sigma(\tilde{u})) \cdot u \cdot s_{\delta_2}(\sigma(\tilde{u})) = \sigma(\tilde{u})$ defines a (unique and nonempty) word u. □

This definition is consistent with Definition 9, as we see later.

Example 24. *One easily checks that:*

1. *the triplet $(\sigma_{TM}, 1, 1)$ (with σ_{TM} the Thue-Morse substitution defined in Subsection 2.1) is a substitute and cut operation from* 21 *to* 11*;*

2. *the triplet $(\sigma_{cex}, 3, 3)$ (where σ_{cex} is the substitution defined in Example 20) is a substitute and cut operation from* 101 *to* 0101*.*

Example 24(2) shows that a finite word is not always a minimal predecessor of its substitute and cut image. Nonetheless:

Lemma 25. *Let $u \in A^*$ and σ a substitution over A. The set of all preimages of u by allowed substitute and cut operations built over σ is finite and contains all minimal predecessors of u by σ.*

Proof. Let $u \in A^*$, σ a substitution and (δ_1, δ_2) a pair of integer such that $(\sigma, \delta_1, \delta_2)$ is a substitute and cut operation allowed on $\tilde{u}$. Then we must have $\delta_1, \delta_2 \leq \max_{a \in A} |\sigma(a)|$; hence the finiteness of the set. Now, if $\tilde{u}$ is a minimal predecessor of u by σ, either $u = \tilde{u} = \epsilon$ and we have ${}^0\sigma^0(\epsilon) = \epsilon$, or $u \neq \epsilon$ and we know from Proposition 19 that all pairs of words p, s such that $\sigma(\tilde{u}) = pus$ will satisfy Condition (2), so we can write $u = {}^{|p|}\sigma^{|s|}(\tilde{u})$. □

4.4 Aside: the graph of all factors of a S-adic system

Let S be a set of substitutions defined over a common alphabet A. Denote by $\mathtt{V}_1$ the set of all factors of the S-adic system X_S:

$$\mathtt{V}_1 := \bigcup_{w \in X_S} \mathcal{F}(w) = \bigcup_{w \text{ S-adique}} \mathcal{F}(w) \qquad \text{(see Lemma 6).}$$

We consider the oriented graph $\mathtt{G}_1$ whose set of vertices is $\mathtt{V}_1$, and whose edges map vertices to their images by all allowed substitute and cut operations $(\sigma, \delta_1, \delta_2)$, with $\sigma \in S$.

Remark 26. *In the particular case $S = \{\sigma\}$, with σ a primitive substitution, the subgraph of $\mathtt{G}_1$ restricted to the vertices of length 1 (i.e. letters in A) is the automaton of prefixes-suffixes, with reversed edges, defined in [CS01].*

Lemma 27 (immediate)**.** *(i) If it is not empty, the graph $\mathtt{G}_1$ has infinitely many vertices.*
(ii) The number of outgoing edges of any vertex is bounded above by $\mathrm{card}(S) \times l^2$*, where* $l = \max_{\sigma\in S, a\in A} |\sigma(a)|$.

We say that a factor $v \in \mathtt{V}_1$ is *accessible* from a factor $u \in \mathtt{V}$ if there exists a finite path in $\mathtt{G}$ which goes from u to v.

Proposition 28. *If $A \subset \mathtt{V}_1$, then any nonempty factor in $\mathtt{V}_1$ is accessible from a letter in A.*

Proof. Assume that $A \subset \mathtt{V}_1$ and pick a nonempty factor u in $\mathtt{V}_1$. By definition, there exists a letter $a \in A$ and a directive sequence $(d_n)_{n\in\mathbb{N}} \in S^{\mathbb{N}}$ such that $u \in \mathcal{F}(w)$ with $w = \lim_n d_0 \circ ... \circ d_{n-1}(a)$. Thus, there exists a [finite] prefix v of w such that $u \in \mathcal{F}(v)$; and, by definition of the convergence (see Subsection 2.1), there exists a rank $n_0 \in \mathbb{N}$ such that v is also a prefix of the finite words $d_0 \circ ... \circ d_{n-1}(a)$ for all n larger than n_0.

Now, we recursively construct three finite sequences: $(u_k)_{k\in\{0,...,n_0\}} \in \mathtt{V}^{n_0+1}$, $(\beta_k)_{k\in\{0,...,n_0-1\}}$ and $(\gamma_k)_{k\in\{0,...,n_0-1\}} \in \mathbb{N}^{n_0}$:

- $u_0 = u$,
- for $k \in \{0, ..., n_0 - 1\}$, we denote by $(p, \tilde{u}, s)$ the triplet of words given by the application of Corollary 21 to the substitution d_k and the word $u_k \in \mathcal{F}(d_k(w_{k+1}))$, with $w_{k+1} := d_{k+1} \circ ... \circ d_{n_0-1}(a)$; we then set $u_{k+1} = \tilde{u}$, $\beta_k = |p|$ and $\gamma_k = |s|$.

Thus, the sequences $(u_k)_{k\in\{0,...,n_0\}}$, $(\beta_k)_{k\in\{0,...,n_0-1\}}$ and $(\gamma_k)_{k\in\{0,...,n_0-1\}}$ are such that for all $k \in \{0, ..., n_0 - 1\}$, u_k is the image of u_{k+1} by the allowed substitute and cut operation (d_k, β_k, γ_k). Furthermore, Corollary 21 ensures that at each step of the recursion, $u_k \in \mathcal{F}(d_k \circ ... \circ d_{n_0-1}(a)) \backslash \{\epsilon\}$; so in particular $u_{n_0} \in \mathcal{F}(a) \backslash \{\epsilon\}$, which implies $u_{n_0} = a$.

Finally, we exhibited a finite path in $\mathtt{G}$, which goes from $a \in A \subset \mathtt{V}_1$ to u; this concludes the proof. □

This result suggests that we could obtain all factors in $\mathtt{V}_1$ by simple knowledge of the alphabet. Hence the question: are all words obtained after a finite composition of allowed substitute and cut operations still in $\mathtt{V}_1$?

Proposition 29. *If $A \subset \mathtt{V}_1$ and if u is the image of a letter by a finite composition of allowed substitute and cut operations, then $u \in \mathtt{V}_1$.*

Proof. Let $a \in A$ and $(\sigma_k, \beta_k, \gamma_k)_{k\in\{0,...,n-1\}} \in (S \times \mathbb{N}^2)^n$ such that for all $k \in \{0, ..., n_0 - 1\}$, the substitute and cut operation $(\sigma_k, \beta_k\gamma_k)$ is allowed on the word ${}^{\beta_{k-1}}\sigma_{k-1}{}^{\gamma_{k-1}} \circ ... \circ {}^{\beta_0}\sigma_0^{\gamma_0}(a)$. Denote $u := {}^{\beta_{n-1}}\sigma_{n-1}{}^{\gamma_{n-1}} \circ ... \circ {}^{\beta_0}\sigma_0^{\gamma_0}(a) \in A^*$; we want to prove that $u \in \mathtt{V}_1$. Since $a \in A \subset \mathtt{V}_1$, there exists w a S-adic word such that $a \in \mathcal{F}(w)$. But then, the infinite word $w' := \sigma_{n-1} \circ ... \circ \sigma_0(w)$ is also S-adic, and we immediately have $u \in \mathcal{F}(w')$, hence $u \in \mathtt{V}_1$. □

The assumption $A \subset \mathtt{V}_1$ is essential.

Example 30. *Consider $S = \{\sigma\}$, where the substitution σ is defined by $\sigma(a) = baa$ and $\sigma(b) = bb$. One easily checks that the S-adic system X_S contains a unique word, which is:*

$$w = bbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbb...$$

Therefore:

$$\mathtt{V}_1 = \{\epsilon, b, bb, bbb, bbbb, bbbbb, ...\}$$

and yet, the words a, baa, baaba, etc, can be obtained from letters by finite compositions of allowed substitute and cut operations.

4.5 The graph of all pairs of factors of S-adic words

4.5.1 Simultaneous substitute and cut operations

In order to study the imbalances of S-adic words, we do not want to reconstruct all pairs of factors in X_S, but all pairs of words (u, v) in which u and v are *simultaneous* factors of a S-adic word.

Definition 31. *Let $u, v, \tilde{u}$ and $\tilde{v}$ in A^*. A* [simultaneous] substitute and cut operation *from the pair $(\tilde{u}, \tilde{v})$ to the pair (u, v) is a quintuplet $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4)$ such that:*

- *$(\sigma, \delta_1, \delta_2)$ is a substitute and cut operation from $\tilde{u}$ to u,*
- *$(\sigma, \delta_3, \delta_4)$ is a substitute and cut operation from $\tilde{v}$ to v.*

We denote it by:

$$^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}(\tilde{u}, \tilde{v}) = (u, v).$$

A quintuplet $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4) \in S \times \mathbb{N}^4$ is an allowed substitute and cut operation from $(\tilde{u}, \tilde{v})$ if there exists a pair $(u, v) \in (A^*)^2$ such that $(u, v) = {}^{\delta_1}_{\delta_3}\sigma^{\delta_2}_{\delta_4}(\tilde{u}, \tilde{v})$.

The crucial point in Definition 31 is that we run the same substitution on both words, but we crop their images independently- these variations will generate imbalance.

Lemma 32. *Let $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4) \in S \times \mathbb{N}^4$ and $(\tilde{u}, \tilde{v}) \in (A^*)^2$. The following assertions are equivalent:*

1. *the quintuplet $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4)$ is an allowed substitute and cut operation from $(\tilde{u}, \tilde{v}) \in (A^*)^2$;*
2. *the triplets $(\sigma, \delta_1, \delta_2)$ and $(\sigma, \delta_3, \delta_4)$ are allowed on $\tilde{u}$ and $\tilde{v}$ respectively;*
3. *both triplets $(\sigma, \delta_1, \delta_2)$ and $(\sigma, \delta_3, \delta_4)$ satisfy Condition (2) in Definition 22.*

Proof. This comes immediately from Proposition-definition 23. □

Example 33. *We consider the set $C = \{c1, c2\}$ of Cassaigne-Selmer substitutions, defined in Example-Definition 4. Let $(\tilde{u}, \tilde{v}) = (23, 33)$.*
There are exactly four substitute and cut operations allowed from the pair $(\tilde{u}, \tilde{v})$:

$$^{0}_{0}c1^{0}_{0} \quad , \quad ^{1}_{0}c1^{0}_{0} \quad , \quad ^{0}_{0}c2^{0}_{0} \quad \textit{and} \quad ^{1}_{0}c2^{0}_{0}.$$

Their respective images are: $(132, 22)$, $(32, 22)$,$(133, 33)$ *and* $(33, 33)$.

4.5.2 The graph of all pairs of factors of S-adic words

Let $\mathtt{G}_2$ be the oriented graph whose set of vertices is

$$\mathtt{V}_2 := \bigcup_{w \in X_S} \{(u,v) \,|u,v \in \mathcal{F}(w)\} = \bigcup_{w \text{ S-adique}} \{(u,v) \,|u,v \in \mathcal{F}(w)\}$$

and whose edges map vertices to their images by all allowed [simultaneous] substitute and cut operations.

We consider two remarkable subsets of $\mathtt{V}_2$:

$$\mathtt{I}_2 := \mathtt{V}_2 \cap \left(\{(\epsilon,\epsilon)\} \cup \cup_{a\in A}\{(a,\epsilon),(\epsilon,a),(a,a)\}\right), \qquad \text{called } \textit{initial states};$$

$$\mathtt{F}_2 := \bigcup_{w \text{ S-adique}} \{(u,v) \,|\, u,v \in \mathcal{F}(w) \text{ and } |u| = |v|\}, \qquad \text{called } \textit{finite states}.$$

Finite states corresponds to the pairs of factors we are interested in when studying the imbalances of infinite words in a S-adic system. To access them, Proposition 34 (below) suggests to traverse the graph $\mathtt{G}$ from its initial states.

Proposition 34. *1. If $\mathtt{G}_2$ is nonempty, then it has infinitely many vertices.*

2. The number of outgoing edges from any vertex is bounded above by $\mathrm{card}(S) \times l^4$*, where* $l = \max_{\sigma\in S, a\in A} |\sigma(a)|$.

3. If $A \subset \mathtt{V}_2$, then any pair of nonempty words in $\mathtt{V}_2$ is accessible from an initial state.

4. If $A \subset \mathtt{V}_2$ and if (u,v) is the image of an initial state by a finite composition of allowed [simultaneous] substitute and cut operations, then $(u,v) \in \mathtt{V}_2$.

Proof. Assertions (1) and (2) are immediate adaptations of Lemma 27.
Assertions (3) and (4) are easy adaptations of Propositions 28 and 29 respectively. The key point is to remember that $(u,v) \in \mathtt{V}_2$ means that u and v are factors of the same S-adic word w, so they can be simultaneously desubstituted by the first substitution of its directive sequence; conversely, the image of a pair in $\mathtt{V}_2$ by an allowed simultaneous substitute and cut operation is still made of factors of a same S-adic word. □

4.5.3 A semi-algorithm to detect imbalances in a S-adic system

Below, we explicitly describe a semi-algorithm to detect high imbalances in X_S.

Algorithm 35.
INPUTS:
- a finite set of substitutions S over a common alphabet A
- a nonnegative integer d

ALGORITHM
$V \leftarrow \mathtt{I}_2$ (initialization)
$V' \leftarrow$ empty set
for $(\tilde{u},\tilde{v})$ in V:
    for $(\sigma,\delta_1,\delta_2,\delta_3,\delta_4)$ substitute and cut operation allowed on (u,v):
        (u,v) $\leftarrow$ image of $(\tilde{u},\tilde{v})$ by the substitute and cut operation
        if (u,v) has not yet appeared in the processing:

if $|u| = |v|$ *and* $\mathrm{imb}(u, v) \geq d$*:*
  return `true`
otherwise:
  put (u,v) in V'
V $\leftarrow$ *V'*
V' $\leftarrow$ *empty set*

Corollary 36 (immediate consequence of Proposition 34)**.** *Let* S *be a finite set of nonerasing substitutions defined over a common alphabet* A*. Assume furthermore that* $A \subset \mathtt{V}_2(S)$*. Then for all* $d \leq \sup_{w \in X_S} \mathrm{imb}(w)$*, Algorithm 35, run on the entry* (S, d)*, will finish.*

Remark 37. *If at each first visit of a pair* (u, v)*, we record the preimage and the substitute and cut operation which gave* (u, v)*, then we are able to retrace a shortest [finite] sequence of substitute and cut operations which lead from one of the initial states to the intended imbalance. Thus, if we denote by* $(\sigma_0, ..., \sigma_{n-1})$ *the substitutions appearing in the labels of this path, and by* $\tau := \sigma_{n-1} \circ ... \circ \sigma_0$ *the substitution obtained by backwards composition of them, then for all S-adic word* w *containing the letter* a*, where* a *is the letter appearing in the first [initial] state of the path, the imbalance of the S-adic word* $\tau(w)$ *is larger or equal than* d*.*

Remark 38. *By contrast, if* $d > \sup_{w \in X_S} \mathrm{imb}(w)$*, the algorithm will indefinitely keep running on the entry* (S, d)*, hence the terminology "semi-algorithm".*

4.6 Construction of the automaton of imbalances

However, to study the imbalance of a S-adic word, it is not necessary to fully describe its factors: the simple knowledge of their abelianized vectors is sufficient. Put in another way, the graph $\mathtt{G}_2$ contains much more information than what is actually needed.

The aim of this subsection is to simplify it. This simplication (and some others...) will make its computation possible (see Section 5).

4.6.1 Obstructions to the abelianization of $\mathtt{G}_2$

We consider the binary relation $\mathcal{R}_{ab}$ defined over pairs of finite words:

$$(u, v)\, \mathcal{R}_{ab}\, (u', v') \text{ if and only if } \mathrm{ab}(u) - \mathrm{ab}(v) = \mathrm{ab}(u') - \mathrm{ab}(v'),$$

which partitions $\mathtt{V}_2$ into equivalence classes:

$$\mathtt{V}_3 := \mathtt{V}_2/\mathcal{R}_{ab} = \bigcup_{w \text{ S-adic}} \{\mathrm{ab}(u) - \mathrm{ab}(v) \mid u, v \in \mathcal{F}(w)\}.$$

Unfortunately, this partition is not compatible with the substitute and cut operation. Indeed, a same vector $x \in \mathtt{V}_3$ may represent two pairs of factors (u, v) and (u', v') for which the sets of allowed substitute and cut operations (and the results they give) are different (see Example 39 below).

Example 39. *We consider the set* C *of Cassaigne-Selmer substitutions, defined in Example-Definition 4. The two pairs* $(u, v) = (132, 132)$ *and* $(u', v') = (\epsilon, \epsilon)$ *in* $\mathtt{V}_2$ *both belong to the equivalence class* $x = (0, 0, 0) \in \mathtt{V}_3$*. However:*
- there are exactly eight substitute and cut operations allowed from the pair (u, v)*:*

$$ {}^0_0c1^0_0 \quad , \quad {}^0_0c1^1_0 \quad , \quad {}^0_0c1^0_1 \quad , \quad {}^0_0c1^1_1$$
$$ {}^0_0c2^0_0 \quad , \quad {}^0_0c2^1_0 \quad , \quad {}^0_0c2^0_1 \quad \textit{and} \quad {}^0_0c2^1_1;$$

- only the first and the fifth ones are also allowed on (u', v')*.*

Fundamentally, a vector $x \in \mathtt{V}_3$ contains no information on the initial and final letters of the pairs of factors it represents. This information is yet essential when applying Condition (2) of Definition 22.

4.6.2 Towards a partial abelianization of $\mathtt{G}_2$

This is why we consider a finer partition of $\mathtt{V}_2$. Let $\mathcal{R}_{pab}$ be the binary relation defined on pairs of finite words by:

$$(u,v)\,\mathcal{R}_{pab}\,(u',v') \text{ if and only if } \begin{cases} (\alpha(u),\omega(u),\alpha(v),\omega(v) = \alpha(u'),\omega(u'),\alpha(v'),\omega(v')) \\ \quad \text{and} \\ \mathrm{ab}(u)-\mathrm{ab}(v) = \mathrm{ab}(u')-\mathrm{ab}(v'). \end{cases}$$

We recall

Notation 10. *The functions α and ω are defined by:*

$$\begin{array}{rccl} \alpha: & A^* & \to & A\cup\{\epsilon\} \\ & u & \mapsto & \begin{cases} u[0] \ \textit{if}\ |u|\geq 1 \\ \epsilon \ \textit{otherwise;} \end{cases} \end{array} \qquad \begin{array}{rccl} \omega: & A^* & \to & A\cup\{\epsilon\} \\ & u & \mapsto & \begin{cases} u[-1] \ \textit{if}\ |u|\geq 2 \\ \epsilon \ \textit{otherwise;} \end{cases} \end{array}$$

where, following `Python`*, $u[0]$ and $u[-1]$ respectively denote the first and last letters of u. For instance, $\alpha(examples) = e$, $\omega(examples) = s$, $\alpha(a) = a$, $\omega(a) = \epsilon$, $\alpha(\epsilon) = \omega(\epsilon) = \epsilon$.*

The functions α and ω are constructed so as to save all the information necessary to apply substitute and cut operations.

Notation 40. *Let $(u,v) \in (A^*)^2$. The 2×2 table*

$$M_{u,v} := \begin{pmatrix} \alpha(u) & \omega(u) \\ \alpha(v) & \omega(v) \end{pmatrix}$$

is called matrix of extremities *of the pair (u,v).*

Example 41. *Let $A = \{a,b,c\}$.*
The pairs of words (acb, cc) and $(acab, cac)$ both belong to the equivalence class $\left(\begin{pmatrix} a & b \\ c & c \end{pmatrix}, (1,1,-1)\right)$.
The pair of words (a, cc) belongs to the class $\left(\begin{pmatrix} a & \epsilon \\ c & c \end{pmatrix}, (1,0,-2)\right)$.
The pair of words (acb, acb) belongs to the class $\left(\begin{pmatrix} a & b \\ a & b \end{pmatrix}, (0,0,0)\right)$, whereas the pair (ϵ,ϵ) belongs to the class $\left(\begin{pmatrix} \epsilon & \epsilon \\ \epsilon & \epsilon \end{pmatrix}, (0,0,0)\right)$. ‘

Let $\mathtt{V}$ denote the quotient set:

$$\mathtt{V} := \mathtt{V}_2/\mathcal{R}_{pab} = \bigcup_{w \text{ S-adic}} \{(M_{u,v}, \mathrm{ab}(u)-\mathrm{ab}(v)) \mid u,v \in \mathcal{F}(w)\}\,.$$

The [simultaneous] substitute and cut operation is compatible with the relation $\mathcal{R}_{pab}$.

Proposition 42. *Let S be a finite set of **nonerasing** substitutions defined over a common alphabet A. Let $X \in \mathtt{V}$. Let (u,v) and (u',v') be two pairs of words in the equivalence class X.*

i) A substitute and cut operation $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4) \in S \times \mathbb{N}^4$ is allowed from (u, v) if and only if it is allowed from (u', v'); in that case, we simply say that $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4)$ is allowed from X.

ii) Furthermore, the images of (u, v) and (u', v') by such a substitute and cut operation belong to a same equivalence class modulo $\mathcal{R}_{pab}$.

At last, the set of all allowed substitute and cut operations from X, as well as the equivalence classes of their images, can be computed from the simple knowledge of X.

It is essential for the substitutions to be non erasing.

Example 43. *Let σ be the [erasing] substitution:*

$$\begin{array}{cccc} \sigma & a & \mapsto & acb \\ & b & \mapsto & b \\ & c & \mapsto & \epsilon. \end{array}$$

We consider the two pairs of words $(\tilde{u}, \tilde{v}) = (cbab, \epsilon)$ and $(\tilde{u}', \tilde{v}') = (cabb, \epsilon)$, which belong to the same equivalence class modulo $\mathcal{R}_{pab}$, given by:

$$\left(\begin{pmatrix} c & b \\ \epsilon & \epsilon \end{pmatrix} (1, 2, 1) \right).$$

Their respective images by the allowed substitute and cut operation ${}^0_0\sigma^0_0$ are $(u, v) = (bacbb, \epsilon)$ and $(u', v') = (acbbb, \epsilon)$, which do not have the same matrix of extremities.

Proof of Proposition 42. i) Observe that the knowledge of $\alpha(u)$ and $\omega(u)$ in Condition (2) of Definition 22 is sufficient to determine the set of substitute and cut operations which are allowed from a word u. Therefore, the knowledge of $M_{u,v}$ is sufficient to determine the set of allowed simultaneous substitute and cut operations from a pair (u, v). Therefore, two pairs of words equal modulo $\mathcal{R}_{pab}$ must have the same set of allowed substitute and cut operations.

ii) Let $(\sigma, \delta_1, \delta_2)$ be an allowed substitute and cut operation from a word $\tilde{u}$. Denote $u = {}^{\delta_1}\sigma^{\delta_2}$. We first show that the knowledge of $\alpha(\tilde{u})$ and $\omega(\tilde{u})$ is sufficient to determine $\alpha(u)$ and $\omega(u)$.
- If $\omega(\tilde{u}) = \epsilon$, then $\tilde{u} \in A \cup \{\epsilon\}$, and the knowledge of $\alpha(\tilde{u}) = \tilde{u}$ is sufficient to determine the whole image $u = {}^{\delta_1}\sigma^{\delta_2}(\tilde{u})$, so in particular $\alpha(u)$ and $\omega(u)$.
- Otherwise, we know that $|\tilde{u}| \geq 2$. In this case, $l_1 = \alpha(\tilde{u})$ and $l_2 = \omega(\tilde{u})$ are respectively the first and last letter of $\tilde{u}$. Since σ is nonerasing, we have $|\sigma(l_1)| \geq 1$ and $|\sigma(l_2)| \geq 1$; since furthermore $\delta_1 < |\sigma(l_1)|$ and $\delta_2 < |\sigma(l_2)|$, the words ${}^{\delta_1}\sigma^0(l_1)$ and ${}^0\sigma^{\delta_2}(l_2)$ are nonempty. These two words are respectively prefix and suffix of u, with moreover $|u| \geq 2$, so we explicitly know the first and last letters of u. Finally, in this case too, the knowledge of $\alpha(\tilde{u})$ and $\omega(\tilde{u})$ is sufficient to determine $\alpha(u)$ and $\omega(u)$.
Now, we show that the knowledge of $\alpha(\tilde{u})$, $\omega(\tilde{u})$ and $\mathrm{ab}(\tilde{u})$ is sufficient to compute $\mathrm{ab}(u)$. Indeed, if M_σ denotes the incidence matrix of σ, we have:

$$\begin{array}{rcl} \mathrm{ab}(u) & = & \mathrm{ab}(\tilde{u})M_\sigma - \mathrm{ab}(p_{\delta_1}(\sigma(\tilde{u}))) - \mathrm{ab}(s_{\delta_2}(\sigma(\tilde{u}))) \\ & = & \mathrm{ab}(\tilde{u})M_\sigma - \mathrm{ab}(p_{\delta_1}(\sigma(\alpha(\tilde{u})))) - \mathrm{ab}(s_{\delta_2}(\sigma(\omega(\tilde{u})))). \end{array}$$

When applied to pairs of words, this says that the knowledge of $M_{\tilde{u},\tilde{v}}$ and $\mathrm{ab}(\tilde{u}) - \mathrm{ab}(\tilde{v})$ (we use here the linearity of the matrix product in the expression above) is sufficient to determine $M_{u,v}$ and $\mathrm{ab}(u) - \mathrm{ab}(v)$, where (u, v) is the image of $(\tilde{u}, \tilde{v})$ by a substitute and cut operation $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4)$. In particular, the images, by an allowed substitute and cut operation, of two pairs of words which equal modulo $\mathcal{R}_{pab}$ are still equal modulo $\mathcal{R}_{pab}$. $\square$

Proposition 42 enable us to define the substitute and cut operations over the quotient set $\mathtt{V} = \mathtt{V}_2/\mathcal{R}_{pab}$. This is detailed in Definition 9.

Example 44. *We consider the set* $C := \{c1, c2\}$ *of Cassaigne-Selmer substitutions (which are nonerasing). We set* $X = \left(\left(\begin{smallmatrix}2 & 3\\ 3 & 3\end{smallmatrix}\right), (0,1,-1)\right)$*. There are exactly four substitute and cut operations allowed from* X*:*

$$ {}^{0}_{0}c1^{0}_{0} \quad , \quad {}^{1}_{0}c1^{0}_{0} \quad , \quad {}^{0}_{0}c2^{0}_{0} \quad \textit{and} \quad {}^{1}_{0}c2^{0}_{0}\,; $$

which respectively give:

$$ \left(\left(\begin{smallmatrix}1 & 2\\ 2 & 2\end{smallmatrix}\right), (1,-1,1)\right), \ \left(\left(\begin{smallmatrix}3 & 2\\ 2 & 2\end{smallmatrix}\right), (0,-1,1)\right), \ \left(\left(\begin{smallmatrix}1 & 3\\ 3 & 3\end{smallmatrix}\right), (1,0,0)\right) \ \textit{and} \ \left(\left(\begin{smallmatrix}3 & 3\\ 3 & 3\end{smallmatrix}\right), (0,0,0)\right). $$

One immediately checks that the pairs of images obtained in Example 33 belong to these equivalence classes.

4.6.3 The automaton of imbalances

As a consequence, we can factorize the graph $\mathtt{G}_2$ by $\mathcal{R}_{pab}$. The quotient graph we obtain is the graph $\mathtt{G}$ described in Section 3.2, that we call *automaton of imbalances*.

Its set of vertices is $\mathtt{V}$; its edges map vertices to their respective images by allowed substitute and cut operations. Among its vertices, we are interested in the subclasses:

- $\mathtt{I} := \mathtt{I}_2/\mathcal{R}_{pab} = \mathtt{V} \cap \left\{\{(\left(\begin{smallmatrix}\epsilon & \epsilon\\ \epsilon & \epsilon\end{smallmatrix}\right), (0)_{a\in A})\} \cup \bigcup_{a\in A}\{(\left(\begin{smallmatrix}a & \epsilon\\ \epsilon & \epsilon\end{smallmatrix}\right), \mathrm{ab}(a)); (\left(\begin{smallmatrix}\epsilon & \epsilon\\ a & \epsilon\end{smallmatrix}\right), -\mathrm{ab}(a)); (\left(\begin{smallmatrix}a & \epsilon\\ a & \epsilon\end{smallmatrix}\right), (0)_{a\in A})\}\right\}$,
- $\mathtt{F} := \mathtt{F}_2/\mathcal{R}_{pab} = \bigcup_{w \text{ S-adic}} \{(M_{u,v}, \mathrm{ab}(u) - \mathrm{ab}(v)) \mid u, v \in \mathcal{F}(w) \text{ and } |u| = |v|\}$;

respectively called *initial* and *final* states of $\mathtt{G}$.

The quotient graph $\mathtt{G}$ inherits from the properties of accessibility of the graph $\mathtt{G}_2$.

Proposition 45. *Let* S *be a finite set of nonerasing substitutions defined over a common alphabet* A*.*

1. *If* $\mathtt{G}$ *is nonempty, then it has infinitely many vertices.*
2. *The number of outgoing edges from any vertex is bounded above by* $\mathrm{card}(S) \times l^4$*, where* $l = \max_{\sigma\in S, a\in A} |\sigma(a)|$.
3. *If each letter in* A *appears in a S-adic word (not necessarily the same), then any vertex in* $\mathtt{V}$ *is accessible from a vertex in* $\mathtt{I}$*.*
4. *If each letter in* A *appears in a S-adic word and if* X *is the image of an element in* $\mathtt{I}$ *by a finite composition of allowed substitute and cut operations, then* $X \in \mathtt{V}$*.*

Proof. Assertions (1) and (2) are proved in Proposition 13. One easily checks that Assertions (3) and (4) are inherited from similar properties in the graph of pairs of factors $\mathtt{G}_2$, namely Assertions (3) and (4) in Proposition 34. □

Therefore, similarly to $\mathtt{G}_2$, the graph $\mathtt{G}$ can be breadth first traversed. Algorithm 46 (below) is an easy adaptation of Algorithm 35; it yet requires the substitutions in S to be nonerasing (remind of Example 43).

Algorithm 46.
INPUTS:
- *a finite set of nonerasing substitutions S over a common alphabet A*
- *a nonnegative integer d*

ALGORITHM
$V \leftarrow$ `I` *(initialization)*
V' $\leftarrow$ empty set
for X in V:
    for $(\sigma, \delta_1, \delta_2, \delta_3, \delta_4)$ substitute and cut operation allowed on X:
        $Y \leftarrow$ image of X by the substitute and cut operation
        if Y has not yet appeared in the processing:
            $(M, (x_a)_{a\in A}) \leftarrow Y$
            if $\sum_{a\in A} x_a = 0$ and $\max_{a\in A} |x_a| \geq d$:
                return `true`
            otherwise:
                put Y in V'
$V \leftarrow$ *V'*
V' $\leftarrow$ empty set

Theorem A. *Let S denote a finite set of nonerasing substitutions over a common alphabet A, and assume that all letters in A appear in some S-adic word (not necessarily the same). If D_S denotes the quantity (possibly infinite):*

$$D_S = \sup_{w\in X_S} \operatorname{imb}(w),$$

then a breadth first search in the automaton of imbalances (viz. Algorithm 46), from its initial states, outputs, for any $d \leq D_S$, a finite sequence of substitutions $(\sigma_i)_{i\in\{1,\ldots,n\}} \in S^*$ such that the imbalance of any word in X_S, whose directive sequence starts with $(\sigma_i)_{i\in\{1,\ldots,n\}}$, is larger than d.

Proof. This comes from Corollary 36 and Remark 37 that hold for Algorithm 35. □

At last, observe that, like Algorithm 35, Algorithm 46 indefinitely keeps running for any instance (S, d) with $d > \sup_{w\in X_S} \operatorname{imb}(w)$.

5 Discussion on the implementation of the semi-algorithm

We saw that, as soon as it is nonempty, the automaton of imbalances has infinitely many vertices (Proposition 13). We implemented its breadth first traversal (Algorithm 46) for the S-adic systems associated with Arnoux-Rauzy (see Example-Definition 5) and Cassaigne-Selmer (see Example-Definition 4) substitutions. In both cases, the exponential growth of the spanning forest is an obstacle to the manifestation of non-trivial properties.

To slow it down, one needs to consider all the opportunities to spare calculations. In particular, we have to take advantage of all the symmetries enjoyed by the language of the system, and find tight sufficient conditions to kill as many unfruitful branches as possible in the trees. This approach is illustrated in Section 3.4 on the (easy) example of the Thue-Morse subshift. Despite such efforts (the symmetries and trim conditions are specific to the studied S-adic system), for interesting systems, the growth remains strong, as illustrated in Figures 2 and 3.

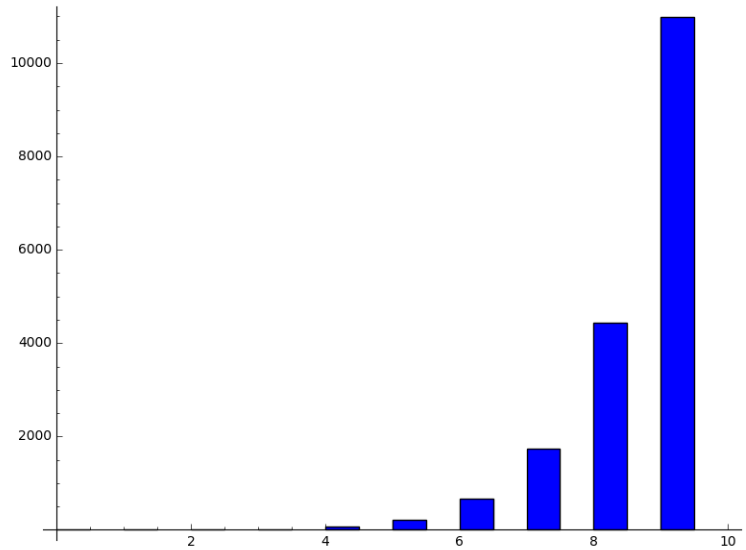


Figure 2: Number of vertices of the spanning forest of the Cassaigne-Selmer system, in function of exploration's depth. We already took advantage of symmetries.

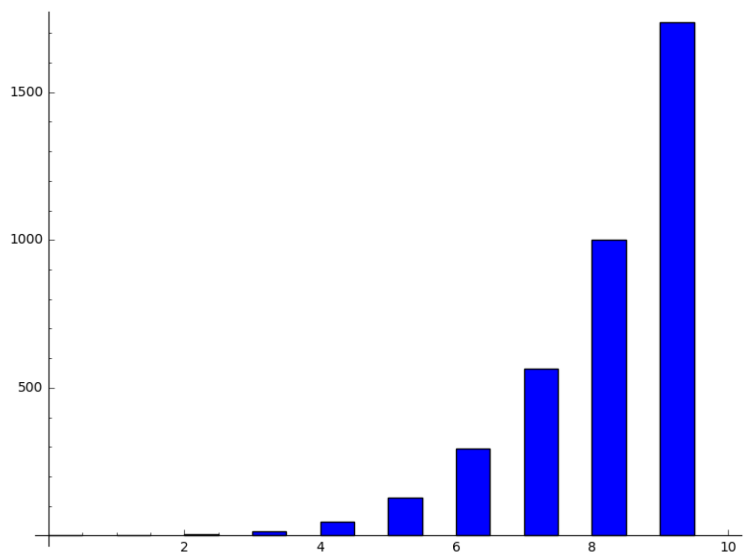


Figure 3: Number of vertices of the spanning forest of the Cassaigne-Selmer system, in function of exploration's depth. Here we take advantage of symmetries, plus an *ad hoc* trim condition.

Here are the results obtained with our final implementation of the automaton of imbalances for the Cassaigne-Selmer system:

- at depth 9, among 1 000 vertices, we found the first imbalance 3;
- at depth 16, among 80 000 vertices, we found the first imbalance 4; the computation (in `Python`) took 20 minutes.

Surprisingly, the study of the labels of the paths leading to these four first occurrences of imbalance turned to be sufficient to guess a general pattern. This pattern gave birth to the families of C-adic words with growing imbalances described in [And18].

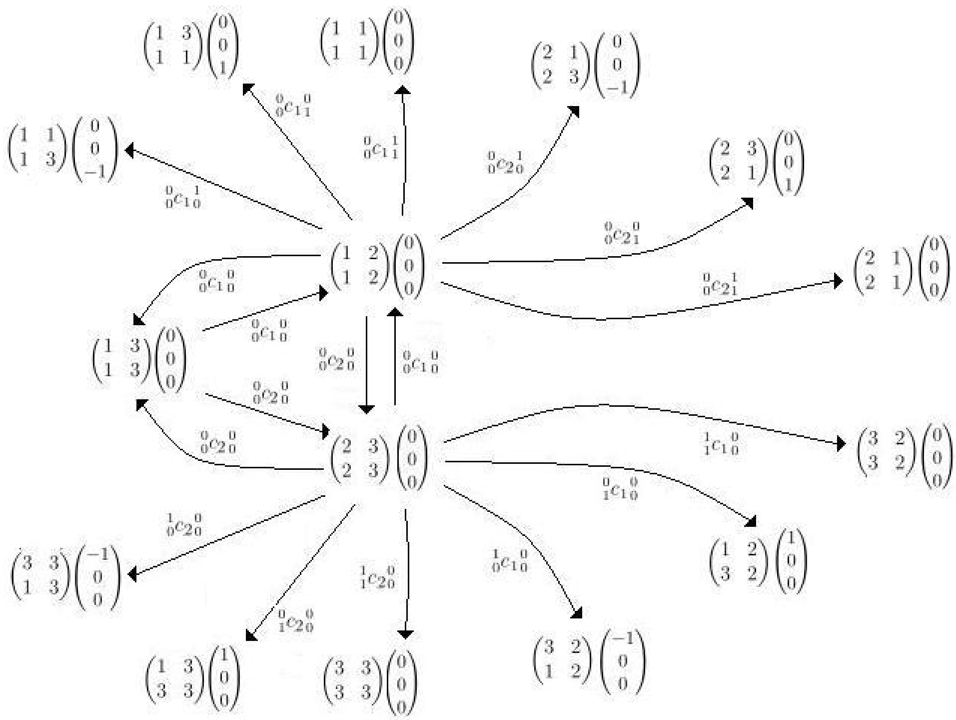


Figure 4: A portion of the automaton of imbalances for the Cassaigne-Selmer system.

References

[And18] Mélodie Andrieu. Autour du déséquilibre des mots C-adiques. *Journées Montoises d'informatique théorique (proceedings)*, pages 51–54, 2018.

[And21] Mélodie Andrieu. A Rauzy fractal unbounded in all directions of the plane. *Comptes Rendus de l'Académie des Sciences*, 359(4):399–407, 2021.

[AR91] Pierre Arnoux and Gérard Rauzy. Représentation géométrique de suites de complexité 2n+1. *Bulletin de la Société Mathématique de France*, 119:199–215, 1991.

[AS13] Pierre Arnoux and Štěpán Starosta. The Rauzy Gasket. In *Further Developments in Fractals and Related Fields*, pages 1–23. Springer, 2013.

[BD14] Valérie Berthé and Vincent Delecroix. Beyond substitutive dynamical systems: S-adic expansions. In *Lecture note 'Kokyuroku Bessatu'*, pages 81–123, 2014.

[Ber11] Valérie Berthé. Multidimensional Euclidean algorithms, numeration and substitutions. *Integers [electronic only]*, 2011.

[CFZ00] Julien Cassaigne, Sébastien Ferenczi, and Luca Q. Zamboni. Imbalances in Arnoux-Rauzy sequences. *Annales de l'Institut Fourier*, 50:1265–1276, 2000.

[CLL17] Julien Cassaigne, Sébastien Labbé, and Julien Leroy. A set of sequences of complexity 2n+1. In *WORDS 2017 Proceedings*, pages 144–156. Springer, 2017.

[CS01] Vincent Canterini and Anne Siegel. Automate des préfixes-suffixes associé à une substitution primitive. *Journal de théorie des nombres de Bordeaux*, 13(2):353–369, 2001.

[DHS13] Vincent Delecroix, Tomáš Hejda, and Wolfgang Steiner. Balancedness of Arnoux-Rauzy and Brun words. In Juhani Karhumäki, Arto Lepistö, and Luca Zamboni, editors, *Combinatorics on Words*, pages 119–131. Springer Berlin Heidelberg, 2013.

[Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[GLR09] Amy Glen, Florence Levé, and Gwénaël Richomme. Directive words of episturmian words: equivalences and normalization. *RAIRO - Theoretical Informatics and Applications*, 43(2):299–319, 2009.