



INSTITUT DE FRANCE  
Académie des sciences

# *Comptes Rendus*

---

## *Mathématique*

Mélodie Andrieu

**A Rauzy fractal unbounded in all directions of the plane**

Volume 359, issue 4 (2021), p. 399-407

Published online: 27 May 2021

<https://doi.org/10.5802/crmath.162>



This article is licensed under the  
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.  
<http://creativecommons.org/licenses/by/4.0/>



*Les Comptes Rendus. Mathématique* sont membres du  
Centre Mersenne pour l'édition scientifique ouverte  
[www.centre-mersenne.org](http://www.centre-mersenne.org)  
e-ISSN : 1778-3569



Combinatorics, Dynamical systems / *Combinatoire, Systèmes dynamiques*

# A Rauzy fractal unbounded in all directions of the plane

Mélodie Andrieu<sup>a</sup>

<sup>a</sup> Institut de Mathématiques de Marseille, I2M, Marseille, France

E-mail: [melodie.andrieu-estevez@univ-amu.fr](mailto:melodie.andrieu-estevez@univ-amu.fr)

**Abstract.** We construct an Arnoux–Rauzy word for which the set of all differences of two abelianized factors is equal to  $\mathbb{Z}^3$ . In particular, the imbalance of this word is infinite – and its Rauzy fractal is unbounded in all directions of the plane.

**Résumé.** Nous construisons explicitement un mot d’Arnoux–Rauzy pour lequel l’ensemble des différences possibles des facteurs abélianisés est égal à  $\mathbb{Z}^3$ . En particulier, le déséquilibre de ce mot est infini, et son fractal de Rauzy n’est borné dans aucune direction du plan.

*Manuscript received 16th September 2020, revised 2nd November 2020 and 5th November 2020, accepted 30th November 2020.*

## 1. Introduction

À l’algorithme de fraction continue soustractif décrit par l’itération de l’application (dite de Farey)

$$\begin{aligned} (\mathbb{R}^+)^2 &\rightarrow (\mathbb{R}^+)^2 \\ (x, y) &\mapsto (x - y, y) \quad \text{si } x \geq y, \\ &\quad (x, y - x) \quad \text{sinon,} \end{aligned}$$

est associée une classe particulière de mots infinis binaires appelés mots sturmiens. Rappelons qu’un *mot* est une suite finie ou infinie d’éléments (*lettres*) pris dans un ensemble fini (*alphabet*). Les mots sturmiens jouissent de nombreuses caractérisations combinatoires, arithmétiques et géométriques (consulter [7] pour une introduction générale). En particulier, ce sont exactement les mots aperiodiques binaires dont le déséquilibre vaut 1, c’est-à-dire dans lesquels tous les facteurs de même longueur (un *facteur* de longueur  $n$  est un sous-mot constitué de  $n$  lettres consécutives) contiennent, à 1 près, le même nombre de 0 (et donc, à 1 près également, le même nombre de 1). Par exemple, un mot commençant par  $w = 001000100100010001001 \dots$  pourrait être sturmien, tandis qu’un mot commençant par  $w = 011011100 \dots$  ne l’est pas, car il contient les facteurs 11 et 00. Cette propriété garantit en particulier que les lettres 0 et 1 sont uniformément distribuées par rapport à une mesure de probabilité  $\nu$  sur  $\{0, 1\}$ , et que l’écart entre la somme de Birkhoff

$$1/N \sum_{n=0}^{N-1} \mathbb{1}_{\{0\}}(w[n]),$$

qui mesure la fréquence de 0 observée parmi les  $N$  premières lettres du mot  $w$ , et sa valeur attendue  $\nu(0)$  (appelée fréquence de 0) est majoré par  $1/N$ . D'un point de vue géométrique, cela signifie que les points

$$P_N := \sum_{n=0}^N e_{w\{n\}},$$

où  $(e_0, e_1)$  désigne la base canonique de  $\mathbb{R}^2$ , restent à une distance bornée de la droite portée par le vecteur fréquence  $(\nu(0), \nu(1))$ . On appelle *ligne brisée* associée à  $w$  la suite  $(P_N)_{N \in \mathbb{N}}$ . En informatique, les lignes brisées associées aux mots sturmiens sont utilisées pour discrétiser les droites de pentes irrationnelles.

Depuis Jacobi, plusieurs algorithmes ont été proposés pour généraliser les fractions continues à des triplets de réels positifs (on peut consulter à ce sujet le livre [9]). De tels algorithmes devraient permettre d'approcher simultanément et efficacement deux réels par une suite de couples de nombres rationnels.

Dans ce document, nous nous intéressons aux mots d'Arnoux–Rauzy, introduits par Arnoux et Rauzy dans [2], qui sont les mots ternaires associés à l'algorithme (défini sur un ensemble de mesure nulle) :

$$F_{AR} : (\mathbb{R}^+)^3 \rightarrow (\mathbb{R}^+)^3$$

$(x, y, z) \mapsto (x - y - z, y, z)$	si $x \geq y + z$ ,
$(x, y - x - z, z)$	si $y \geq x + z$ ,
$(x, y, z - x - y)$	si $z \geq x + y$ .

Parce qu'ils conservent de nombreuses propriétés combinatoires des mots sturmiens, les mots d'Arnoux–Rauzy sont souvent présentés comme leur généralisation. En particulier, on peut montrer qu'ils admettent un vecteur fréquence des lettres. Aussi, une façon d'étudier la ligne brisée (tridimensionnelle) associée à un mot d'Arnoux–Rauzy consiste à la projeter, parallèlement au vecteur fréquence, sur le plan diagonal  $\Delta_0 : x + y + z = 0$ . On appelle *fractal de Rauzy de  $w$*  l'adhérence de cet ensemble de points.

Jusqu'en 2000, on a pensé que, comme pour les mots sturmiens, le déséquilibre des mots d'Arnoux–Rauzy était borné, ou au moins fini. Cassaigne, Ferenczi et Zamboni [4] ont contredit cette conjecture en construisant un mot d'Arnoux–Rauzy de déséquilibre infini – un mot donc, dont la ligne brisée s'écarte régulièrement et de plus en plus loin de sa direction moyenne, ou, dit encore autrement, un mot dont le fractal de Rauzy n'est pas borné.

Aujourd'hui, on ne sait presque rien sur les propriétés géométriques et topologiques de ces fractals de Rauzy déséquilibrés. Le théorème d'Oseledets [8] suggère toutefois que ces fractals sont contenus dans une bande du plan; en effet, si les exposants de Lyapounov associés au produit de matrices donné par l'algorithme existent, l'un de ces exposants au moins doit être négatif puisque leur somme est nulle.

Dans cette note, nous prouvons que cette intuition est fautive.

**Théorème 1.** *Il existe un mot d'Arnoux–Rauzy dont le fractal de Rauzy n'est borné dans aucune direction du plan.*

La construction que nous présentons s'adapte à la classe des mots associée à l'algorithme de fraction continue multidimensionnelle de Cassaigne–Selmer, introduite dans [5], ainsi qu'aux mots épisturmiens stricts, qui sont la généralisation des mots d'Arnoux–Rauzy. Rappelons qu'un mot sur un alphabet contenant  $d$  lettres est *épisturmien strict* si son langage est clos par miroir et s'il admet, pour chaque longueur  $n$ , un unique facteur multi-prolongeable à droite, et si ce facteur peut de plus être prolongé par chacune des  $d$  lettres de l'alphabet.

**Théorème 1’.** *Il existe un mot  $C$ -adique  $w_\infty$  dont le fractal de Rauzy n’est borné dans aucune direction du plan.*

**Théorème 1’.** *Soit  $d \geq 3$ . Il existe un mot épisturmien strict  $w_\infty$  sur l’alphabet  $\{1, \dots, d\}$  tel que pour tout hyperplan  $\mathcal{H}$  de  $\mathbb{R}^d$ , la distance des points de la ligne brisée  $(ab(p_n(w_\infty)))_{n \in \mathbb{N}}$  à l’hyperplan  $\mathcal{H}$  n’est pas bornée.*

Les démonstrations des Théorèmes 1’ et 1’’ reposent sur des techniques similaires à celles du Théorème 1, et sont intégralement rédigées dans [1].

Par ailleurs, nous proposons une preuve élémentaire du :

**Théorème 2.** *Le vecteur fréquence des lettres d’un mot d’Arnoux–Rauzy a des coordonnées rationnellement indépendantes.*

Ce résultat, conjecturé par Arnoux et Starosta en 2013 [3], a été démontré très récemment par des moyens plus sophistiqués par Dynnikov, Hubert et Skripchenko [6].

Le Théorème 2 est en fait vrai en toute dimension (voir [1] pour une preuve complète) :

**Théorème 2’.** *Soit  $d \geq 2$ . Le vecteur fréquence des lettres d’un mot épisturmien strict sur l’alphabet  $\{1, \dots, d\}$  a des coordonnées rationnellement indépendantes.*

## 1. Introduction (short English version)

Until 2000, it was believed that, as for Sturmian words, the imbalance of Arnoux–Rauzy words was bounded – or at least finite. Cassaigne, Ferenczi and Zamboni disproved this conjecture by constructing an Arnoux–Rauzy word with infinite imbalance, i.e. a word whose broken line deviates regularly and further and further from its average direction [4]. Today, we know virtually nothing about the geometrical and topological properties of these unbalanced Rauzy fractals. The Oseledets theorem suggests that these fractals are contained in a strip of the plane: indeed, if the Lyapunov exponents of the matricial product associated with the word exist, one of these exponents at least is nonpositive since their sum equals zero. This article aims at disproving this belief.

**Theorem 1.** *There exists an Arnoux–Rauzy word whose Rauzy fractal is unbounded in all directions of the plane.*

Theorem 1 also holds, on one hand, for  $C$ -adic words, which are the infinite words over  $\{1, 2, 3\}$  associated with the Cassaigne–Selmer multidimensional continued fraction algorithm introduced in [5] and, on the other hand, for strict episturmian words, which are the generalization of Arnoux–Rauzy words. We recall that a strict episturmian word is a word whose language is closed by mirror and which admits, for each length, a unique right-special factor – which is moreover prolonged by each letter in the alphabet.

**Theorem 1’.** *There exists  $w_\infty$  a  $C$ -adic word whose Rauzy fractal is unbounded in all directions of the plane.*

**Theorem 1’.** *Let  $d \geq 3$ . There exists  $w_\infty$  a strict episturmian word over the alphabet  $\{1, \dots, d\}$  such that for any hyperplane  $\mathcal{H}$  in  $\mathbb{R}^d$ , the distance between  $\mathcal{H}$  and the broken line  $(ab(p_n(w_\infty)))_{n \in \mathbb{N}}$  is unbounded.*

The proofs of Theorems 1’ and 1’’ are based on techniques similar to those of Theorem 1; they can be found in [1].

Besides, we propose an elementary proof of:

**Theorem 2.** *The vector of letter frequencies of any Arnoux–Rauzy word has rationally independent entries.*

This theorem completes the works of Arnoux and Starosta, who conjectured it in 2013, to prove that the Arnoux–Rauzy continued fraction algorithm detects all kind of rational dependencies [3]. Note that it has been recently proved by Dynnikov, Hubert and Skripchenko using quadratic forms [6].

Again, with a similar proof (see [1]), this result holds in arbitrary dimension:

**Theorem 2’.** *Let  $d \geq 2$ . The vector of letter frequencies of any strict episturmian word over  $\{1, \dots, d\}$  has rationally independent entries.*

## 2. Preliminaries

We denote by  $\mathfrak{A}^*$  the set of all finite words over an alphabet  $\mathfrak{A}$ . A finite word  $u = u[0]u[1] \dots u[n-1]$ , where  $u[k]$  denotes the  $(k+1)^{\text{th}}$  letter of  $u$ , is a *factor of length  $n$*  of a (finite or infinite) word  $w$  if there exists a nonnegative integer  $i$  such that for all  $k \in \{0, \dots, n-1\}$ ,  $w[i+k] = u[k]$ ; in the particular case  $i = 0$ , we say that  $u$  is the *prefix of length  $n$*  of  $w$ , and denote it by  $u = p_n(w)$ . We denote by  $\mathcal{F}_n(w)$  the set of factors of  $w$  of length  $n$  and by  $\mathcal{F}(w)$  its set of factors of all lengths.

A *substitution* is an application mapping letters to finite words:  $\mathfrak{A} \mapsto \mathfrak{A}^*$ , that we extend into a morphism on the free monoid for the concatenation operation  $\mathfrak{A}^*$  on one hand, and on the set of infinite words  $\mathfrak{A}^{\mathbb{N}}$  on the other hand. Three substitutions will be of high interest in this paper:  $\sigma_1, \sigma_2$  and  $\sigma_3$  defined over  $A = \{1, 2, 3\}$  by:

$$\begin{aligned} \sigma_i : A &\rightarrow A^* \\ i &\mapsto i \\ j &\mapsto ij \quad \text{for } j \in A \setminus \{i\}. \end{aligned}$$

They are called *Arnoux–Rauzy substitutions*; we denote  $AR = \{\sigma_1, \sigma_2, \sigma_3\}$ . The set  $AR$  can be seen as a three letter alphabet -it should not be confused with  $A = \{1, 2, 3\}$  over which the substitutions are defined. As much as we can, we refer to the elements of  $AR^*$  or  $AR^{\mathbb{N}}$  as “sequences” instead of “words”; nonetheless, some tools like the notions of factor and prefix will turn out to be useful for this second alphabet as well, especially in Section 3.

The set  $\mathfrak{A}^{\mathbb{N}}$  of infinite words over  $\mathfrak{A}$  is endowed with the distance  $\delta$ : for all  $w, w' \in \mathfrak{A}^{\mathbb{N}}$ ,  $\delta(w, w') = 2^{-n_0}$ , where  $n_0 = \min\{n \in \mathbb{N} \mid w[n] \neq w'[n]\}$  if  $w \neq w'$ , and  $\delta(w, w') = 0$  otherwise. We say that a sequence of finite words  $(u_n)_{n \in \mathbb{N}} \in (\mathfrak{A}^*)^{\mathbb{N}}$  *converges* to an infinite word  $w \in \mathfrak{A}^{\mathbb{N}}$  if for any sequence of infinite words  $(v_n)_{n \in \mathbb{N}} \in (\mathfrak{A}^{\mathbb{N}})^{\mathbb{N}}$ , the sequence of infinite words  $(u_n \cdot v_n)_{n \in \mathbb{N}} \in (\mathfrak{A}^{\mathbb{N}})^{\mathbb{N}}$  converges to  $w$ .

If  $(s_n)_{n \in \mathbb{N}} \in AR^{\mathbb{N}}$  is a sequence containing infinitely many occurrences of each Arnoux–Rauzy substitution  $\sigma_1, \sigma_2$  and  $\sigma_3$ , then the sequence of finite words  $(s_0 \circ \dots \circ s_{n-1}(\alpha))$ , with  $\alpha \in A$ , converges to an infinite word  $w_0$  which does not depend on  $\alpha$ . The infinite words  $w_0$  obtained this way are called *standard Arnoux–Rauzy words*. An infinite word  $w$  is an *Arnoux–Rauzy word* if it has the same set of factors than a standard Arnoux–Rauzy word  $w_0$ . One can show that the standard Arnoux–Rauzy word  $w_0$  and the *directive sequence*  $(s_n)_{n \in \mathbb{N}}$  associated with  $w$  are unique. This definition of Arnoux–Rauzy words is equivalent to the more usual one: an infinite word is an *Arnoux–Rauzy word* if it has complexity  $2n+1$  and admits exactly one right and one left special factor of each length.

Given a finite word  $u \in \mathfrak{A}^*$  and a letter  $\alpha \in \mathfrak{A}$ , we denote by  $|u|_\alpha$  the number of occurrences of  $\alpha$  in  $u$ . The *abelianized vector* of  $u$ , sometimes called *Parikh vector* of  $u$ , is the vector  $\text{ab}(u) = (|u|_\alpha)_{\alpha \in \mathfrak{A}}$ , which counts the number of times that each letter occurs in the finite word  $u$ . At this point, it is useful to order the alphabet. For the convenience of typing, we choose to

represent abelianized words as line vectors. Observe that the sum of the entries of  $\text{ab}(u)$  is equal to the *length* of the word  $u$ , that we denote by  $|u|$ . Now, given a substitution  $s : \mathfrak{A} \rightarrow \mathfrak{A}^*$ , the *incidence matrix* of  $s$  is the matrix  $M_s$  whose  $i^{\text{th}}$  row is the abelianized of the image by  $s$  of the  $i^{\text{th}}$  letter in the alphabet. For instance, the incidence matrices of the Arnoux–Rauzy substitutions are:

$$M_{\sigma_1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad M_{\sigma_2} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad M_{\sigma_3} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \in GL_3(\mathbb{Z}).$$

Abelianized words and incidence matrices are made to satisfy:  $\text{ab}(s(u)) = \text{ab}(u)M_s$  for any substitution  $s : \mathfrak{A} \rightarrow \mathfrak{A}^*$  and any finite word  $u \in \mathfrak{A}^*$ .

If  $w \in \mathfrak{A}^{\mathbb{N}}$  is an infinite word and  $\alpha \in \mathfrak{A}$  is a letter, the frequency of  $\alpha$  in  $w$  is the limit, if it exists, of the proportion of  $\alpha$  in the sequence of growing prefixes of  $w$ :

$$f_w(\alpha) = \lim_{n \rightarrow \infty} \frac{|p_n(w)|_\alpha}{n}.$$

We denote by  $f_w = (f_w(\alpha))_{\alpha \in \mathfrak{A}}$  the *vector of letter frequencies* of  $w$ , if it exists. When the vector of letter frequencies exists, as it is the case for any Arnoux–Rauzy word, it is natural to study the difference between the predicted frequencies of letters and their observed occurrences. Given an infinite word  $w \in \mathfrak{A}^{\mathbb{N}}$  for which the vector of letter frequencies is defined, we consider the *discrepancy function*:

$$\begin{aligned} &\mathbb{N} \rightarrow \mathbb{R} \\ &n \mapsto \max_{\alpha \in \mathfrak{A}} \left| |p_n(w)|_\alpha - n f_w(\alpha) \right|. \end{aligned}$$

The discrepancy is linked to a combinatorial property: the imbalance. The imbalance of an infinite word  $w$  is the quantity (possibly infinite) :

$$\text{imb}(w) = \sup_{n \in \mathbb{N}} \sup_{u, v \in \mathcal{F}_n(w)} \|\text{ab}(u) - \text{ab}(v)\|_\infty.$$

The imbalance of an infinite word  $w$  is finite if and only if its discrepancy function is bounded. Geometrically, the discrepancy is linked to the diameter of the Rauzy fractal. Let  $\Delta_0$  denotes the plane of  $\mathbb{R}^3$  with equation  $x + y + z = 0$ . For  $w$  an Arnoux–Rauzy word, denote by  $f_w$  its letter frequencies vector and by  $\pi_w$  the (oblique) projection onto  $\Delta_0$  associated with the direct sum:  $\mathbb{R}f_w \oplus \Delta_0 = \mathbb{R}^3$ . The *Rauzy fractal* of  $w$ , denoted by  $\mathcal{R}_w$ , is the closure of the image of the set of abelianized prefixes of  $w$  (the *broken line of  $w$* ) by the projection  $\pi_w$ :

$$\mathcal{R}_w = \overline{\cup_{k \in \mathbb{N}} \{ \pi_w(\text{ab}(p_k(w))) \}} \subset \Delta_0.$$

Note that the statement of our main result (Theorem 1) does not depend on the choice of the plane we project onto.

### 3. Results

**Lemma 3.** *For any  $(a, b, c) \in \mathbb{Z}^3$ , there exists  $s \in AR^*$  and there exist  $u, v \in \mathcal{F}(s(1))$  that satisfy  $\text{ab}(u) - \text{ab}(v) = (a, b, c)$ .*

**Remark 4 (Abuse of notation).** If  $s = s_0 \cdot \dots \cdot s_{n-1} \in AR^*$ , and if  $w \in A^* \cup A^{\mathbb{N}}$ , then  $s(w)$  denotes the image of the word  $w$  by the substitution  $s_0 \circ \dots \circ s_{n-1}$ .

**Proof.** Section 4 is devoted to the proof of Lemma 3. □

Therefore, all standard Arnoux–Rauzy words –and thereby all Arnoux–Rauzy words– whose directive sequence starts with the prefix  $s$  will admit  $(a, b, c)$  as difference of abelianized factors.

**Lemma 5.** For any  $p \in AR^*$  and any  $(a, b, c) \in \mathbb{Z}^3$ , there exists  $s \in AR^*$  and there exist  $u, v \in \mathcal{F}(p \cdot s(1))$  that satisfy  $\text{ab}(u) - \text{ab}(v) = (a, b, c)$ .

**Proof.** Let  $p \in AR^*$  and  $(a, b, c) \in \mathbb{Z}^3$ . Denote by  $M_p$  the incidence matrix of the substitution associated with  $p$  (following Remark 4), which is a product of the Arnoux–Rauzy matrices  $M_{\sigma_1}, M_{\sigma_2}$  and  $M_{\sigma_3}$ , and thus belongs to  $GL_3(\mathbb{Z})$ . By Lemma 3, there exists  $s \in AR^*$  and there exist  $u$  and  $v \in \mathcal{F}(s(1))$  such that  $\text{ab}(u) - \text{ab}(v) = (a, b, c) M_p^{-1}$ . But then,  $p(u)$  and  $p(v)$  are factors of  $\mathcal{F}(p \cdot s(1))$  and satisfy  $\text{ab}(p(u)) - \text{ab}(p(v)) = (\text{ab}(u) - \text{ab}(v)) M_p = (a, b, c)$ .  $\square$

We now construct a standard Arnoux–Rauzy word for which all triplets of integers can be obtained as a difference of two of its abelianized factors.

**Proposition 6.** There exists an Arnoux–Rauzy word  $w_\infty$  such that for all  $(a, b, c) \in \mathbb{Z}^3$ , there exist  $u$  and  $v \in \mathcal{F}(w_\infty)$  satisfying  $\text{ab}(u) - \text{ab}(v) = (a, b, c)$ .

**Proof.** Let  $\varphi : \mathbb{N} \rightarrow \mathbb{Z}^3$  a bijection (that can be chosen explicitly). We construct an infinite word  $d \in AR^\mathbb{N}$  as the limit of the sequence of finite words  $(p_k)_{k \in \mathbb{N}} \in (AR^*)^\mathbb{N}$  that we define by recurrence as follows. We first set  $p_0$  as the prefix given by Lemma 3 for  $(a, b, c) = \varphi(0)$ . Now, for  $k \in \mathbb{N}$ , we set  $p_{k+1} = p_k \cdot \sigma_1 \cdot \sigma_2 \cdot \sigma_3 \cdot s$ , where  $s \in AR^*$  is given by applying Lemma 5 to the word  $p_k \cdot \sigma_1 \cdot \sigma_2 \cdot \sigma_3 \in AR^*$  and the vector  $\varphi(k+1) \in \mathbb{Z}^3$ . By construction, the sequence of finite words  $(p_k)_{k \in \mathbb{N}}$  converges to an infinite sequence  $d$  which contains infinitely many occurrences of  $\sigma_1, \sigma_2$  and  $\sigma_3$ . This guarantees that the sequence of finite words  $(d_0 \circ \dots \circ d_{n-1}(1))_{n \in \mathbb{N}}$  converges to an Arnoux–Rauzy word, that we denote by  $w_\infty$ . Finally, for any  $k \in \mathbb{N}$ , since the directive sequence of  $w_\infty$  starts with the prefix  $p_k$ , there exist  $u_k, v_k \in \mathcal{F}(w_\infty)$  such that  $\text{ab}(u_k) - \text{ab}(v_k) = \varphi(k)$ .  $\square$

**Corollary 7.** The imbalance of the word  $w_\infty$  is infinite.

**Proof.** For any  $n \in \mathbb{N}$ , there exist  $u_n$  and  $v_n \in \mathcal{F}(w_\infty)$  such that  $\text{ab}(u_n) - \text{ab}(v_n) = (n, 0, -n)$ ; this implies both  $|u_n| = |v_n|$  and  $|u_n|_1 - |v_n|_1 = n$ . The imbalance of  $w_\infty$  is thus infinite.  $\square$

The imbalance of a word, which is a combinatorial quantity, is linked to the geometrical shape of its associated broken line. More precisely: a word  $w$  admitting frequencies has an infinite imbalance if and only if its Rauzy fractal is unbounded. We now propose to show that the word  $w_\infty$  actually satisfies a stronger property: its Rauzy fractal is unbounded in *all directions* of the plane. This relies on the following proposition.

**Proposition 8.** Let  $w \in A^\mathbb{N}$ . If for all  $\mathbf{d} \in \mathbb{Z}^3 \cap \Delta_0$ , where  $\Delta_0$  denotes the plane of  $\mathbb{R}^3$  with equation  $x + y + z = 0$ , there exist  $u$  and  $v \in \mathcal{F}(w)$  such that  $\text{ab}(u) - \text{ab}(v) = \mathbf{d}$ , then, for any plane  $\Pi$  and for any  $D \in \mathbb{R}^+$ , there exists  $k \in \mathbb{N}$  such that the euclidean distance between the point  $\text{ab}(p_k(w))$  and the plane  $\Pi$  is larger than  $D$ .

**Proof.** Without loss of generality, we can assume that  $\Pi$  contains  $(0, 0, 0)$ .

If  $\Pi = \Delta_0$ , then for any  $D \in \mathbb{R}^+$ ,  $\text{dist}(\text{ab}(p_k(w)), \Pi) > D$ , with  $k = \lfloor D\sqrt{3}/3 \rfloor + 1$ .

Let  $\Pi \neq \Delta_0$ . By contradiction, assume that there exists  $D \in \mathbb{R}^+$  such that for all nonnegative integer  $k$ ,  $\text{dist}(\text{ab}(p_k(w)), \Pi) \leq D$ . Let  $\mathbf{d} \in \mathbb{Z}^3 \cap \Delta_0$  with  $\text{dist}(\mathbf{d}, \Pi) > 4D$ , and factors  $u, v \in \mathcal{F}(w)$  such that  $\text{ab}(u) - \text{ab}(v) = \mathbf{d}$ . Then, without loss of generality, we have  $\text{dist}(\text{ab}(u), \Pi) > 2D$ . Let  $t \in A^*$  be such that  $tu$  is a prefix of  $w$ . Then we have  $\text{dist}(\text{ab}(t), \Pi) > D$  or  $\text{dist}(\text{ab}(tu), \Pi) > D$ , a contradiction.  $\square$

**Remark 9.** Proposition 8 and its proof remain valid by replacing  $\Delta_0$  by any other plane whose intersection with  $\mathbb{Z}^3$  is not trapped between two parallel lines.

**Theorem 1.** There exists an Arnoux–Rauzy word whose Rauzy fractal is unbounded in all directions of the plane.

**Proof.** We obtain, by applying Proposition 8 to the word  $w_\infty$  described in Proposition 6 and to planes spanned by  $f_w$  and a vector of  $\Delta_0$ , that the Rauzy fractal associated with  $w_\infty$  cannot be trapped between two parallel lines.  $\square$

**4. Proof of Lemma 3**

We consider the infinite oriented graph whose vertices are the elements of  $\mathbb{Z}^3$  and whose edges map triplets to their images by one the 15 following applications. For  $\delta \in \{-2, -1, 0, 1, 2\}$  and  $i \in \{1, 2, 3\}$ , consider:

$$\begin{aligned} \tau_{i,\delta} : \mathbb{Z}^3 &\rightarrow \mathbb{Z}^3 \\ (x_j)_{j \in \{1,2,3\}} &\mapsto (y_j)_{j \in \{1,2,3\}} \quad \text{where } y_i = x_1 + x_2 + x_3 + \delta \text{ and } y_j = x_j \text{ for } j \neq i. \end{aligned}$$

Our aim is to show that all vertices can be reached from the triplet  $O = (0, 0, 0) \in \mathbb{Z}^3$ , moving through a finite number of edges (see Definition 11 and Proposition 12 below.) The motivation lies in the following lemma.

**Lemma 10.** *Let  $d \in \mathbb{Z}^3$ . If there exist  $n \in \mathbb{N}$  and a finite sequence  $(i_l, \delta_l)_{0 \leq l \leq n-1} \in (\{1, 2, 3\} \times \{-2, -1, 0, 1, 2\})^n$  such that  $d = \tau_{i_{n-1}, \delta_{n-1}} \circ \dots \circ \tau_{i_0, \delta_0}(O)$ , then there exist  $s_1 \in AR^*$  and  $u, v \in \mathcal{F}(\sigma_{i_{n-1}} \circ \dots \circ \sigma_{i_0}(s_1(1)))$  satisfying  $\text{ab}(u) - \text{ab}(v) = d$ .*

**Proof.** Let  $d \in \mathbb{Z}^3$ . Assume that there exist  $n \in \mathbb{N}$  and  $(i_l, \delta_l)_{0 \leq l \leq n-1} \in (\{1, 2, 3\} \times \{-2, -1, 0, 1, 2\})^n$  such that  $d = \tau_{i_{n-1}, \delta_{n-1}} \circ \dots \circ \tau_{i_0, \delta_0}(O)$ . We are going to build iteratively two finite sequences of finite words  $(u_l)$  and  $(v_l)$ , where  $l \in \{0, \dots, n\}$ , and  $s_1 \in AR^*$ , such that for all  $l$ , the words  $u_l$  and  $v_l$  are factors of  $\sigma_{i_{l-1}} \circ \dots \circ \sigma_{i_0}(s_1(1))$ , and such that  $\text{ab}(u_n) - \text{ab}(v_n) = d$ .

First, we choose  $s_1 \in AR^*$  that satisfies  $|s_1(1)| \geq 2n$ , and we set  $u_0 = v_0 = p_n(s_1(1))$  (prefix of length  $n$  of  $s_1(1)$ ). Then, assuming that  $u_l$  and  $v_l \in \mathcal{F}(\sigma_{i_{l-1}} \circ \dots \circ \sigma_{i_0}(s_1(1)))$  are built, we set  $\tilde{u}_{l+1} = \sigma_{i_l}(u_l)$  and  $\tilde{v}_{l+1} = \sigma_{i_l}(v_l)$ . From  $\tilde{u}_{l+1}$  and  $\tilde{v}_{l+1}$ , we now define  $u_{l+1}$  and  $v_{l+1}$  according to the following table.

$\delta$	choice for $u_{l+1}$	choice for $v_{l+1}$
0	$\tilde{u}_{l+1}$	$\tilde{v}_{l+1}$
1	$\tilde{u}_{l+1} \cdot i_l$	$\tilde{v}_{l+1}$
2	$\tilde{u}_{l+1} \cdot i_l$	$v_{l+1}$ such that $i_l \cdot v_{l+1} = \tilde{v}_{l+1} (*)$
-1	$\tilde{u}_{l+1}$	$\tilde{v}_{l+1} \cdot i_l$
-2	$u_{l+1}$ such that $i_l \cdot u_{l+1} = \tilde{u}_{l+1} (*)$	$\tilde{v}_{l+1} \cdot i_l$

The steps marked with (\*) (removal of the initial  $i_l$ ) are always possible since, for all  $l \in \{0, \dots, n-1\}$ , the words  $u_l$  and  $v_l$  (and thus  $\tilde{u}_{l+1}$  and  $\tilde{v}_{l+1}$ ) are nonempty.

In all of these cases, the words  $u_{l+1}$  and  $v_{l+1}$  are factors of  $\sigma_l \circ \dots \circ \sigma_0(s_1(1))$  and satisfy  $\text{ab}(u_{l+1}) - \text{ab}(v_{l+1}) = \tau_{i_l, \delta_l}(\text{ab}(u_l) - \text{ab}(v_l))$ . In particular, at step  $l = n-1$ , the finite words  $u_n$  and  $v_n$  are factors of  $\sigma_{i_{n-1}} \circ \dots \circ \sigma_{i_0}(s_1(1))$  and satisfy  $\text{ab}(u_n) - \text{ab}(v_n) = \tau_{i_{n-1}, \delta_{n-1}} \circ \dots \circ \tau_{i_0, \delta_0}(O) = d$ .  $\square$

In the sequel, it is convenient to introduce some vocabulary from graph theory.

**Definition 11.** *A triplet  $(a, b, c) \in \mathbb{Z}^3$  is accessible from a triplet  $(d, e, f)$  if there exist a nonnegative integer  $n$  and a finite sequence  $(i_l, \delta_l)_{0 \leq l \leq n-1} \in (\{1, 2, 3\} \times \{-2, -1, 0, 1, 2\})^n$  such that  $(a, b, c) = \tau_{i_{n-1}, \delta_{n-1}} \circ \dots \circ \tau_{i_0, \delta_0}((d, e, f))$ .*

**Proposition 12.** *All triplets in  $\mathbb{Z}^3$  are accessible from  $O$ .*

The proof of Proposition 12 lies on the Lemmas 13 and 14.

**Lemma 13.** *The triplet  $(a, b, c) \in \mathbb{Z}^3$  is accessible from  $O$  if and only if  $(-a, -b, -c)$  is also accessible from  $O$ . Similarly,  $(x_j)_{j \in \{1,2,3\}}$  is accessible from  $O$  if and only if for all  $s \in \mathfrak{S}_3$ , where  $\mathfrak{S}_3$  denotes the symmetric group acting on three elements, the triplet  $(x_{s(j)})_{j \in \{1,2,3\}}$  is accessible from  $O$ .*

**Proof.** For the first assertion, change  $\delta_l$  into  $-\delta_l$  in the finite sequence of edges going from  $O$  to  $(a, b, c)$ . For the second assertion, change  $i_l$  into  $s(i_l)$  in the finite sequence of edges going from  $O$  to  $(x_j)_{j \in \{1,2,3\}}$ .  $\square$



**Lemma 14.** *Let  $a \in \mathbb{N}$ . The triplet  $(a, -a, -a) \in \mathbb{Z}^3$  is accessible from  $O$ .*

**Proof.** The lemma is trivially true for  $a = 0$ . By recurrence, consider an arbitrary nonnegative integer  $a$  such that the triplet  $(a, -a, -a)$  is accessible from  $O$ . One can check that  $(a + 1, -a - 1, a + 1) = \tau_{1,1} \circ (\tau_{3,2})^{2a+1} \circ \tau_{2,-1}((a, -a, -a))$ . So the triplet  $(a + 1, -a - 1, a + 1)$  is accessible from  $O$ . But then, Lemma 13 indicates that  $(a + 1, -a - 1, -a - 1)$  is accessible from  $O$ .  $\square$

**Proof of Proposition 12.** The proof relies on the four following observations.

- The vertices  $(a, b, -a)$  and  $(a, -a, c)$  are accessible from  $O$  for all  $a, b, c \in \mathbb{Z}$ . Indeed, it suffices to write  $(a, b, -a) = (\tau_{2,0})^{a+b}((a, -a, -a))$  and  $(a, -a, c) = (\tau_{3,0})^{a+c}((a, -a, -a))$  and remember that  $(a, -a, -a)$  is accessible from  $O$  by Lemma 14.
- The vertex  $(a, c, c) = \tau_{2,0}((a, -a, c))$  is also accessible from  $O$ ,
- If  $a \geq b > c > -a$ , then  $(a, -a + b - c, c) = (\tau_{2,0})^{-1}(a, b, c)$  is closer to  $(a, -a, -a)$  in sup norm, and we have  $|-a + b - c|, |c| < a$ ,
- If  $a \geq c > b > -a$ , then  $(a, b, -a - b + c) = (\tau_{3,0})^{-1}((a, b, c))$  is closer to  $(a, -a, -a)$  than  $(a, b, c)$ , and we have  $|b|, |-a - b + c| < a$ .

Let  $a, b, c \in \mathbb{Z}^3$ . By Lemma 13, it suffices to deal with the case  $|b|, |c| \leq |a|$ , and  $a > 0$ . Following the observations above, we recursively construct a finite sequence  $(i_l)_{0 \leq l \leq n-1} \in \{2, 3\}^n$  such that  $(a, b, c) = \tau_{i_{n-1}, 0} \circ \dots \circ \tau_{i_0, 0}((a, -a, -a))$ . Since  $(a, -a, -a)$  is accessible from  $O$  (Lemma 14), the vertex  $(a, b, c)$  is also accessible from  $O$ .  $\square$

**Proof of Lemma 3.** Lemma 3 follows from Proposition 12, Definition 11 and Lemma 10.  $\square$

**Remark 15.** The graph  $\mathcal{G}$  is a simplification, exploiting the remarkable properties of the substitutions  $\sigma_1, \sigma_2$  and  $\sigma_3$ , of the imbalance automaton, introduced in [1] for a much wider range of S-adic systems (ie class of words obtained from a set of substitutions through *directive sequences*).

## 5. The vector of letter frequencies of $w_\infty$ has rationally independent entries

We sketch an elementary proof of the much wider result:

**Theorem 2.** *The vector of letter frequencies of any Arnoux–Rauzy word has rationally independent entries.*

The proof is inspired from a similar result that holds for C-adic words [5].

**Proof.** Let  $w$  an Arnoux–Rauzy word; denote by  $(s_n)_{n \in \mathbb{N}}$  its directive sequence and by  $f$  its letter frequencies vector. We recall that for all nonnegative integer  $n$ ,  $s_n = \sigma_i$  if and only if the  $i^{\text{th}}$  entry of  $F_{AR}^n(f)$  ( $F_{AR}$  is defined in Section 1) is greater than the sum of the two others. By contradiction, assume that the entries of  $f$  are not rationally independent.

First, observe that if for some  $r \in \mathbb{N}$ , the  $i^{\text{th}}$  entry of  $F_{AR}^r(f)$  is zero, then it will remain zero; and from this point on the directive sequence will not contain the substitution  $\sigma_i$ , which is conflicting with the definition of Arnoux–Rauzy words and the uniqueness of the directive sequence. Thus, for all  $n \in \mathbb{N}$ , all entries of  $F_{AR}^n(f)$  are positive. Let  $l_0$  a nonzero integer column vector such that  $f l_0 = 0$  (recall that  $f$  is a line vector). Let  $l_m = M_{s_{m-1}} \dots M_{s_0} l_0$ . The Arnoux–Rauzy matrices being invertible,  $l_m$  is also a nonzero integer column vector; it satisfies

$$F_{AR}^m(f) l_m = f \cdot M_{s_0}^{-1} \cdot M_{s_{m-1}}^{-1} l_m = f l_0 = 0.$$

Denote  $l_m = (a, b, c)^t$  and consider  $D_m = \max(|b - a|, |c - b|, |c - a|) \in \mathbb{N}$  the difference between the maximum and the minimum entry of  $l_m$ , that we call *spread* of  $l_m$ . We claim that the sequence of nonnegative integers  $(D_m)_{m \in \mathbb{N}}$  is non-increasing and that it furthermore decreases infinitely often - and here will be the contradiction.

Indeed, the vector  $l_{m+1}$  is of the form  $Ml_m$ , where  $M$  is one of the three Arnoux–Rauzy matrices  $M_{\sigma_1}$ ,  $M_{\sigma_2}$  or  $M_{\sigma_3}$ , which give respectively:  $l_{m+1} = (a, a+b, a+c)^t$ ,  $l_{m+1} = (a+b, b, c+b)^t$  and  $l_{m+1} = (a+c, b+c, c)^t$ . One can easily show, observing that the extreme entries of  $l_m$  have opposite signs, that in all cases  $D_m \geq D_{m+1}$ . Similarly, we write  $l_{m+2} = M_{s_{m+1}}M_{s_m}l_m$ . A quick argument shows that as soon as  $s_{m+1} \neq s_m$ , which happens infinitely many times by definition of Arnoux–Rauzy words, we have  $D_m > D_{m+2}$ .  $\square$

## References

- [1] M. Andrieu, “Exceptional trajectories in the symbolic dynamics of multidimensional continued fraction algorithms”, PhD Thesis, Institut de Mathématiques de Marseille, Marseille, France, 2021.
- [2] P. Arnoux, G. Rauzy, “Représentation géométrique de suites de complexité  $2n+1$ ”, *Bull. Soc. Math. Fr.* **119** (1991), no. 2, p. 199-215.
- [3] P. Arnoux, Š. Starosta, “The Rauzy Gasket”, in *Further Developments in Fractals and Related Fields. Mathematical foundations and connections*, Trends in Mathematics, Springer, 2013, p. 1-23.
- [4] J. Cassaigne, S. Ferenczi, L. Q. Zamboni, “Imbalances in Arnoux-Rauzy sequences”, *Ann. Inst. Fourier* **50** (2000), p. 1265-1276.
- [5] J. Cassaigne, S. Labbé, J. Leroy, “A Set of Sequences of Complexity  $2n+1$ ”, in *Combinatorics on words. 11th international conference, WORDS 2017 Proceedings*, Springer, 2017, p. 144-156.
- [6] I. Dynnikov, P. Hubert, A. Skripchenko, “Dynamical systems around the Rauzy gasket and their ergodic properties”, in preparation, <https://arxiv.org/abs/2011.15043>, 2020.
- [7] M. Lothaire, *Combinatorics on Words*, 2nd ed., Encyclopedia of Mathematics and Its Applications, vol. 17, Cambridge University Press, 1997, Foreword by Roger Lyndon.
- [8] V. I. Oseledets, “A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems”, *Trans. Mosc. Math. Soc.* **19** (1968), p. 197-231.
- [9] F. Schweiger, *Multidimensional Continued Fractions*, Oxford Science Publications, Oxford University Press, 2000.