Autour du déséquilibre des mots C-adiques

Mélodie Andrieu

28 mai 2018

Résumé

Nous étudions une propriété combinatoire, le déséquilibre, d'une classe particulière de mots sur l'alphabet $\{a, b, c\}$: les mots C-adiques. En particulier, nous exhibons des familles de mots C-adiques de déséquilibres arbitrairement grands, et même des mots C-adiques de déséquilibre infini. Ces constructions ont été obtenues par l'exploration d'un automate et l'étude de ses chemins.

1 Motivations

À l'algorithme de fraction continue soustractif décrit par l'itération de l'application

$$\begin{array}{lcll} (\mathbb{R}^+)^2 & \to & (\mathbb{R}^+)^2 & \\ (x,y) & \mapsto & (x-y,y) & \text{si } x \geq y \\ & & (x,y-x) & \text{sinon} \end{array}$$

est associée une classe particulière de mots infinis binaires : les mots sturmiens. Ceux-ci jouissent de deux caractérisations combinatoires : d'une part, ce sont exactement les mots de complexité $n+1$, c'est-à-dire les mots qui admettent $n+1$ facteurs de longueur n pour tout entier n ; d'autre part, ce sont les mots apériodiques dont le déséquilibre vaut 1, c'est-à-dire les mots apériodiques dans lesquels chaque lettre apparaît, à une unité près, un même nombre de fois dans tous les facteurs d'une longueur donnée.

Plusieurs tentatives ont été faites pour généraliser les fractions continues à des triplets de réels positifs. Un tel algorithme pourrait permettre d'approcher simultanément deux réels par une suite de couples de nombres rationnels.

Dans ce document, nous nous interrogeons sur les mots C-adiques, qui sont les mots ternaires associés à l'algorithme [3] :

$$\begin{array}{lcll} (\mathbb{R}^+)^3 & \to & (\mathbb{R}^+)^3 & \\ (x,y,z) & \mapsto & (x-z,z,y) & \text{si } x \geq z \\ & & (y,x,z-x) & \text{sinon.} \end{array}$$

Tout comme les mots d'Arnoux-Rauzy, ces mots sont de complexité $2n+1$. L'intérêt de cet algorithme est qu'il admet comme instance n'importe quel triplet de réels positifs, contrairement à l'algorithme d'Arnoux-Rauzy qui n'est défini que sur un ensemble de mesure de Lebesgue nulle.

Aussi, il est naturel de s'interroger sur l'existence d'une borne uniforme pour le déséquilibre. Hélas, comme pour les mots d'Arnoux-Rauzy [2], nous pouvons construire des familles de mots C-adiques de déséquilibre aussi grand que souhaité, et même, par un lemme de pompage, des mots de déséquilibre infini.

2 Déséquilibre, mots C-adiques

Soit u un mot fini sur l'alphabet ternaire $A = \{a, b, c\}$ et $\alpha \in A$ une lettre. On désigne par $|u|_\alpha$ le nombre d'occurrences de la lettre α dans le mot u. Le *vecteur de Parikh* de u est le vecteur $\chi(u) = (|u|_a, |u|_b, |u|_c)$, qui compte les multiplicités de chacune des lettres de l'alphabet. Remarquons que la somme des coordonnées de ce vecteur est égale à la longueur du mot u, que l'on note $|u|$. Étant donné deux mots de même longueur u et v, on appelle *vecteur de déséquilibre de u et v* la différence de leurs vecteurs de Parikh. La somme des coordonnées d'un tel vecteur est nulle. Le déséquilibre d'un mot infini w est la quantité (éventuellement infinie) :

$$d = \sup_{n\in\mathbb{N}} \sup_{u,v\in F_n(w)} ||\chi(u) - \chi(v)||_\infty,$$

qui s'écrit encore :

$$d = \sup_{n\in\mathbb{N}} \sup_{u,v\in F_n(w)} \max_{\alpha\in\{a,b,c\}} ||u|_\alpha - |v|_\alpha|.$$

Elle mesure les iniquités de répartition entre les lettres dans un mot donné.

Dans toute la suite, on s'intéressera aux substitutions $c_1 : a \mapsto a, b \mapsto ac, c \mapsto b$ et $c_2 : a \mapsto b, b \mapsto ac, c \mapsto c$, qui proviennent de l'algorithme de fraction continue 1 [3]. On notera aussi $C = \{c_1, c_2\}$.

Un *mot C-adique* est un mot infini de la forme $w = lim_{n\to\infty} s_1 \circ ... \circ s_n(w')$, où $(s_k)_{k\in\mathbb{N}} \in C^{\mathbb{N}}$ porte le nom de *suite directrice* de w, et où w' est un mot infini quelconque sur A ; avec la condition supplémentaire que chacune des substitutions c_1 et c_2 apparaisse une infinité de fois dans la suite directrice.

Remarque : pour S un ensemble de substitutions, on peut étudier les mots S-adiques dans un cadre plus général [1].

Enfin, pour $s \in C$, nous introduisons les applications ${}^{-}s$ et s^{-} qui, à un mot fini non vide u, associent le mot $s(u)$ auquel on efface la première (resp. la dernière) lettre. Ces applications ne sont pas des morphismes.

3 Construction de mots de C-adiques de déséquilibres arbitrairement grands

Pour tout entier n, nous souhaitons exhiber un mot C-adique de déséquilibre supérieur à n. Pour ce faire, nous allons construire par récurrence une suite $(w_n)_{n\in\mathbb{N}}$ de mots C-adiques, et donner sur chacun d'eux deux facteurs u_n et v_n de même longueur, dont le vecteur de déséquilibre est de norme n. Cela assure que w_n a pour déséquilibre au moins n. L'intuition de ces constructions provient de l'exploration d'un automate et de l'étude de ses chemins.

Construction de (w_n).

Soit w_0 n'importe quel mot C-adique, par exemple $c_1 \circ c_2 \circ c_1 \circ c_2 \circ ...(a)$. Posons $w_1 = c_2 \circ c_2 \circ c_2(w_0)$ et pour tout $n \geq 1$:

$$\begin{cases} w_{n+1} = c_1^{2n+2} \circ c_2(w_n) & \text{si } n \text{ est impair} \\ w_{n+1} = c_2^{2n+2} \circ c_1(w_n) & \text{sinon.} \end{cases}$$

Pour tout entier n, w_n est un mot C-adique.

Construction des suites de facteurs (u_n) et (v_n).

La lettre b apparaît dans w_0 (la complexité nous le garantit), donc $acc = c_2 \circ c_2 \circ c_2(b)$ apparaît dans w_1. Posons $u_1 = ac$, $v_1 = cc$ et, pour tout $n \geq 1$:

$$\begin{cases} u_{n+1} = c_1^- \circ (c_1 \circ c_1)^n \circ c_1 \circ c_2(u_n) & \text{si } n \text{ est impair} \\ v_{n+1} = c_1 \circ (c_1 \circ^- c_1)^n \circ c_1 \circ c_2(v_n) \end{cases}$$

$$\begin{cases} u_{n+1} = c_2 \circ (c_2 \circ c_2^-)^n \circ c_2 \circ c_1(u_n) & \text{sinon.} \\ v_{n+1} =^- c_2 \circ (c_2 \circ c_2)^n \circ c_2 \circ c_1(v_n) \end{cases}$$

Les mots u_n et v_n sont bien facteurs de w_n, pour chaque n.

Lemme 1. *Soit $n \geq 1$.*

— *Si n est impair, alors u_n commence par a et termine par c, v_n commence et termine par c, et $\chi(u_n) - \chi(v_n) = (n, 1-n, -1)$.*

— *Sinon, alors u_n commence et termine par a, v_n commence par a et termine par c, et $\chi(u_n) - \chi(v_n) = (1, n-1, -n)$.*

Theorème 1. *Il existe des mots C-adiques de déséquilibre arbitrairement grand.*

4 Construction de mots C-adiques de déséquilibre infini

Nous venons de construire une famille de mots C-adiques pour laquelle le déséquilibre n'est pas borné. Toutefois, le déséquilibre des mots pris individuellement peut l'être. Pour construire un mot dont le déséquilibre est infini, nous allons recourir à un lemme de pompage.

La première étape consiste à montrer que lorsque l'on compose un mot C-adique par l'une ou l'autre des substitutions c_1 et c_2, on ne rééquilibre pas le mot au-delà d'une certaine proportion.

Proposition 1. *Si w est un mot C-adique de déséquilibre $Des(w) \geq 3n$, alors $c_1(w)$ (resp. $c_2(w)$) est un mot C-adique de déséquilibre $Des(w) \geq n$.*

Pour tout entier naturel m, on note désormais $C^m := \{s_0 \circ ... \circ s_{m-1} \in \{c_1, c_2\}^m\}$, et $C^* := \cup_{m \in \mathbb{N}} C^m$.

Corollaire 1 (Lemme de pompage)**.** $\forall s \in C^*, \forall n \in \mathbb{N}, \exists \sigma \in C^*$ *tel que pour tout mot C-adique w, $Des(s \circ \sigma(w)) \geq n$.*

Theorème 2. *Il existe un mot C-adique de déséquilibre infini.*

Démonstration. D'après le corollaire 1, je peux construire une suite $(\sigma_k)_{k \in \mathbb{N}} \in (C^*)^{\mathbb{N}}$ telle que pour tout entier naturel n et pour tout mot C-adique w, le déséquilibre du mot $\sigma_0 \circ ... \circ \sigma_n(w)$ vaut au moins n. Ainsi, le mot limite $w_\infty = \lim_{n \to \infty} \sigma_0 \circ ... \sigma_n(w_0)$ (s'il existe, sinon prendre une valeur d'adhérence de la suite), où w_0 est un mot C-adique quelconque, est lui même un mot C-adique (chaque σ_k fourni par le corollaire contient c_1 et c_2), et son déséquilibre vaut au moins n, pour tout n ; il est donc de déséquilibre infini. □

Références

[1] Valérie Berthé, Wolfgang Steiner, and Jörg Thuswaldner. Geometry, dynamics, and arithmetic of s-adic shifts. *preprint https ://arxiv.org/abs/1410.0331*, 2014.

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

[3] Julien Cassaigne, Sébastien Labbé, and Julien Leroy. A set of sequences of complexity 2n+1. In *Combinatorics on Words - 11th International Conference, WORDS 2017, Montréal, QC, Canada, September 11-15, 2017, Proceedings*, pages 144–156.