

Etude des motifs et répétitions

Recherche de mots infinis denses en facteurs carrés distincts

Maher Mallem,
encadré par Srečko Brlek au LaCIM
Stage de L3 informatique, ENS Paris-Saclay

Août 2017

Table des matières

1	Définitions et notations	3
2	Le contexte	5
2.1	Contexte humain	5
2.1.1	Le laboratoire	5
2.1.2	Le déroulement du stage	5
2.2	Contexte scientifique	6
2.2.1	La recherche sur les carrés	6
2.2.2	Quelques différences notables entre carrés et palindromes	7
3	La contribution	8
3.1	Cas des suites strictement croissantes	8
3.1.1	Preuve d'absence de double carrés	8
3.1.2	Une condition suffisante pour la densité	10
3.1.3	Les suites exponentielles	11
3.2	Cas des suites croissantes	12
3.2.1	Absence de double carrés	12
3.2.2	Impact des plats	12
3.2.3	Une conjecture pour la densité	13
4	Conclusion	14

1 Définitions et notations

Généralités sur les mots

On pose Σ un alphabet fini.

On supposera pour toute la suite que $\{0, 1\} \subseteq \Sigma$.

Une lettre est un élément de Σ .

On note la concaténation par $\cdot$, ou simplement par l'absence de symboles.

Un mot est soit le mot vide noté ε , soit une concaténation de lettres (finie ou infinie).

Soit w un mot.

On note $|w|$ sa longueur (si w est infini, on pose $|w| = \infty$).

On note $\text{Alph}(w)$ l'ensemble des lettres apparaissant dans w .

Si w est fini, si $n \in \mathbb{N}$, on pose $w^n = \underbrace{w \cdot \dots \cdot w}_{n \text{ fois}}$, avec pour convention $w^0 = \varepsilon$.

Les facteurs

Soit w un mot (fini ou infini).

x est un facteur de w lorsqu'il existe des mots u, y tels que $w = uxy$.

Si $a, b \in \mathbb{N}^*$, $a \leq b \leq |w|$, on note $w[a, b]$ le facteur de w entre la a -ème lettre et la b -ème lettre incluses.

x est un préfixe de w lorsqu'il existe un mot y tel que $w = xy$.

Si $n \in \mathbb{N}^*$, $n \leq |w|$, on note $w[1, n]$ le préfixe de w constitué de ses n premières lettres.

Les carrés

Soit w un mot (fini ou infini).

w est un carré lorsqu'il existe un mot u tel que $w = uu$.

On considèrera que le mot vide ε est un carré.

On note $S(w)$ l'ensemble des facteurs carrés de w .

$|S(w)|$ est notamment le nombre de carrés distincts apparaissant dans w .

On note $d(w)$ la densité en carrés distincts de w , si elle existe.

Si w est fini : $d(w) = \frac{|S(w)|}{|w|}$.

Si w est infini, c'est la limite, si elle existe, des densités en carrés distincts de ses préfixes : $d(w) = \lim_{n \rightarrow \infty} d(w[1, n])$.

w est dit dense en carrés distincts lorsque $d(w)$ existe et vaut 1.

Les positions

Soit w un mot. Soit $i \in \mathbb{N}$, $1 \leq i \leq |w|$.

La $i^{\text{ème}}$ lettre de w en partant de la gauche est dite de position i .

La position finale d'un facteur de w est la position de sa lettre la plus à droite.

On note $S_i(w)$ l'ensemble des facteurs carrés de w dont leur apparition la plus à gauche est de position finale i . $S_i(w)$ correspond aussi aux facteurs carrés de w tels que le plus petit préfixe de w où ils apparaissent est $w[1, i]$.

Remarque 1. Si w est un mot fini, alors $S(w) = \bigcup_{1 \leq i \leq |w|} S_i(w)$, et c'est une union

disjointe.

La position i est une lacune lorsque $S_i(w) = \emptyset$.

La position i est un double carré lorsque $|S_i(w)| = 2$.

Les plats

Soit $u = (u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels. Soit $i \in \mathbb{N}^*$.

i est un plat lorsque $u_i = u_{i+1}$.

On note M_u l'ensemble des plats de u : $M_u = \{i \in \mathbb{N}^* | u_i = u_{i+1}\}$.

La construction des mots étudiés

Soit $u = (u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels.

On pose, pour n dans $\mathbb{N}^*$: $q_u(n) = 0^{u_{n+1}} 1 0^{u_n} 1 0^{u_{n+1}} 1$

On pose :
$$\begin{cases} Q_u(0) = \varepsilon \\ \forall n \in \mathbb{N}^*, Q_u(n) = Q_u(n-1) \cdot q_u(n) = q_u(1) \cdot [...] \cdot q_u(n) \end{cases}$$

Le mot infini relatif à la suite u est $Q_u(\infty) = \lim_{n \rightarrow \infty} Q_u(n) = q_u(1) \cdot q_u(2) \cdot q_u(3) \cdot [...]$

Remarque 2.
$$\begin{cases} \forall n \in \mathbb{N}^*, |q_u(n)| = 2u_{n+1} + u_n + 3 \\ \forall n \in \mathbb{N}^*, |Q_u(n)| = \sum_{1 \leq i \leq n} 2u_{i+1} + u_i + 3 \end{cases}$$

2 Le contexte

2.1 Contexte humain

2.1.1 Le laboratoire

Le Laboratoire de Combinatoire et d'Informatique Mathématique (LaCIM) de l'Université du Québec à Montréal (UQAM) se situe en plein coeur de Montréal au Canada. Les sujets de recherche y sont à l'interface entre mathématiques, combinatoire et informatique théorique. Le laboratoire est très ouvert à l'international, j'ai pu moi-même rencontrer des personnalités venant aussi bien de Belgique et d'Italie que du Mexique et des Etats-Unis. C'est un laboratoire à échelle humaine - une vingtaine de chercheurs - avec une ambiance très conviviale. Ayant réalisé mon stage pendant les vacances d'été, peu de chercheurs étaient présents. Malgré cela, le laboratoire reste actif avec un séminaire étudiant chaque semaine, un café convivial tous les mercredis après-midi et des séances d'initiation à SageMath : un logiciel tout-en-un basé sur Python pour faire des mathématiques et du calcul formel. Ce n'est pas un hasard étant donné que beaucoup de contributeurs de ce logiciel viennent du LaCIM. SageMath contient une pléthore de bibliothèques et peut s'avérer très utile, je recommande sa découverte. Côté locaux, une salle informatique d'une quinzaine de postes est à disposition du personnel, au choix entre Ubuntu et OS X. Par contre pas de partage des données entre ordinateurs pour le moment, ce qui peut être compensé par l'utilisation d'un GitHub si l'on n'est pas certain de rester sur le même poste.

2.1.2 Le déroulement du stage

Je suis resté au LaCIM du 26 juin au 6 août. Mon directeur de stage Srečko Brlek ne pouvait pas être présent pour les deux premières semaines. A mon arrivée, j'ai été accueilli par ses deux post-doctorants : Elise Vandomme et Francesco Dolce. Après un briefing sur mes connaissances, ils m'ont aiguillé sur les facteurs carrés distincts, et notamment la conjecture des carrés, qui fait l'objet de leur recherche.

N'ayant aucune expérience antérieure dans ce domaine, les deux premières semaines m'ont servi à prendre mes marques, me renseigner sur l'état de l'art et décider de la question précise que je pourrais aborder autour de la conjecture des carrés. J'ai réalisé un programme Python naïf pour calculer les facteurs carrés dans un mot et leur nombre, cela pour expérimenter et gagner en intuition. Une chose qui m'avait grandement frappé était ma difficulté à trouver par moi-même des mots avec beaucoup de facteurs carrés distincts. J'arrivais rarement à dépasser la moitié de la longueur du mot, alors qu'il est possible d'obtenir un nombre de facteurs carrés distincts du même ordre de grandeur que la longueur. C'est donc sur cette voie que je me suis engagé pour le reste du stage.

Sur les quatre dernières semaines, mon directeur de stage était présent, et nous organisons moi, Srečko, Elise et Francesco, une séance de mise au point au moins une fois par semaine,

et avec au moins deux d'entre eux à chaque fois. Ce mode de fonctionnement m'a beaucoup aidé à m'organiser et à m'entraîner à présenter mes résultats ou mes intuitions, aussi bien au tableau qu'à l'oral.

2.2 Contexte scientifique

2.2.1 La recherche sur les carrés

La combinatoire des mots est, malgré une apparente simplicité de nombre de ses questions, un domaine de recherche d'actualité. La plupart des articles de recherche que j'ai rencontré sont plutôt récents par rapport à l'histoire de l'informatique, le plus vieux datant de 1989. Ceci peut s'expliquer par l'avancée de la puissance des ordinateurs, qui ont pu permettre de multiplier les expérimentations et d'établir des conjectures solides. Gardons à l'esprit, par exemple, que le nombre de mots de longueur $n \in \mathbb{N}$ est exponentiel en n . Ce simple fait complique grandement la détermination de propriétés générales sur les mots.

La recherche sur les carrés ne fait pas exception, et est même le cas d'école du sujet coriace de combinatoire des mots. On pourra trouver un résumé intéressant de l'état de l'art sur le sujet dans [CIR09]. Ce qui va nous intéresser particulièrement est le nombre de facteurs carrés distincts qu'un mot peut contenir. Le premier résultat notable fut établi en 1998 par Fraenkel et Simppson dans [FS98]. Ils commencent par montrer le résultat suivant :

Lemme 2.2.1. *Soit w un mot fini. Alors, pour toute position $1 \leq i \leq |w|$, il existe au plus deux facteurs carrés de w dont l'apparition la plus à gauche est de position finale i .*

Ce qui revient à dire, avec nos notations : $\forall 1 \leq i \leq |w|, |S_i(w)| \leq 2$. En partant d'un préfixe de w , on peut donc avoir zéro, un ou deux nouveaux facteurs carrés à chaque ajout de la lettre suivante dans w . Ceci permet de majorer leur nombre par deux fois la longueur du mot, quand celui-ci est fini :

Corollaire 2.2.1 (Fraenkel et Simpson, 1998). *Un mot fini de longueur n a au plus $2n$ facteurs carrés distincts.*

On ne trouve cependant jamais de mot de longueur n avec plus de n facteurs distincts expérimentalement, ce qui pousse à conjecturer n comme majorant :

Conjecture 2.2.1 (Conjecture des carrés). *Un mot fini w de longueur n a au plus $n - |\text{Alph}(w)| + 1$ facteurs carrés distincts.*

Le " $-|\text{Alph}(w)|$ " vient du fait qu'à la première apparition en partant de la gauche d'une lettre dans w , on ne peut pas avoir un nouveau facteur carré, où chacune de ces lettres apparaît un nombre pair de fois. Et le " $+1$ " sert à compter le mot vide, qui par convention est compté comme un facteur carré.

Cette conjecture est ouverte depuis une vingtaine d'années. Le meilleur résultat obtenu à ce jour est $\lfloor \frac{11n}{6} \rfloor$, démontré dans [DFT15] en 2015.

En plus des mots finis, il est possible de s'intéresser aux mots infinis en introduisant une notion de densité en facteurs carrés distincts. On peut par exemple citer l'étude du mot de Fibonacci dans [FS99] et celle du mot de Thue-Morse dans [Br189]. Ces deux mots échouent

cependant à atteindre la densité maximale prévue par la conjecture des carrés, à savoir 1. Un exemple de mot infini dense en carrés distincts est donné dans [FS98], mais c'est le seul exemple que j'ai pu trouver dans la littérature.

2.2.2 Quelques différences notables entre carrés et palindromes

Pour bien comprendre à quel point l'étude des carrés est spécifique et avec ses difficultés propres, il peut être judicieux de la comparer avec celle des palindromes, qui est complètement résolue.

Pour les palindromes, on a le lemme suivant :

Lemme 2.2.2. *Soit w un mot fini. Alors, pour toute position $1 \leq i \leq |w|$, il existe au plus un facteur palindromique de w dont l'apparition la plus à gauche est de position finale i .*

La preuve n'est pas très difficile, c'est d'ailleurs un bon exercice. En appliquant ce lemme pour toutes les positions du mot, on en déduit le corollaire suivant :

Corollaire 2.2.2. *Un mot fini de longueur n a au plus n facteurs palindromiques distincts.*

Ce majorant est atteint par exemple par les mots composés d'une seule lettre.

Pour les carrés par contre, on a vu que les résultats qu'on peut établir sont bien moins forts. La principale différence - et difficulté - est que contrairement aux palindromes, il est possible d'avoir deux nouveaux carrés sur une même position. Il n'est pas aisé de comprendre l'impact des doubles carrés sur la structure d'un mot et sa densité en carrés. C'est d'ailleurs par ce chemin-là que l'amélioration du majorant à $\lfloor \frac{11n}{6} \rfloor$ a été obtenue dans [DFT15]. Il reste cependant encore beaucoup à faire en recherche sur les carrés.

3 La contribution

Mon objectif était de tenter de généraliser l'exemple de mot infini dense en facteurs carrés distincts proposé par Fraenkel et Simpson dans [FS98] :

Soit : $\forall n \in \mathbb{N}^*, q(n) = 0^{n+1} 10^n 10^{n+1} 1$

Soit :
$$\begin{cases} Q_u(0) = \varepsilon \\ \forall n \in \mathbb{N}^*, Q(n) = Q(n-1) \cdot q(n) = q(1) \cdot [\dots] \cdot q(n) \end{cases}$$

Le mot infini est défini par : $Q(\infty) = \lim_{n \rightarrow \infty} Q(n) = q(1) \cdot q(2) \cdot q(3) \cdot [\dots]$

L'idée de la généralisation est de modifier la longueur des séries de '0' dans la définition des $q(n)$. On se fixe $u = (u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels. Pour $n \in \mathbb{N}^*$, dans la définition de $q(n)$, on remplace l'exposant n par u_n et l'exposant $(n+1)$ par u_{n+1} .

Dans le cas de l'exemple de Fraenkel et Simpson, la suite en question est simplement l'identité : $(u_n)_{n \in \mathbb{N}^*} = (n)_{n \in \mathbb{N}^*}$. Il s'agit d'une suite strictement croissante pour les exposants. Une idée peut donc être de se restreindre dans un premier temps aux suites strictement croissantes pour espérer retrouver une répartition analogue des facteurs carrés, ce qui s'avère être le cas.

3.1 Cas des suites strictement croissantes

Dans toute cette partie, on fixe $(u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels strictement croissante.

3.1.1 Preuve d'absence de double carrés

Une première propriété intéressante de l'exemple de Fraenkel et Simpson est qu'aucune de ses positions n'est un double carré. Cela permet, pour toute position i du mot, de majorer le cardinal de $S_i(Q(\infty))$ par 1, donc d'après la remarque 1 de majorer, pour tout n dans $\mathbb{N}^*$, $S(Q(\infty)[1, n])$ par n . On peut se demander si cette propriété est également vraie en utilisant n'importe quelle suite strictement croissante.

On procède de même que Fraenkel et Simpson pour la démonstration. On commence par remarquer la chose suivante :

Lemme 3.1.1. *Tout facteur carré dans $Q_u(\infty)$ contient au plus quatre fois la lettre '1'.*

Démonstration. Par l'absurde :

Si ce n'est pas le cas, alors il existe dans $Q_u(\infty)$ un facteur carré uu avec plus de quatre fois la lettre '1', donc tel que u a au moins trois fois la lettre '1'.

Il y a alors nécessairement un facteur v de u de la forme $10^x 10^y 1$ pour certains entiers x et y . v est un facteur de u et uu est censé apparaître dans $Q_u(\infty)$. Donc il devrait y avoir au moins deux occurrences disjointes de v dans $Q_u(\infty)$.

Or, d'après la construction de $Q_u(\infty)$, il a deux possibilités pour v : soit $v = 10^{u_{n+1}} 10^{u_n} 1$

pour un certain n , soit $v = 10^{u_n} 10^{u_{n+1}} 1$ pour un certain n .

- Si $v = 10^{u_{n+1}} 10^{u_n} 1$ pour un certain n : v n'apparaît qu'une fois dans $Q_u(\infty)$, donc il n'y a pas au moins deux occurrences de v : contradiction.
- Si $v = 10^{u_n} 10^{u_{n+1}} 1$: v apparaît au plus deux fois dans $Q_u(\infty)$: une fois dans la partie $q_u(n)$, et une fois entre $q_u(n-1)$ et $q_u(n)$ pour $n \geq 2$:

$$Q_u(\infty) = Q_u(n-2) \cdot \underbrace{0^{u_n} 10^{u_{n-1}} 10^{u_n} 1}_{q_u(n-1)} \cdot \underbrace{0^{u_{n+1}} 10^{u_n} 10^{u_{n+1}} 1}_{q_u(n)} \cdot [...]$$

Dans le cas où $n = 1$, v n'apparaît qu'une fois et on conclut de la même manière que le point précédent. Pour $n \geq 2$, v apparaît deux fois, mais ces deux occurrences partagent une lettre '1' : celle tout à droite pour l'un, et tout à gauche pour l'autre. Donc il n'y a pas au moins deux occurrences disjointes de v dans $Q_u(\infty)$: contradiction. $\square$

Ceci établi, cela signifie que les facteurs carrés de $Q_u(\infty)$ sont de l'une des formes suivantes, pour n dans $\mathbb{N}^*$:

- $(0)_n : \{0^{2n}\}$
- $(2)_n : \{(0^{u_n-i} 10^i)^2, 0 \leq i \leq u_n\}$
- $(4\vee)_n : \{(0^{u_{n+1}-i} 10^{u_n} 10^i)^2, 0 \leq i \leq u_{n+1}\}$
- $(4\wedge)_n : \{(0^{u_n-i} 10^{u_{n+1}} 10^i)^2, 0 \leq i \leq u_n\}$

On peut vérifier sans mal qu'ils apparaissent effectivement tous dans $Q_u(\infty)$. Il reste maintenant à déterminer pour chacun de ces facteurs carrés c pour quelle position i on a $c \in S_i(Q_u(\infty))$ - id est la position finale de la première apparition vers la gauche de chaque carré.

Lemme 3.1.2. *Soit $n \in \mathbb{N}, n \geq 2$. Les positions finales des premières apparitions de carrés dans la partie $q_u(n)$ du mot infini $Q_u(\infty)$ est la suivante :*

$$Q_u(\infty) = Q_u(n-1) \cdot \underbrace{0 \dots 0}_{\substack{(4\wedge)_{n-1} \\ 0 \leq i \leq u_{n-1}}} \underbrace{[...]}_{\substack{(2)_n \\ \max(u_{n-1}+1, u_n-u_{n-1}) \leq i \leq u_n}} \underbrace{[...]}_{\substack{(0)_i \\ i \text{ pair}, u_n+1 \leq i \leq u_{n+1}}} \underbrace{0 \dots 0}_{u_{n+1}} 1$$

$$\underbrace{0 \dots 0}_{\substack{(2)_{n+1} \\ u_{n+1}-u_n \leq i \leq u_n}} 1 \underbrace{0 \dots 0}_{\substack{(2)_n \\ 0 \leq i \leq u_n-u_{n-1}-1}} \underbrace{[...]}_{\substack{(4\vee)_n \\ u_{n+1}-u_{n-1} \leq i \leq u_{n+1}}} \underbrace{0 \dots 0}_{u_{n+1}} 1 \cdot [...]$$

Ceci fait, on remarque qu'il n'y a bel et bien pas de doubles carrés dans la partie $q_u(n)$ du mot infini $Q_u(\infty)$, pour $n \geq 2$. Il ne reste plus qu'à traiter le cas $n = 1$ pour conclure.

Théorème 3.1.1. *Si $(u_n)_{n \in \mathbb{N}^*}$ est une suite d'entiers strictement croissante, alors $Q_u(\infty)$ est sans double carrés.*

Démonstration. On montre pour tout n dans $\mathbb{N}^*$ que chaque ensemble de positions au niveau d'un $q_u(n)$ est sans doubles carrés.

- Pour $n = 1$: les seuls carrés présents sont les $(0)_i$ pour $0 \leq i \leq u_2$ dans la série de '0' à gauche, et les $(2)_1$ pour $0 \leq i \leq u_1$, qui ont leur position finale sur la série de '0' à droite. Il n'y a donc pas de doubles carrés.

— Pour $n \geq 2$: on applique le lemme 3.1.2 pour montrer qu'il n'y a pas de doubles carrés. $\square$

Cette absence de doubles carrés sera très utile par la suite, on peut d'ores déjà remarquer la chose suivante :

Corollaire 3.1.1 (Majoration de $d(Q_u(\infty))$). *Soit $(u_n)_{n \in \mathbb{N}^*}$ est une suite d'entiers strictement croissante Si $d(Q_u(\infty))$ existe, alors $d(Q_u(\infty)) \leq 1$.*

Démonstration. D'après le théorème 3.1.1 : $\forall i \in \mathbb{N}^*, S_i(Q_u(\infty)) \leq 1$.

Donc en utilisant la remarque 1 :

$$\forall n \in \mathbb{N}^*, S(Q_u(\infty)[1, n]) \leq \sum_{1 \leq i \leq n} S_i(Q_u(\infty)) \leq n = |Q_u(\infty)[1, n]|.$$

Donc : $\forall n \in \mathbb{N}^*, d(Q_u(\infty)[1, n]) \leq 1$.

D'où le résultat. $\square$

3.1.2 Une condition suffisante pour la densité

On cherche à encadrer la densité en carrés distincts des préfixes de $Q_u(\infty)$.

Lemme 3.1.3. *Soit $u = (u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels strictement croissante.*

On suppose qu'il existe un rang n_0 tel que : $\forall n \geq n_0, u_{n+1} \leq 2u_n$.

Alors :

$$\forall n \geq n_0, \frac{\sum_{n_0 \leq i \leq n} 3u_i + u_{i-1} - u_{i+1}}{\sum_{1 \leq i \leq n} 2u_{i+1} + u_i + 3} \leq d(Q_u(n)) \leq 1$$

Démonstration.

— Pour la majoration : on procède comme dans la démonstration du corollaire 3.1.1.

— Pour la minoration : soit $n \geq n_0$. Soit $n_0 \leq i \leq n$.

Alors, d'après le lemme 3.1.2, on a un certain nombre de nouveaux carrés dans la partie $q_u(i)$ de $Q_u(n)$:

— ceux de type $(4 \wedge)_{i-1}$: ils sont u_{i-1}

— ceux de type $(2)_{i+1}$: ils sont $2u_i - u_{i+1}$ (rendu possible par l'hypothèse du lemme : $2u_i - u_{i+1} \geq 0$)

— ceux de type $(2)_i$ dans la série de '0' la plus à droite : ils sont $u_i - u_{i-1}$

— ceux de type $(4 \vee)_i$: ils sont u_{i-1}

En sommant, cela nous donne au moins $3u_i + u_{i-1} - u_{i+1}$ positions avec un nouveau carré dans la partie $q_u(i)$ de $Q_u(n)$.

Donc : $S(Q_u(n)) \geq \sum_{n_0 \leq i \leq n} 3u_i + u_{i-1} - u_{i+1}$.

On conclut en divisant par $|Q_u(n)|$ et en utilisant la remarque 2. $\square$

On peut en déduire la condition suffisante suivante :

Théorème 3.1.2. *Soit $(u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels strictement croissante.*

Si $u_n \sim u_{n+1}$, alors $d(Q_u(\infty))$ existe et vaut 1.

Démonstration. $u_n \sim u_{n+1}$.

Donc il existe un rang n_0 tel que : $\forall n \geq n_0, \frac{u_{n+1}}{u_n} \leq 2$.

Soit $n \geq n_0$.

On peut alors appliquer le lemme 3.1.3 pour encadrer $d(Q_u(n))$. Or, $(u_n)_{n \in \mathbb{N}^*}$ est strictement croissante. Donc : $\forall n \in \mathbb{N}^*, u_n \geq n$. Donc la série de terme général u_n diverge. Donc, en sachant que $u_n \sim u_{n+1}$, on peut montrer que :

$$\frac{\sum_{n_0 \leq i \leq n} 3u_i + u_{i-1} - u_{i+1}}{\sum_{1 \leq i \leq n} 2u_{i+1} + u_i + 3} \rightarrow 1$$

. Cependant cela ne permet en l'état que de minorer la densité des préfixes des étapes de construction de $Q_u(\infty)$, et non de tous les préfixes de $Q_u(\infty)$.

Pour minorer la densité des préfixes entre $Q_u(n)$ et $Q_u(n+1)$, il suffit d'utiliser la quantité :

$$\frac{\sum_{n_0 \leq i \leq n} 3u_i + u_{i-1} - u_{i+1}}{\sum_{1 \leq i \leq n+1} 2u_{i+1} + u_i + 3}$$

(on va désormais jusqu'à $n+1$ dans la somme au dénominateur), cette quantité tendant aussi vers 1 pour les mêmes raisons que celles évoquées précédemment.

Donc, par le théorème des gendarmes, $d(Q_u(\infty))$ existe et vaut 1. $\square$

Une application directe est les suites polynomiales.

Corollaire 3.1.2. *Soit P une fonction polynomiale de degré supérieur ou égal à 1.*

Soit $(u_n)_{n \in \mathbb{N}^} = (\lfloor P(n) \rfloor)_{n \in \mathbb{N}^*}$.*

Alors $Q_u(\infty)$ est dense en carrés distincts.

Démonstration. Une suite $(u_n)_{n \in \mathbb{N}^*}$ polynomiale est strictement croissante à partir d'un certain rang et vérifie : $u_n \sim u_{n+1}$.

- Si $(u_n)_{n \in \mathbb{N}^*}$ est strictement croissante dès le départ : on peut appliquer le théorème 3.1.2 et conclure.
- Si $(u_n)_{n \in \mathbb{N}^*}$ est strictement croissante seulement à partir d'un rang m_0 : remarquons d'abord que les carrés mis en évidence dans le lemme 3.1.2 apparaissent dans notre cas pour $n \geq m_0 + 1$. Donc le lemme 3.1.3 et le théorème 3.1.2, restent valables au prix de remplacer n_0 par $\max(n_0, m_0)$ dans l'encadrement du lemme 3.1.3, et on peut bien démontrer que $Q_u(\infty)$ est dense en carrés distincts. $\square$

3.1.3 Les suites exponentielles

Il s'agit d'un exemple de suites strictement croissantes non concluantes.

Lemme 3.1.4. *Il y a au moins $\lfloor \frac{u_{n+1} - u_n}{2} \rfloor$ lacunes dans la partie $q_u(n)$ de $Q_u(\infty)$.*

Démonstration. A chaque début d'une partie $q_u(n)$ dans $Q_u(\infty)$, on a une série de '0' de longueur inédite : à partir du $(u_n + 1)^{\text{ème}}$ '0' jusqu'au $(u_{n+1})^{\text{ème}}$, les $i^{\text{èmes}}$ avec i impair sont des lacunes. D'où les $\lfloor \frac{u_{n+1} - u_n}{2} \rfloor$ lacunes au moins. $\square$

Ce résultat élémentaire élémentaire, en le combinant avec le théorème 3.1.1, suffit à montrer que les suites exponentielles sont non concluantes.

Corollaire 3.1.3. *Si $\alpha \in]1, +\infty[$, si $(u_n)_{n \in \mathbb{N}^*} = (\lfloor \alpha^n \rfloor)_{n \in \mathbb{N}^*}$, alors $Q_u(\infty)$ n'est pas dense en carrés distincts.*

Démonstration. En utilisant le théorème 3.1.1 et le lemme 3.1.4 on montre que :

$$\forall n \in \mathbb{N}^*, d(Q_u(n)) \leq 1 - \frac{\sum_{1 \leq i \leq n} \lfloor \frac{|\alpha^{i+1}| - |\alpha^i|}{2} \rfloor}{\sum_{1 \leq i \leq n} 2\lfloor \alpha^{i+1} \rfloor + \lfloor \alpha^i \rfloor + 3}$$

Or, on peut montrer sans grande difficulté que :

$$\frac{\sum_{1 \leq i \leq n} \lfloor \frac{|\alpha^{i+1}| - |\alpha^i|}{2} \rfloor}{\sum_{1 \leq i \leq n} 2\lfloor \alpha^{i+1} \rfloor + \lfloor \alpha^i \rfloor + 3} \rightarrow c > 0$$

c étant une constante.

Donc $d(Q_u(n))$ ne tend pas vers 1.

Donc $d(Q_u(\infty))$, si elle existe, ne peut pas valoir 1. □

3.2 Cas des suites croissantes

Dans toute cette partie, on fixe $(u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers naturels croissante.

3.2.1 Absence de double carrés

Cette propriété, démontrée pour les suites strictement croissantes, s'avère vérifiée pour toutes les suites croissantes.

Théorème 3.2.1. *Si $(u_n)_{n \in \mathbb{N}^*}$ est une suite d'entiers croissante, alors $Q_u(\infty)$ est sans double carrés.*

La preuve, purement technique et peu intéressante, ne sera pas détaillée ici. Remarquons simplement que, comme pour les suites strictement croissantes, cette propriété nous permet d'obtenir une majoration intéressante de la densité.

Corollaire 3.2.1. *Soit $(u_n)_{n \in \mathbb{N}^*}$ est une suite d'entiers croissante. Si $d(Q_u(\infty))$ existe, alors $d(Q_u(\infty)) \leq 1$.*

3.2.2 Impact des plats

On peut se demander s'il y a un bénéfice à passer d'une suite strictement croissante à une suite croissante admettant des plats. La réponse est sans conteste négative, les plats étant source de nombreuses lacunes supplémentaires.

Lemme 3.2.1. *Soit $n \geq 2$. Si $u_n = u_{n+1}$, alors il y a au moins u_{n-1} lacunes dans la partie $q_u(n)$ de $Q_u(\infty)$.*

Démonstration. Il s'agit des u_{n-1} '0' les plus à droite dans la deuxième série de '0'. On montre que si un carré a pour position finale l'une de ses positions i , alors il apparaît déjà $u_n + 1$ crans plus à gauche, et donc ce carré n'est pas dans $S_i(Q_u(\infty))$. □

On peut se servir de ce résultat pour montrer que des suites sont non concluantes, comme les suites "escalier" où les marches restent du même ordre de grandeur.

Corollaire 3.2.2. *Si $\alpha \in]0, 1[$, si $(u_n)_{n \in \mathbb{N}^*} = (\lfloor \alpha n \rfloor)_{n \in \mathbb{N}^*}$, alors $Q_u(\infty)$ n'est pas dense en carrés distincts.*

Démonstration. En combinant le corollaire 3.2.1 et le théorème 3.2.1, on montre que :

$$\forall n \in \mathbb{N}, n \geq 2, d(Q_u(n)) \leq 1 - \frac{\sum_{1 \leq i \leq n} \lfloor \alpha(i-1) \rfloor - \sum_{1 \leq i \leq \lfloor \alpha n \rfloor} i}{\sum_{1 \leq i \leq n} 2 \lfloor \alpha(i+1) \rfloor + \lfloor \alpha i \rfloor + 3}$$

Or, on peut montrer sans grand mal que :

$$\frac{\sum_{1 \leq i \leq n} \lfloor \alpha(i-1) \rfloor - \sum_{1 \leq i \leq \lfloor \alpha n \rfloor} i}{\sum_{1 \leq i \leq n} 2 \lfloor \alpha(i+1) \rfloor + \lfloor \alpha i \rfloor + 3} \rightarrow c \in]0, 1[$$

Donc $d(Q_u(n))$ ne tend pas vers 1.

Donc $d(Q_u(\infty))$, si elle existe, ne peut pas valoir 1. □

3.2.3 Une conjecture pour la densité

Maintenant que l'on a montré que les plats ont un impact négatif sur la densité, on peut se demander si c'est la seule cause suffisamment significative pour contrarier la convergence de la densité de $Q_u(\infty)$ vers 1. La conjecture suivante prévoit que c'est le cas.

Conjecture 3.2.1. *Soit $(u_n)_{n \in \mathbb{N}^*}$ une suite d'entiers croissante.*

On suppose qu'il existe $(v_n)_{n \in \mathbb{N}^}$ telle que $\begin{cases} u_n \sim v_n, \\ Q_v(\infty) \text{ est dense en carrés distincts.} \end{cases}$*

Alors :

$$Q_u(\infty) \text{ est dense en carrés distincts} \Leftrightarrow \frac{\sum_{2 \leq i \leq n, i \in M_u} u_{i-1}}{\sum_{1 \leq i \leq n} u_i} \rightarrow 0$$

On peut d'ores et déjà démontrer le sens direct à l'aide du théorème 3.2.1 et du lemme 3.2.1. Il s'agit du même raisonnement que pour montrer que les suites "escalier" précédentes sont non concluantes. J'invite donc la lectrice/le lecteur à s'essayer à cette preuve.

4 Conclusion

La recherche sur les carrés s'avère être une expérience étonnante et enrichissante. Ce stage fut une grande opportunité pour découvrir en pratique le monde de la recherche : état de l'art, collaboration, autonomie dans sa réflexion. Cela m'a permis également de me rendre compte à quel point six semaines est une durée très courte pour de la recherche : j'ai dû mettre en suspens nombre de mes développements avec l'arrivée de l'échéance. Ceci inclut :

- explorer ce qu'il se passe entre les suites strictement croissantes de croissance polynomiale et celles de croissance exponentielle ;
- prouver ou donner un contre-exemple à la conjecture pour les suites croissantes ;
- établir des résultats pour des suites non nécessairement croissantes.

Au niveau humain, je garde un très bon souvenir de mon passage au LaCIM. Je remercie particulièrement Elise Vandomme et Francesco Dolce pour m'avoir accueilli à mon arrivée au LaCIM, et pour avoir suivi de très près mes avancées. Je remercie Nadia Lafrenière pour m'avoir particulièrement aidé à me familiariser avec l'état de la recherche sur les carrés. Je remercie Emile Nadeau pour son code efficace sur le calcul des facteurs carrés d'un mot fini. Je remercie tout l'effectif du LaCIM pour leur ouverture d'esprit, ainsi que leur entrain pour les mathématiques et la combinatoire. Enfin je remercie Srečko Brlek pour m'avoir écouté, pris au sérieux, communiqué son expérience et rendu mon séjour des plus agréables.

Bibliographie

- [Brl89] Srećko Brlek. Enumeration of factors in the thue-morse word. *Discrete Applied Mathematics*, 24(1) :83 – 96, 1989.
- [CIR09] Maxime Crochemore, Lucian Ilie, and Wojciech Rytter. Repetitions in strings : Algorithms and combinatorics. *Theoretical Computer Science*, 410(50) :5227 – 5235, 2009. Mathematical Foundations of Computer Science (MFCS 2007).
- [DFT15] Antoine Deza, Frantisek Franek, and Adrien Thierry. How many double squares can a string contain ? *Discrete Applied Mathematics*, 180 :52 – 69, 2015.
- [FS98] Aviezri S. Fraenkel and Jamie Simpson. How many squares can a string contain ? *J. Comb. Theory Ser. A*, 82(1) :112–120, April 1998.
- [FS99] Aviezri S. Fraenkel and Jamie Simpson. The exact number of squares in fibonacci words. *Theoretical Computer Science*, 218(1) :95 – 106, 1999.
- [FS14] Aviezri S. Fraenkel and Jamie Simpson. Corrigendum : Corrigendum to "the exact number of squares in fibonacci words" [theoret. comput. sci. 218 (1) (1999) 95-106]. *Theor. Comput. Sci.*, 547 :122–, August 2014.
- [GS04] Dan Gusfield and Jens Stoye. Linear time algorithms for finding and representing all the tandem repeats in a string. *Journal of Computer and System Sciences*, 69(4) :525 – 546, 2004.
- [Ili05] Lucian Ilie. A simple proof that a word of length n has at most $2n$ distinct squares. *Journal of Combinatorial Theory, Series A*, 112(1) :163 – 164, 2005.