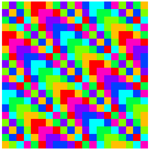
Combinatoire des mots : résultats classiques et avancées récentes

Michel Rigo

http://www.discmath.ulg.ac.be/
http://orbi.ulg.ac.be/

Université de Liège



MATH0470-1 Combinatoire des mots (30h Th, 20h Pr)
Master en sciences mathématiques

D'une manière générale, la combinatoire est la branche des mathématiques qui étudie les "configurations" d'un ensemble "discret" (et généralement fini). On s'intéresse alors au dénombrement ou à l'énumération effective des objets de cet ensemble, à la structure (éventuellement algébrique) de celui-ci ou encore à ses propriétés extrémales (éléments maximaux, ...).

Comme son nom l'indique, la combinatoire des mots s'intéresse plus spécialement aux mots (finis ou infinis), i.e., aux suites de symboles ou de lettres appartenant à un alphabet fini.

Un théorème élémentaire

- Alphabet = ensemble (fini)
- Un mot fini $w : \{1, \ldots, n\} \to A$
- Un mot infini (à droite) = suite à valeurs dans A

$$w : \mathbb{N} \to A$$

- Un facteur $w[i, j]$

$$w_1 \cdots w_i \cdots w_j \cdots w_n, \quad 1 \leq i \leq j \leq n$$

- A^* muni de la concaténation, neutre ε
- Un carré uu

Théorème

Si $\#A = 2$, tout mot suffisamment long a un carré comme facteur.

- Peut-on éviter les cubes sur 2 lettres ?
  Existe-t-il un mot infini sans cube ?
- Peut-on éviter les carrés sur 3 lettres ?

Etant donné un *motif*, e.g. $XYYXY$, ce motif est-il évitable ?
Pour une *propriété* donnée, e.g. contenir un carré abélien, cette propriété est-elle évitable ? Quelle taille minimale d'alphabet ?

Règle de bon sens

Considérer des mots infinis calculables (machines de Turing).
Au plus l'algorithme permettant d'engendrer le mot est *simple*, au mieux on pourra en maîtriser les propriétés (preuves).

$\rightsquigarrow$ Complexité de Kolmogorov

- Peut-on éviter les cubes sur 2 lettres ?
  Existe-t-il un mot infini sans cube ?
- Peut-on éviter les carrés sur 3 lettres ?

Etant donné un *motif*, e.g. $XYYXY$, ce motif est-il évitable ? Pour une *propriété* donnée, e.g. contenir un carré abélien, cette propriété est-elle évitable ? Quelle taille minimale d'alphabet ?

Règle de bon sens

Considérer des mots infinis calculables (machines de Turing). Au plus l'algorithme permettant d'engendrer le mot est *simple*, au mieux on pourra en maîtriser les propriétés (preuves).

$\rightsquigarrow$ Complexité de Kolmogorov

Applications

- Théorie des nombres
- Dynamique symbolique
- Géométrie discrète
- Algorithmique du texte
- Bio-informatique

aacctgaaaa	aagtgtgaaa	aaaattttgt
ggatttgtgg	ataaaacaag	gtttttgcta
atttttgcta	ataaaaaaaa	tttataaaga
gattcgtgaa	agcaaagatt	gtggataacg
aaaaactcaa	gaaaaatttc	ttgacgggtc
ttatcccatc	ttctataatt	taggtgtatc
tatagggat	ataggctttt	ttagatagat
taggaggtgt	gaaaatgaaa	agaacctacc
aaccaaaaaa	ccggcggcgt	aaaagagtgc
acggcttctt	aaagcggatg	cggaccccag

Figure: Carboxydothermus hydrogenoformans Z-2901 genome

Maladies d'Huntington et de Kennedy : $(\mathrm{cag})^j, j > 36$

On se limite aux réels dans $[0,1[$

- x est rationnel $\Leftrightarrow \{x\}$ est rationnel;
- x est algébrique $\Leftrightarrow \{x\}$ est algébrique.

Développement (glouton) en base $b \geq 2$ d'un réel x

$$x = \sum_{i=1}^{+\infty} d_i \, b^{-i} \qquad \bullet \, d_1 \, d_2 \, d_3 \cdots$$

Bijection entre $[0,1[$ et $\{0,\ldots,b-1\}^{\mathbb{N}} \setminus \{0,\ldots,b-1\}^*(b-1)^{\mathbb{N}}$.

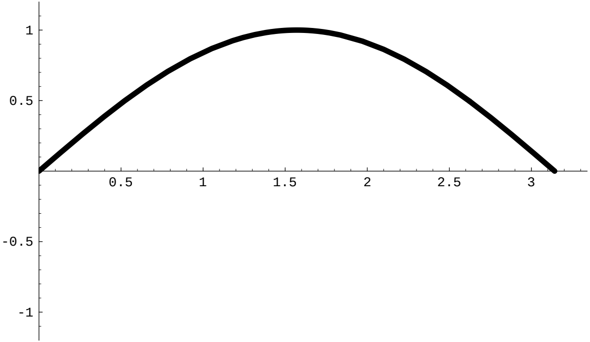
Proposition

Un réel $x \in [0,1[$ est rationnel si et seulement si son développement en base b est ultimement périodique.

Exercice (J.-P. Allouche'99)

Etudier le signe, sur l'intervalle $[0, \pi]$ de la fonction

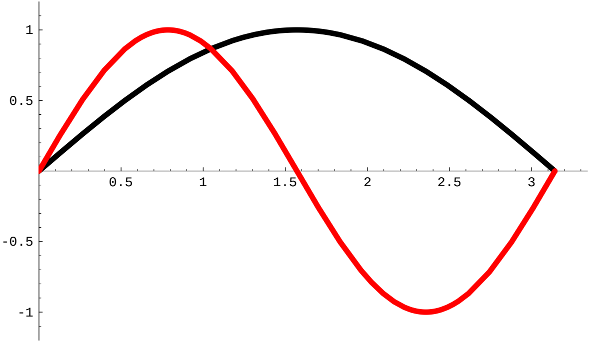
$$F_n(x) = \sin(x)\sin(2x)\sin(4x)\cdots\sin(2^n x)$$



Exercice (J.-P. Allouche'99)

Etudier le signe, sur l'intervalle $[0,\pi]$ de la fonction

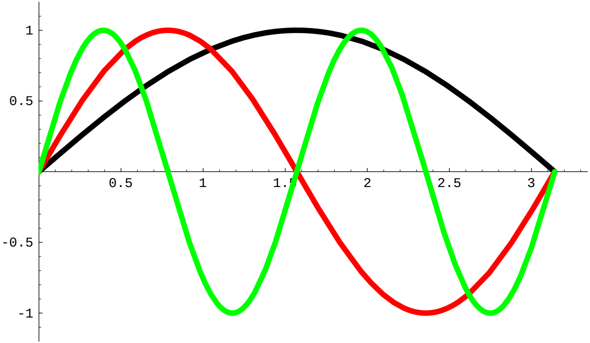
$$F_n(x) = \sin(x)\sin(2x)\sin(4x)\cdots\sin(2^n x)$$



EXERCICE (J.-P. ALLOUCHE'99)

Etudier le signe, sur l'intervalle $[0, \pi]$ de la fonction

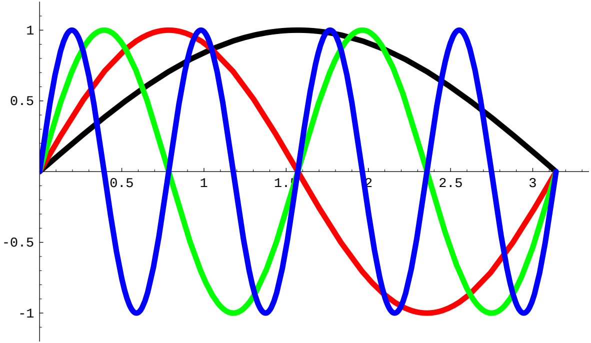
$$F_n(x) = \sin(x)\sin(2x)\sin(4x)\cdots\sin(2^n x)$$



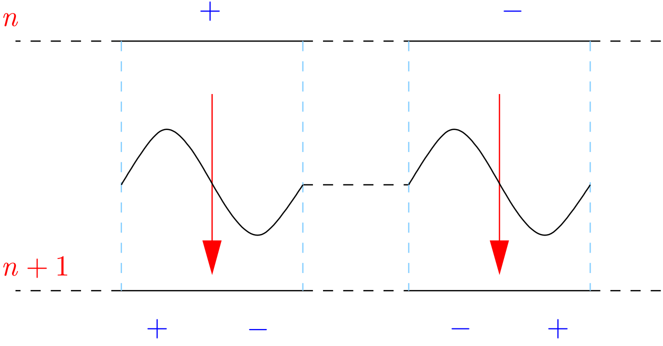
Exercice (J.-P. Allouche'99)

Etudier le signe, sur l'intervalle $[0, \pi]$ de la fonction

$$F_n(x) = \sin(x)\sin(2x)\sin(4x)\cdots\sin(2^n x)$$



$n=0$	$+$
$n=1$	$+ \quad -$
$n=2$	$+ \quad - \quad - \quad +$
$n=3$	$+ \quad - \quad - \quad + \quad - \quad + \quad + \quad -$



$$\boxed{+ \mapsto +-, \quad - \mapsto -+}$$

On peut définir une suite en *itérant un morphisme* (prolongeable)

$$f : a \mapsto ab, \quad b \mapsto ba$$

$$\begin{array}{rcl} f(a) & = & ab \\ f^2(a) & = & abba \\ f^3(a) & = & abbabaab \\ f^4(a) & = & abbabaabbaababba \\ f^5(a) & = & abbabaabbaababbabaababbaabbabaab \end{array}$$

Moyen de construction de suites

Avec un minimum de topologie, cette suite de mots $(f^n(a))_{n\geq 0}$ converge vers un mot infini limite.

Les classiques :

Le mot de Thue–Morse

$a \mapsto ab$, $b \mapsto ba$ (longueur constante)

$$\mathbf{t} = abbabaabbaababbabaababbaabbabaab \cdots$$

$\mathbf{t}$ ne contient aucun chevauchement $cvcvc$, $c \in \{a, b\}$, $v \in \{a, b\}^*$

$$abb|ab|a|abb|a|ab|abb|ab|a|ab|abb|a|abb| \cdots \to 3213123212313 \cdots$$

Le mot de Fibonacci

$0 \mapsto 01$, $1 \mapsto 0$ (longueur non constante)

$$\mathbf{f} = 0100101001001010010100100101001001 \cdots$$

Un mot périodique

$a \mapsto abc,\ b \mapsto abc,\ c \mapsto abc$

$$abcabcabcabcabcabcabcabcabc\cdots$$

La suite caractéristique des carrés parfaits

$a \mapsto abcc,\ b \mapsto bcc,\ c \mapsto c$

$$abccbccccbccccccbccccccccbcccc\cdots$$

$a, b \mapsto 1,\ c \mapsto 0$ (ici, codage supplémentaire)

La suite caractéristique des puissances de 2

$a \mapsto ab,\ b \mapsto bc,\ c \mapsto cc$

$$abbcbcccbcccccccbccccccccccc\cdots$$

$b \mapsto 1,\ a, c \mapsto 0$ (ici, codage supplémentaire)

Un mot infini sans "cube additif"

$$0 \mapsto 03,\ \ 1 \mapsto 43,\ \ 3 \mapsto 1,\ \ 4 \mapsto 01$$

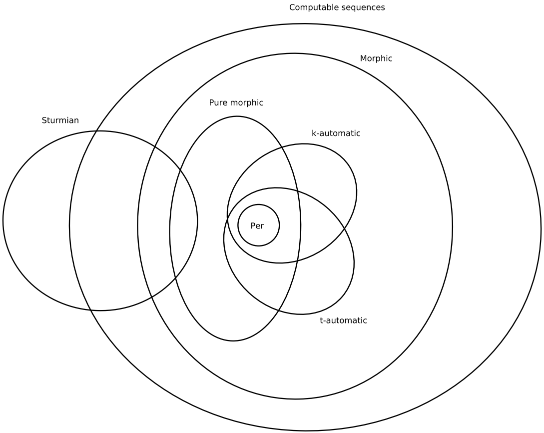
$$031430110343430310110110314303434303434\cdots$$

mot sans cube additif, e.g., 041340.

J. Cassaigne, J. D. Currie, L. Schaeffer, J. Shallit

Taxonomie :

- Point fixe d'un morphisme de longueur constante
- Point fixe d'un morphisme
- Codage (projection sur un alphabet plus petit)
- On peut se limiter aux morphismes non effaçants



L-Systems (Aristid Lindenmayer)

parallel rewriting systems originally introduced in 1968 to model the development of multicellular organisms

https://commons.wikimedia.org/w/index.php?curid=125410

Exemple K. Mahler (1937)

$0.1234567891011\cdots$ est transcendant et n'est pas un nombre de Liouville.

x est un nombre de Liouville, si $\forall n$, $\exists p, q \in \mathbb{N}$, $q > 1$ tels que $0 < |x - p/q| < 1/q^n$. Tout nombre de Liouville is transcendant.

Conjecture Cobham (1968) / Hartmanis–Stearns

Le développement en base b d'un nombre algébrique *irrationnel* ne peut pas être obtenu en itérant un morphisme de longueur constante (+ codage éventuel).

$\rightsquigarrow$ les morphismes de longueurs constantes produisent des nombres rationnels ou transcendants.

Une condition (précise) de bégaiement

Théorème (Adamczewski–Bugeaud)

Si le développement en base b de x débute par des carrés arbitrairement longs, alors x est rationnel ou transcendant.

$$\overline{010}\,\underline{010}\,1001001010010100100101001001010010100100101001$$

$$\overline{01001}\,\underline{01001}\,001010010100100101001001010010100100101001$$

$$\overline{01001010}\,\underline{01001010}\,010100100101001001010010100100101001$$

$$\overline{0100101001001}\,\underline{0100101001001}\,01001001010010100100101001$$

$$\overline{010010100100101001010}\,\underline{010010100100101001010}\,0100101001$$

$$x_\varphi = \sum_{i=1}^{+\infty} f_i\, b^{-i} \text{ est transcendant.}$$

Fonction de complexité (en facteurs)

$$abcabcabcabcabcabcabc\cdots$$

$$\mathsf{p}(n) = 3, \quad \forall n \geq 1$$

$$\mathsf{p}(n) \leq \mathsf{p}(n+1), \forall n$$

$\rightsquigarrow$ lien avec l'entropie d'un système dynamique

Adamczewski–Bugeaud (2004)

Soit x un nombre algébrique irrationnel.
Le développement $\mathbf{w}$ en base b de x vérifie

$$\lim_{n\to+\infty} \frac{\mathsf{p}_{\mathbf{w}}(n)}{n} = +\infty.$$

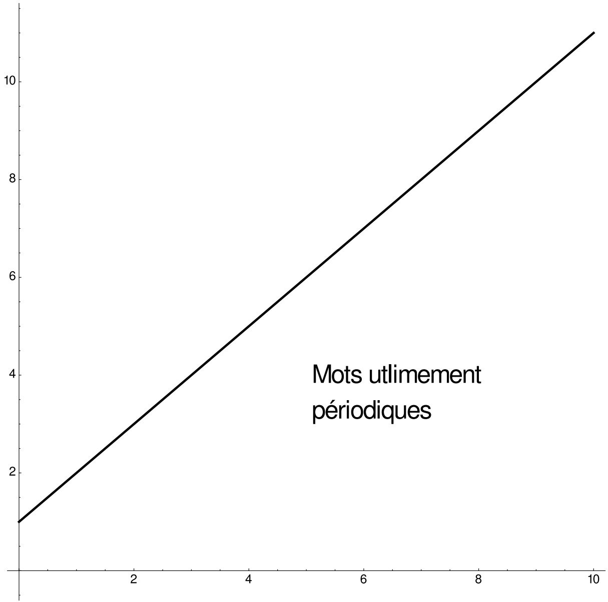
$\rightsquigarrow$ Si $\frac{\mathsf{p}_{\mathbf{w}}(n)}{n}$ est borné, alors x est transcendant (ou rationnel).

Les nombres rationnels ont un développement ultimement périodique.

Morse–Hedlund

Soit $\mathbf{w}$ un mot infini sur un alphabet fini.
Le mot $\mathbf{w}$ est ultimement périodique uv^ω si et seulement si

- il existe k tel que $\mathsf{p}_{\mathbf{w}}(k) \leq k$
- il existe C tel que $\mathsf{p}_{\mathbf{w}}(n) \leq C$, $\forall n \geq 0$
- il existe n_0 tel que $\mathsf{p}_{\mathbf{w}}(n) = \mathsf{p}_{\mathbf{w}}(n_0)$, $\forall n \geq n_0$
- il existe m tel que $\mathsf{p}_{\mathbf{w}}(m) = \mathsf{p}_{\mathbf{w}}(m+1)$



fonction de complexité $\mathsf{p}(n)$

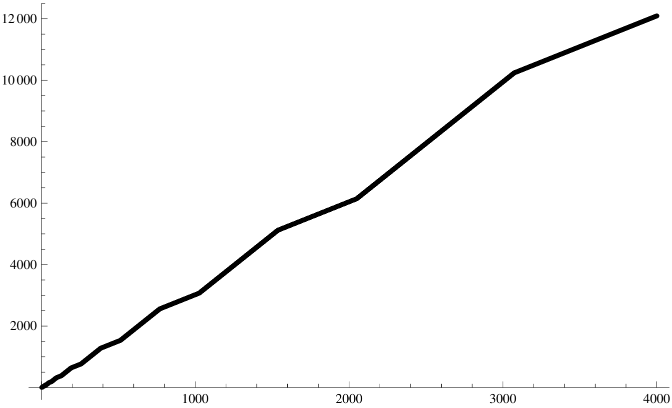
Revenons à la conjecture de Cobham...

Thue–Morse

		00101
00	0010	00110
01	0011	01001
10	0100	01011
11	0101	01100
001	0110	01101
010	1001	10010
011	1010	10011
100	1011	10100
101	1100	10110
110	1101	11001
		11010

$(\mathsf{p}_{\mathbf{t}}(n))_{n\geq 0} = 1, 4, 6, 10, 12, 16, 20, 22, 24, 28, 32, 36, \ldots$ `A005942`

$$\mathsf{p}_{\mathbf{t}}(n) = \begin{cases} 4n - 2\cdot 2^m - 4, & \text{if } 2\cdot 2^m < n \leq 3\cdot 2^m; \\ 2n + 4\cdot 2^m - 2, & \text{if } 3\cdot 2^m < n \leq 4\cdot 2^m. \end{cases}$$



Complexité des suites automatiques[1] :
(morphisme de longueur constante k)

- Soit $\mathbf{x}$ un point fixe d'un morphisme k-uniforme $f : A^* \to A^*$ (un codage est superflu) $|f(b)| = k$ pour tout $b \in A$.
- Soit u un facteur de longueur n apparaissant dans $\mathbf{x}$.
- Il existe i tel que $k^{i-1} \leq n < k^i$.
  Notons que $|f^i(b)| = k^i$ pour tout $b \in A$.
- On considère la factorisation de $\mathbf{x}$ en blocs de longueur k^i de la forme $f^i(b)$.

[1] On ne parlera pas d'automates dans cet exposé

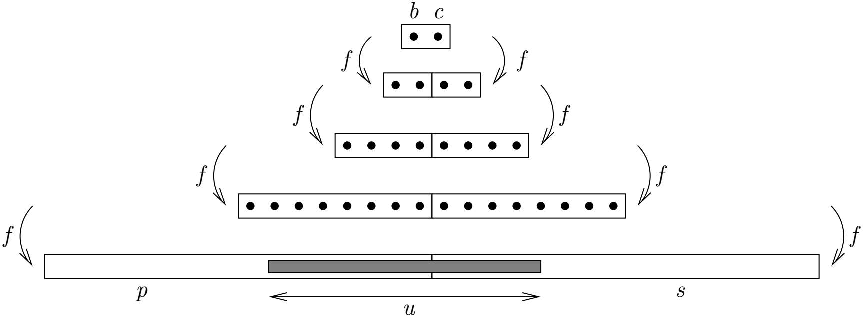


FIGURE: Itération d'un morphisme 2-uniforme.

Le facteur u apparaît dans un $f^i(b)$ ou en chevauche deux.
Il existe deux lettres b et c telles que $f^i(bc) = pus$ avec $|p| < k^i$.

$$\mathsf{p}_{\mathbf{w}}(n) \leq k^i (\# A)^2 = k(\# A)^2 k^{i-1} \leq \underbrace{k(\# A)^2}_{constante} n$$

Théorème

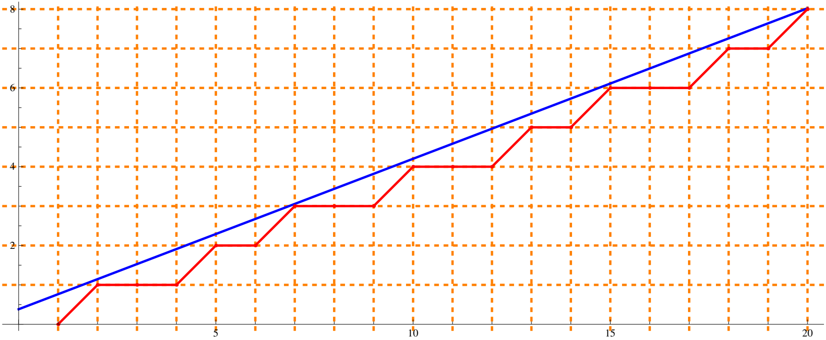
La complexité en facteurs d'une suite automatique est en $\mathcal{O}(n)$.

Au vu du théorème d'Adamczewski–Bugeaud (2004),
les nombres réels "automatiques" sont

- soit rationnels (on les engendre tous),
- soit transcendants.

Géométrie discrète : discrétisation de droites (pente irrationnelle)

$$\lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\,\alpha + \beta \rfloor$$



$$abaababaabaababaababaabaababaabaababaababaabaabab \cdots$$

Mots sturmiens :

Codage d'une droite $y = \alpha x + \beta$ SSI codage de l'orbite

$$\{R^n(\beta) \mid n \geq 0\}$$

pour la transformation

$$R : [0,1[\to [0,1[,\ x \mapsto x + \alpha \pmod 1)$$

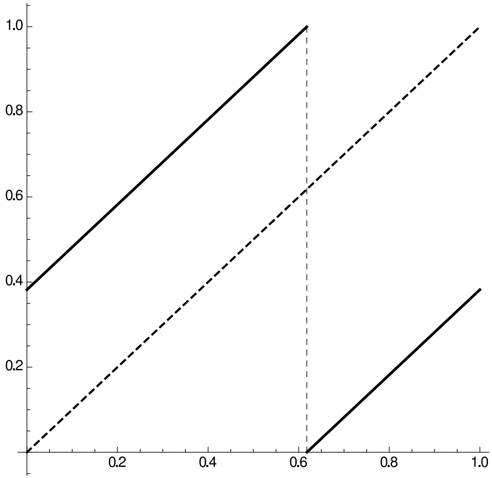
et la partition

$$[0,1[= [0, 1-\alpha[\cup[1-\alpha, 1[.$$

Remarques

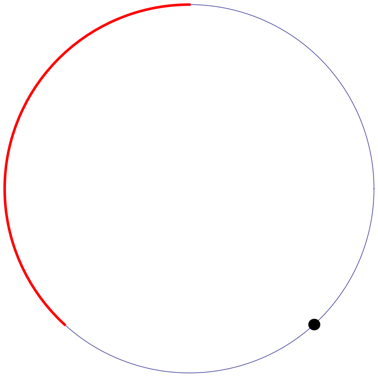
- $\alpha \in \mathbb{Q} \Rightarrow$ orbite périodique
- $\alpha \notin \mathbb{Q} \Rightarrow$ orbite dense (thm. de Kronecker)
- mot de Fibonacci $\alpha = \beta = 1/\varphi^2 \simeq 0,382$

$x \mapsto x + \alpha \pmod 1$ avec $\alpha = 1/\varphi^2 \simeq 0,382$



Réalisation de systèmes dynamiques discrets

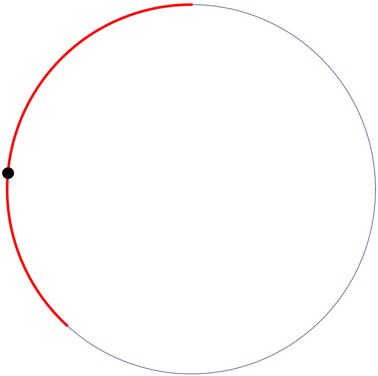
$\mathbb{T} = \mathbb{R}/\mathbb{Z},\ R : \mathbb{T} \to \mathbb{T},\ x \mapsto x + \alpha \pmod 1$



$I_a = [0, 1 - \alpha[,\ I_b = [1 - \alpha, 1[,\ a$

Réalisation de systèmes dynamiques discrets

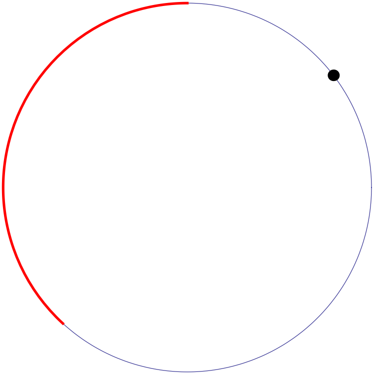
$\mathbb{T} = \mathbb{R}/\mathbb{Z},\ R : \mathbb{T} \to \mathbb{T},\ x \mapsto x + \alpha \pmod 1$



$I_a = [0, 1 - \alpha[,\ I_b = [1 - \alpha, 1[,\ ab$

Réalisation de systèmes dynamiques discrets

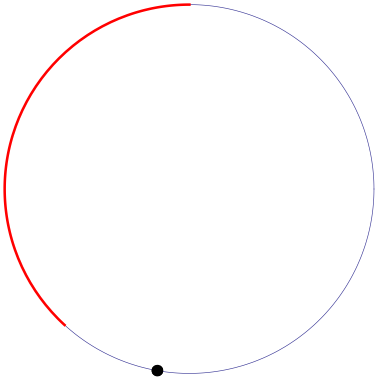
$\mathbb{T} = \mathbb{R}/\mathbb{Z},\ R : \mathbb{T} \to \mathbb{T},\ x \mapsto x + \alpha \pmod 1$



$I_a = [0, 1-\alpha[,\ I_b = [1-\alpha, 1[,\ aba$

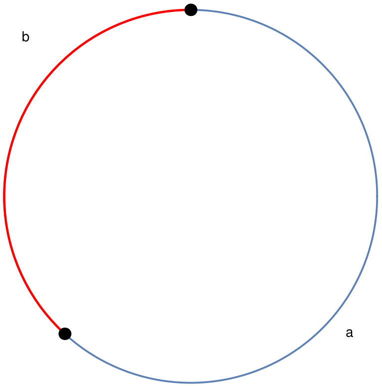
Réalisation de systèmes dynamiques discrets

$\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $R : \mathbb{T} \to \mathbb{T}$, $x \mapsto x + \alpha \pmod 1$



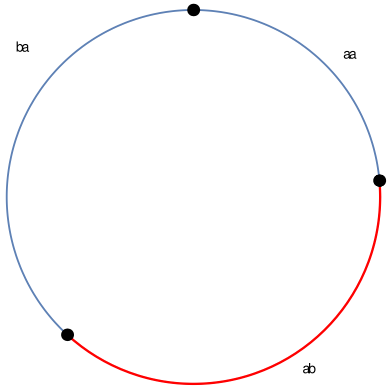
$I_a = [0, 1-\alpha[$, $I_b = [1-\alpha, 1[$, $abaa$

I_b



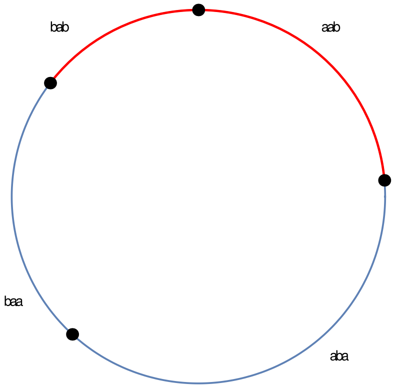
$I_a, \ I_b$

$R^{-1}(I_b)$



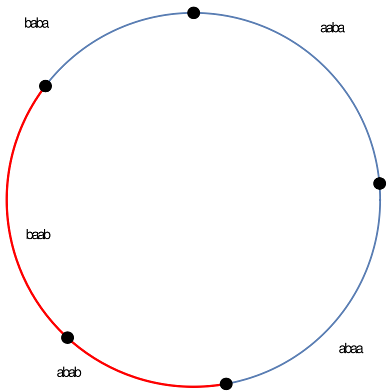
$I_a \cap R^{-1}(I_a),\ I_a \cap R^{-1}(I_b),\ I_b \cap R^{-1}(I_a)$

$R^{-2}(I_b)$



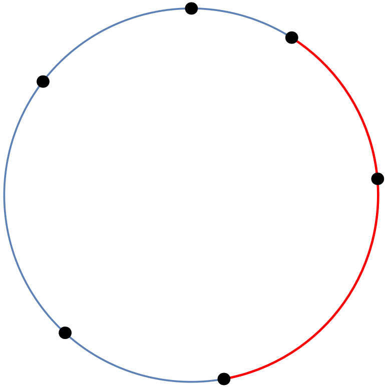
$$I_a \cap R^{-1}(I_a) \cap R^{-2}(I_b),\ I_a \cap R^{-1}(I_b) \cap R^{-2}(I_a),$$
$$I_b \cap R^{-1}(I_a) \cap R^{-2}(I_a),\ I_b \cap R^{-1}(I_a) \cap R^{-2}(I_b)$$

$R^{-3}(I_b)$



$I_{aab} \cap R^{-3}(I_a),\ I_{aba} \cap R^{-3}(I_a),\ I_{aba} \cap R^{-3}(I_b)$
$I_{baa} \cap R^{-3}(I_b),\ I_{bab} \cap R^{-3}(I_a)$

$R^{-4}(I_b)$



I_{aabaa}, I_{aabab}, I_{abaab},
I_{ababa}, I_{baaba}, I_{babaa}

On vient de démontrer que la complexité en facteurs d'un mot sturmien est exactement

$$\mathsf{p}(n) = n + 1, \forall n \geq 0.$$

Il s'agit même d'une caractérisation : mots non périodiques de complexité minimale (cf. thm. Morse–Hedlund).

$\rightsquigarrow$ Les "réels sturmiens" sont transcendants.

Des moyens de constructions simples :

- propriétés riches et variées
- preuves et techniques “raisonnables”

On peut même, dans certains, cas avoir des preuves automatiques !

M. Presburger (1929)

La théorie du premier ordre $Th(\langle \mathbb{N}, + \rangle)$ est décidable.

Preuve: $\langle \mathbb{N}, + \rangle$ admet l'élimination des quantificateurs

$\rightarrow$ *vérifier un nombre fini d'égalités (éventuellement modulo m) ou d'inégalités de combinaisons linéaires d'entiers et de variables.*

$$\boxed{\;=,\ (\exists x),\ \neg,\ \vee,\ (\forall x),\ \wedge,\ \rightarrow,\ \leftrightarrow,\ \leq, <}$$

$$x \leq y \equiv (\exists z)(x + z = y), \qquad x < y \equiv (x \leq y) \wedge \neg(x = y)$$

Exemple de formule close

$$(\exists x)(\exists y)(\forall z)\Big\{(x+y=z \vee x=y+y) \\ \rightarrow (\forall u)[(x=u) \vee \neg(y=u+z)]\Big\}$$

Exemple

La formule close suivante est vraie

$$(\forall x)(\exists y)[x=y+y \vee x=\mathcal{S}(y+y)]$$

où $\mathcal{S}(x)=y \equiv (x<y) \wedge (\forall z)(x<z \rightarrow (y \leq z))$.

On peut définir des *constantes*

$$x = 0 \equiv (\forall y)(x \leq y), \quad 1 = \mathcal{S}(0), \quad 2 = \mathcal{S}(1), \ldots$$

la *multiplication par une constante* et des *congruences*

$$2x \equiv x + x, \quad k.x = \underbrace{x + \cdots + x}_{k \text{ fois}},$$

$$x \equiv_k y \;\equiv\; (\exists z)(x = y + k.z \vee y = x + k.z).$$

Un exemple moins trivial (Frobenius)

Les Chicken McNuggets sont vendus par 6, 9, or 20.
Le plus grand nombre de nuggets qui ne peut pas être commandé est 43.

$$(\forall n)(n > 43 \to (\exists x, y, z \geq 0)(n = 6x + 9y + 20z))$$

$$\wedge \neg((\exists x, y, z \geq 0)(43 = 6x + 9y + 20z)).$$

On peut définir des parties de $\mathbb{N}$

Les nombres impairs

$$\varphi(x) \equiv (\exists y)[\overbrace{x}^{\text{variable libre}} = \mathcal{S}(y+y)]$$

$$\{n \in \mathbb{N} \mid \langle \mathbb{N}, + \rangle \models \varphi(n)\} = 2\mathbb{N} + 1$$

Remarque

Une partie de $\mathbb{N}$ est définissable dans $\langle \mathbb{N}, + \rangle$ SSI elle est ultimement périodique

Expansion

Soit $k \geq 2$ un entier.

J.R. Büchi 1960

$Th(\langle \mathbb{N}, +, V_k \rangle)$ est décidable.

$V_k(x)$ est la plus grande puissance de k divisant x;
$V_k(0) = 1$.

$$V_2(28) = 4, \quad V_2(64) = 64, \quad V_2(21) = 1.$$

Caractérisation des suites k-automatiques

Un mot infini $\mathbf{x}$ sur A est k-automatique SSI, pour tout $a \in A$, il existe une formule $\varphi_{\mathbf{x},a}(n)$ de $\langle \mathbb{N}, +, V_k \rangle$ définissant l'ensemble $\{n \geq 0 \mid \mathbf{x}(n) = a\}$.

Exemple 1 dans $\langle \mathbb{N}, + \rangle$

Soient $A = \{a, b\}$ et $\varphi_a(n) \equiv (\exists y)(n = 2y)$.
On obtient le mot $(ab)^\omega = abababab \cdots$ qui est k-automatique pour tout $k \geq 2$.

$$f : a \mapsto aba, \;\; b \mapsto bab$$

Exemple 2 dans $\langle \mathbb{N}, +, V_2 \rangle$

Soient $A = \{a, b, c\}$ et

$$\varphi_b(n) \equiv V_2(n) = n, \quad \varphi_c(n) \equiv (n \geq 1) \wedge \neg\varphi_b(n).$$

$$f : a \mapsto ab, \;\; b \mapsto bc, \;\; c \mapsto cc, \quad g : b \mapsto 1, \; a, c \mapsto 0$$

$$f^\omega(a) = abbcbcccbcccccccbcccc \cdots$$

$g(f^\omega(a))$ est la suite caractéristique de $\{2^n \mid n \geq 1\}$.

Sans ces techniques logiques (Honkala 1986)

Ultimate periodicity problem

INSTANCE: un morphisme k-uniforme f et un codage g
DECIDER si oui ou non $\mathbf{x} = g(f^\omega(a))$ est ultimement périodique ?

Puisque $\mathbf{x}$ est k-automatique, pour tout a dans A, on dispose d'une formule $\varphi_{\mathbf{x},a}(n)$ satisfaite ssi $\mathbf{x}(n) = a$.

$$\mathbf{x}(i) = \mathbf{x}(j) \equiv \bigvee_{a\in A} (\varphi_{\mathbf{x},a}(i) \wedge \varphi_{\mathbf{x},a}(j))$$

$$(\exists p)(\exists N)(\forall i \geq N)\, \mathbf{x}(i) = \mathbf{x}(i+p).$$

Que l'on peut alors décider par la théorie des automates.

Re-formulation par E. Charlier, N. Rampersad, J. Shallit (2012)

Théorème

Soit $k \geq 2$.
Si l'on peut exprimer une propriété d'une suite k-automatique $\mathbf{x}$ en utilisant :

des quantificateurs (au premier ordre), des opérateurs logiques, des variables entières, l'addition, la soustraction, l'indexation dans $\mathbf{x}$ et la comparaison d'entiers ou d'éléments de $\mathbf{x}$,

alors cette propriété est décidable.

A. Thue (1906)

Le mot de Thue–Morse $\mathbf{t}$ est sans chevauchement.

$$\neg(\exists i)(\exists \ell \geq 1)[(\forall j < \ell)(\mathbf{t}(i+j) = \mathbf{t}(i+\ell+j)) \wedge \mathbf{t}(i) = \mathbf{t}(i+2\ell)]$$

The proof of the theorem uses two lemmas.

LEMMA 2.2.5. *Let* $X = \{ab, ba\}$; *if* $x \in X^*$, *then* $axa \notin X^*$ *and* $bxb \notin X^*$.

Proof. By induction on $|x|$. If $|x| = 0$, then indeed $aa, bb \notin X^*$. Let $x \in X^*$, $x \neq 1$ and suppose $u = axa \in X^*$ (the case $bxb \in X^*$ is similar). Then $u = x_1 x_2 \cdots x_r$, with $x_1, \ldots, x_r \in X$; consequently $x_1 = ab$ and $x_r = ba$. Thus $u = abyba$ with $y = x_2 \cdots x_{r-1} \in X^*$. But now by induction $x = byb$ is not in X^*, contrary to the assumption. ■

LEMMA 2.2.6. *Let* $w \in A^+$. *If* w *has no overlapping factor, then* $\mu(w)$ *has no overlapping factor*.

Proof. Assume that $\mu(w)$ has an overlapping factor for some $w \in A^*$. We show that w also has an overlapping factor.

By asumption, there are $x, v, y \in A^*$, $c \in A$ with

$$\mu(w) = xcvcvcy$$

Note that $|cvcvc|$ is odd, but $\mu(w) \in X^*$ with $X = \{ab, ba\}$: therefore $|\mu(w)|$ is even and $|xy|$ is odd. Thus

- Either: $|x|$ is even, and $x, cvcv, cy \in X^*$,
- Or: $|x|$ is odd, and $xc, vcvc, y \in X^*$.

This implies that $|v|$ is odd, since otherwise we get from $cvcv \in X^*$ (resp. $vcvc \in X^*$) that both v, cvc are in X^*, which contradicts Lemma 2.2.5.

In the case $|x|$ is even, it follows that cv is in X^* and $w = rsst$ with $\mu(r) = x$, $\mu(s) = cv$, $\mu(t) = cy$. But then s and t start with the same letter c and ssc is an overlapping factor in w.

In the case $|x|$ is odd, similarly $vc \in X^*$, and $w = rsst$ with $\mu(r) = xc$, $\mu(s) = vc$, $\mu(t) = y$. Here r and s end with c and css is an overlapping factor in w. ■

Proof of Theorem 2.2.3. Assume that $\mathbf{t}$ has an occurrence of an overlapping factor. Then it occurs in a left factor $\mu^k(a)$ for some $k > 0$. On the other hand, since a has no overlapping factor, by iterated application of Lemma 2.2.6 no $\mu^n(a)(n \geqslant 0)$ has an overlapping factor. Contradiction. ■

De nombreuse propriétés pour les suites automatiques sont décidables :

- facteurs bordés (arbitrairement longs)
- mot récurrent
- récurrence linéaire
- $\mathrm{Fac}(\mathbf{x}) \subset \mathrm{Fac}(\mathbf{y})$
- $\mathrm{Fac}(\mathbf{x}) = \mathrm{Fac}(\mathbf{y})$
- existence d'un facteur non bordé de longueur n

$\vdots$

Remarque

Applicable aussi à d'autres familles de mots contenant le mot de Fibonacci.

Computer Science > Formal Languages and Automata Theory

Automatic Theorem Proving in Walnut

Hamoon Mousavi

(Submitted on 18 Mar 2016)

Walnut is a software package that implements a mechanical decision procedure for deciding certain combinatorial properties of some special words referred to as automatic words or automatic sequences. Walnut is written in Java and is open source. It is licensed under GNU General Public License.

Subjects: **Formal Languages and Automata Theory (cs.FL)**; Logic in Computer Science (cs.LO); Mathematical Software (cs.MS); Combinatorics (math.CO)

Cite as: **arXiv:1603.06017 [cs.FL]**

(or **arXiv:1603.06017v1 [cs.FL]** for this version)

Submission history

From: Seyyed Hamoon Mousavi Haji [view email]

[v1] Fri, 18 Mar 2016 23:53:10 GMT (684kb,D)

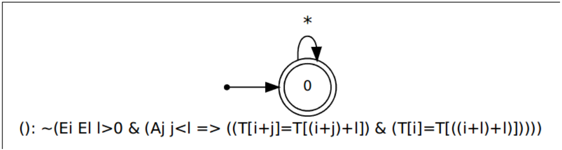
Preuve automatique que le mot de Thue–Morse est sans chevauchement

$$\neg(\exists i)(\exists \ell \geq 1)[(\forall j < \ell)(\mathbf{t}(i+j) = \mathbf{t}(i+\ell+j)) \wedge \mathbf{t}(i) = \mathbf{t}(i+2\ell)]$$

Jusqu'à 97 états pour une étape intermédiaire

```
rigo@X1:~/Walnut/Walnut/Walnut/bin$ java Main.prover
eval test "~(Ei El l>0 & (Aj j<l => ((T[i+j]=T[(i+j)+l]) & (T[i]=T[((i+l)+l)]))))":
l>0 has 2 states: 14ms
 j<l has 2 states: 0ms
  T[(i+j)]=T[((i+j)+l)] has 12 states: 136ms
   T[i]=T[((i+l)+l)] has 6 states: 39ms
    (T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)]) has 72 states: 41ms
     (j<l=>(T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)])) has 97 states: 62ms
      (A j (j<l=>(T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)]))) has 1 states: 186ms
       (l>0&(A j (j<l=>(T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)])))) has 1 states: 1ms
        (E l (l>0&(A j (j<l=>(T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)]))))) has 1 states: 0ms
         (E i (E l (l>0&(A j (j<l=>(T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)])))))) has 1 states: 1ms
          ~(E i (E l (l>0&(A j (j<l=>(T[(i+j)]=T[((i+j)+l)]&T[i]=T[((i+l)+l)])))))) has 1 states: 0ms
total computation time: 506ms
```

GraphViz / `xdot ../Result/test.gv`



Bibliographie succincte

- J.-P. Allouche, J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge Univ. Press (2003)
- V. Berthé, M.R., *Combinatorics, Automata and Number Theory*, Encycl. of Math. and its Appl. **135**, Cambridge Univ. Press (2010)
- V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and p-recognizable sets of integers, *Bull. Belg. Math. Soc.* **1** (1994), 191–238
- M.R., *Formal languages, Automata and Numeration Systems*, ISTE-Wiley (2014)