

THÉODORE TAPSOBA

Automates calculant la complexité de suites automatiques

Journal de Théorie des Nombres de Bordeaux, tome 6, n° 1 (1994),
p. 127-134

http://www.numdam.org/item?id=JTNB_1994__6_1_127_0

© Université Bordeaux 1, 1994, tous droits réservés.

L'accès aux archives de la revue « Journal de Théorie des Nombres de Bordeaux » (<http://jtnb.cedram.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

Automates calculant la complexité de suites automatiques

par THÉODORE TAPSOBA

RÉSUMÉ – Le point fixe u d'une substitution injective uniforme de module σ sur un alphabet A est examiné du point de vue du nombre $P(u, n)$ de ses blocs distincts de longueur n . Lorsque u est minimal et A de cardinal deux, nous construisons un automate pour la suite $n \rightarrow P(u, n+1) - P(u, n)$.

ABSTRACT – A fixed point u of an injective substitution of constant length σ on an alphabet A is considered in relation with the number $P(u, n)$ of its distinct n -blocks. When u is minimal and A a set of two elements, we prove that the sequence $n \rightarrow P(u, n+1) - P(u, n)$ is obtained by an automaton which is built explicitly.

1. Introduction

L'étude des mots infinis remonte au moins aux travaux de Thue ([13], [14]). Les régularités de ces mots ont dans un premier temps retenu l'attention ([13], [1], [7]) ; par la suite s'est posé le problème de la détermination du nombre $P(n)$ de facteurs distincts de longueur n . On a évidemment la majoration $P(n) \leq (\text{Card}(A))^n$. Pour les mots infinis qui sont points fixes de substitutions uniformes sur un alphabet A , A. Cobham a montré [4] qu'il existe une constante C , dépendant du mot infini, telle que $P(n) \leq Cn$ et N. Bleuzen-Guernalec [2] a précisé la constante en la majorant par $\sigma(\text{Card}(A))^2$ ou σ désigne le module de la substitution.

Nous construisons ici un automate pour la suite $n \rightarrow P(n+1) - P(n)$ lorsque le mot infini est minimal et point fixe d'une substitution injective uniforme sur un alphabet de cardinal deux.

2. Préliminaires

Soit A^* le monoïde engendré par un ensemble fini non vide appelé alphabet. Les éléments de A sont appelés lettres et ceux de A^* mots. Pour tout mot v de A , $|v|$ désigne la longueur de v , c'est-à-dire le nombre de ses lettres. L'élément neutre de A^* est le mot vide noté ε . C'est le mot de longueur zéro. Soit $a \in A$; on note simplement a^* pour l'ensemble $\{a\}^*$

des mots finis formés avec la seule lettre a . Le mot infini formé avec la lettre a est noté a^∞ .

Un mot v est dit facteur de w si $w = xvy$ avec x et y dans A^* . On écrit alors $v|w$. Si $x = \varepsilon$ (resp. $y = \varepsilon$), v est dit préfixe (resp. suffixe) de w .

$p_k(w)$ (resp. $s_k(w)$) désigne le préfixe (resp. le suffixe) de w de longueur k ($\leq |w|$). Un préfixe ou un suffixe de w est dit strict s'il est différent de w .

On appelle substitution une application $f : A \rightarrow A^*$. Cette application se prolonge de manière naturelle en morphisme de monoïde $A^* \rightarrow A^*$. Une substitution f est dite uniforme de module σ si $\sigma = |f(i)|$ pour toute lettre i de A , croissante si $|f(i)| \geq 2$. S'il existe une lettre a de A telle que $f(a) = am$ avec $|m| > 0$, alors l'ensemble des mots infinis de préfixe a possède un point fixe $u = amf(m)f^2(m) \cdots f^k(m) \cdots$. Une telle suite est automatique. On sait en effet ([5], [4]) que les suites projetées lettre à lettre de points fixes de substitutions de longueur constante sont exactement les suites engendrées par automates finis.

Soit u un mot fini ou infini. L'ensemble des facteurs finis de u est noté $F(u)$ et celui des facteurs de longueur n est noté $F_n(u)$. Il est clair que tout facteur d'un mot v de $F(u)$ est un mot de $F(u)$ et qu'il existe une lettre a telle que $va \in F(u)$. Le facteur v de u sera dit spécial si pour toute lettre i de A , vi est facteur de u . $FS(u)$ désigne l'ensemble des facteurs spéciaux de u et $FS_n(u)$ celui des facteurs spéciaux de longueur n .

Soit S le décalage défini par $S(a_0a_1a_2 \cdots) = a_1a_2 \cdots$ et soit Ω l'adhérence de l'ensemble $\{S^k(u), k \in \mathbb{N}\}$ où la distance d est définie par :

$$d(v, w) = \exp(-\inf\{n \in \mathbb{N}; v_n \neq w_n\}).$$

La suite u est associée au système dynamique (Ω, T) , (où T est la restriction de S à Ω), et est dite minimale si les seuls fermés invariants de Ω sous l'action de S sont l'ensemble vide et Ω .

La caractérisation suivante est classique [6] :

PROPOSITION 1. *Le mot u est minimal si et seulement si pour tout facteur m de u , il existe un entier j dépendant de m tel que :*

$$(1) \quad \forall k \in \mathbb{N}, m|u_k u_{k+1} \cdots u_{k+j}.$$

Remarque 1. La condition (1) de la proposition 1 exprime que tout mot du langage de u apparaît dans u avec des lacunes bornées (par j).

Lorsque u est point fixe d'une substitution f , nous avons (voir [10]) un critère de minimalité en fonction de f :

PROPOSITION 2. *Soit u point fixe de la substitution f sur l'alphabet A . Si a est préfixe de u avec $|f(a)| \geq 2$ et si toutes les lettres de A sont dans u , alors les propriétés suivantes sont équivalentes :*

- (i) u est minimal et $\lim |f^k(b)| = +\infty$ pour toutes les lettres b de A ,
- (ii) il existe $L \leq \text{Card}(A)$ tel que pour tout $b \in A$, $a|f^L(b)$,
- (iii) pour tout $b \in A$, il existe $k(b) \in \mathbb{N}$ tel que $a|f^{k(b)}(b)$.

Lorsque $A = \{1, 2\}$, on a un critère de minimalité très simple :

PROPOSITION 3. *Soit u un mot infini point fixe d'une substitution f sur $A = \{1, 2\}$ tel que 1 soit préfixe de u et $u \neq 1^\infty$.*

- (i) *Supposons f croissante, alors : u minimale $\Leftrightarrow f(2) \notin 2^*$.*
- (ii) *Supposons $|f(1)| \geq 2$ et $f(2) = 2$, alors : u minimale $\Leftrightarrow f(1) \in 1A^*1$.*

Preuve

(i) Si le mot est minimal, n'étant pas formé que de 1, il admet les facteurs $f^k(2)$. Donc si $f(2) \in 2^*$, le mot 1 apparaît dans u avec des lacunes arbitrairement grandes, contrairement à la proposition 1. Réciproquement si $f(2) \notin 2^*$, alors $1|f(2)$ et la minimalité résulte de la proposition 2.

(ii) Si 2 est suffixe de $f(1)$ alors on peut écrire $f(1) = 1B12^{(b)}$ avec $b \geq 1$ et plus généralement $f^k(1) = 1B_k12^{(kb)}$. Cette forme s'obtient par récurrence puisque $f^{k+1}(1) = (1B12^{(b)})f(B_k)(1B12^{(b)})2^{(kb)}$. Ainsi les mots 2^s , $s \geq 0$, sont-ils tous facteurs de u qui n'est donc pas minimale. Supposons maintenant que 1 soit suffixe de $f(1)$. Posons $r = |f(1)|$. Le mot $2^{(r-1)}$ n'est pas facteur de $f(1)$ et s'il n'est pas facteur d'un mot m , il n'est pas facteur du mot $f(m)$. Il en résulte que 1 apparaît dans u avec des lacunes bornées par $r - 1$. Par suite $f^k(1)$ apparaît aussi dans u avec des lacunes bornées et donc u est minimale. \square

Dans tout ce qui suit, u désigne une suite minimale non périodique, point fixe d'une substitution injective uniforme de module σ sur un alphabet A .

3. Quelques propriétés des facteurs et facteurs spéciaux

Soit m un facteur de u . Il peut se factoriser sous la forme :

$$(2) \quad m = xf(v)y$$

avec les conditions

- (3) x suffixe strict d'un mot $f(v_1)$, y préfixe strict d'un mot $f(v_2)$
et v_1vv_2 facteur de u .

Un facteur m de u est dit mot rythmé s'il n'existe qu'un seul triplet (v, x, y) vérifiant les conditions (2) et (3).

PROPOSITION 4. *S'il existe un facteur R de u rythmé et de longueur $\geq \sigma$, alors tout facteur de u dont R est un facteur est aussi un mot rythmé.*

Preuve

Soit R un facteur de u rythmé et de longueur σ . Soit m un facteur de u tel que R soit un facteur de m . La factorisation de u en blocs de σ lettres, c'est-à-dire $u = f(u_0)f(u_1)f(u_2)\dots$ donne une factorisation de R lu dans m . Celle-ci étant unique, elle détermine alors un préfixe B de R , de longueur $|B| < \sigma$ et tel que si $R = u_k u_{k+1} \dots u_{k+\sigma}$, alors $k + |B|$ est un multiple de σ . La position de $u_{k+|B|}$ dans R , et par suite dans m , fournit une factorisation de m induite par celle de u , mais celle-ci ne dépend pas de la lecture de m dans u . Soit $m = \alpha(m_1)(m_2)\dots(m_t)\beta$ cette factorisation unique où les mots m_i sont dans $f(A)$ et les mots α et β ont des longueurs strictement inférieures à σ . Le caractère injectif de f donne la factorisation rythmée de m . \square

PROPOSITION 5. *Tout suffixe d'un facteur spécial est spécial.*

Preuve

Soit $w = xv$ un facteur spécial. Pour toute lettre i de A , $xvi = wi$ est un facteur de u et vi est donc un facteur de u . Ainsi v est-il spécial. \square

Le corollaire suivant est immédiat :

COROLLAIRE. *Si $FS_p(u)$ est vide, alors pour tout $n \geq p$, $FS_n(u)$ est vide.*

Le résultat assez classique suivant, cité dans [11], a été démontré dans [12] :

PROPOSITION 6. *Il existe L_0 dépendant de σ et de $\text{Card}(A)$ tel que tout facteur m de u de longueur plus grande que L_0 soit un mot rythmé.*

Nous supposons à présent $A = \{1, 2\}$.

4. Calcul automatique

Posons $\alpha = |P|$ où P est le plus grand préfixe commun à $f(1)$ et $f(2)$ et montrons ceci :

LEMME. Soit k le plus petit entier tel que $\sigma k + \alpha \geq n$. Alors les facteurs spéciaux de longueur $n > L_0$ sont les suffixes de longueur $(n - \alpha)$ des images des facteurs spéciaux de longueur k auxquels on a concaténé à droite P .

Preuve

Montrons dans un premier temps que la construction ci-dessus donne des facteurs spéciaux : soit v un facteur spécial de longueur k . Alors, pour toute lettre i de A , vi est un facteur de u . Il en est de même pour son image qui est de longueur $\sigma k + \sigma$.

Le préfixe $f(v)P$ de longueur $\sigma k + \alpha$ de l'image $f(vi)$ est spécial par définition de P . Par la proposition 5, il en est de même pour tous ses suffixes.

Montrons à présent que tout facteur spécial de longueur $n > L_0$ peut être obtenu par la construction du lemme : soit w un facteur spécial de longueur $n > L_0$. Par la proposition 6, w est rythmé : $w = xf(v)y$, où x est un suffixe strict de l'image par f d'une lettre j et y préfixe strict de l'image par f d'une lettre k . Ainsi w est-il un facteur de $f(jvk)$ avec jvk facteur de u . Comme w est spécial, wi est un facteur de u pour toute lettre i de A . Chacun des $\text{Card}(A)$ facteurs wi est un préfixe d'exactly un $f(k)$, la correspondance étant bijective. On a donc $y = P$ et jv spécial. Finalement $w = xf(v)P$ où $xf(v)$ est un suffixe de $f(jv)$ avec jv spécial. \square

Remarque 2. Pour deux lettres distinctes i et j , désignons par $P_{i,j}$ le plus grand préfixe commun à $f(i)$ et $f(j)$. Pour un alphabet A de cardinal quelconque, si pour tout couple (i, j) de lettres distinctes de A on a $P_{i,j} = P$, (P étant le plus grand préfixe commun aux images de toutes les lettres de A), alors le lemme ci-dessus reste vrai.

Soit $P(n)$ le nombre de facteurs distincts de longueur n et soit $q(n) = P(n+1) - P(n)$. On a le résultat suivant :

THÉORÈME 1. Soit f une substitution injective uniforme de module σ sur $A = \{1, 2\}$, admettant un point fixe u minimal non périodique. Alors la suite $n \rightarrow q(n)$ ne prend qu'un nombre fini de valeurs.

Preuve

On a $q(n) = \text{Card}(F_{n+1}) - \text{Card}(F_n) = \text{Card}(FS_n)$. Soit L_0 la constante de la proposition 6 et soit $E = \{q(n) ; 1 \leq n \leq L_0\}$. Posons $\mu = |S|$ où S est le plus grand suffixe commun à $f(1)$ et $f(2)$. Si l'entier n est écrit sous la forme $n = \sigma k + \alpha + r$, ($1 \leq r \leq \sigma$), on sait (lemme ci-dessus) que les facteurs spéciaux de longueur n sont obtenus à partir des facteurs spéciaux de longueur $(k+1)$.

Deux facteurs spéciaux distincts de longueur $(k+1)$ ayant le même suffixe de longueur k donneront (par la construction du lemme) le même facteur spécial de longueur n si $r \leq \mu$. Pour $r > \mu$, il y a autant de facteurs spéciaux de longueur n que de facteurs spéciaux de longueur $(k+1)$. Il vient donc que :

$$\text{Card}(FS_n) = \sum_i \text{Card} \{s_r(f(a)), a = p_1(m), m \in FS_{k+1}, s_k(m) = m_i\},$$

$$(\leq \text{Card}(FS_{k+1})).$$

L'entier $q(n)$ est donc parfaitement déterminé dès que $q(k+1)$ et r sont connus. Si $(k+1) > L_0$, on lui applique le lemme pour obtenir $q(k+1)$. Ainsi pour tout $n > L_0$, $q(n) \in E$. \square

Pour $n > L_0$, si Φ_r désigne l'application de E dans E qui à $q(n)$ fait correspondre $q(k+1)$ lorsque $n = \sigma k + \alpha + r$, ($1 \leq r \leq \sigma$), et $[x]$ la partie entière du réel x , on a donc le résultat suivant :

THÉORÈME 2. *Pour tout entier $n > L_0$, il existe un σ -automate (dont on a un programme formel ci-dessous) qui donne la valeur de $q(n)$ à partir de la décomposition de n sous la forme $n = \sigma k + \alpha + r$, ($1 \leq r \leq \sigma$) :*

```

 $n := n_{\text{initial}}$ 
 $\Phi :=$  Application identité dans  $E$ 
Tant que  $n > L_0$ , faire
 $k \leftarrow \begin{cases} (n - \alpha)\sigma^{-1} & \text{si entier} \\ 1 + [(n - \alpha)\sigma^{-1}] & \text{sinon} \end{cases}$ 
 $r \leftarrow n - \sigma k + 1$ 
 $\Phi \leftarrow \Phi \circ \Phi_r$ 
 $n \leftarrow k$ 
Fin de faire
 $q(n_{\text{initial}}) := \Phi(q(n))$ .

```

5. Exemple

Soit f la substitution sur $\{1, 2\}$ définie par $f(1) = 12$ et $f(2) = 21$. f est injective. Le point fixe de f de préfixe 1 est un mot infini u minimal et non périodique. C'est la célèbre suite de Morse [9]. On observe facilement que les seuls mots rythmés dans u de longueur 3 sont 221, 122, 112, 211, les autres ne sont pas rythmés (121, 212) ou ne sont pas des facteurs (222, 111). Soit m un mot de longueur 4. Supposons qu'il n'admette pas de facteur rythmé de longueur 3. Alors 121 ou 212 sont les seuls facteurs possibles de

longueur 3 de sorte que $m = 1212$ ou $m = 2121$. On vérifie directement que ces deux mots sont des facteurs de u .

Supposons $m = 1212 = 1f(2)2$. Étant facteur de u , le mot $21f(2)21 = f(222)$ est aussi facteur de u et ne provient que de l'image par f de 222 . Par suite le mot 222 est lui aussi facteur de u , ce qui est faux. Ainsi 1212 est rythmé, de même que 2121 . Les autres facteurs de longueur 4 contiennent nécessairement l'un des quatre mots rythmés de longueur 3. Finalement on a montré le résultat suivant (dont une version légèrement différente de trouve dans [10]) :

PROPOSITION 7. *Dans le mot de Morse, tout facteur de longueur supérieure ou égale à quatre est rythmé et la valeur quatre est optimale.*

Un dénombrement direct donne $q(1) = q(2) = 2$ et $q(3) = 4$; on a alors $E = \{2, 4\}$. Notons que $f(1)$ et $f(2)$ n'ont ni suffixe commun ni préfixe commun. Ainsi pour $n > 4$, $q(n) = \text{Card}(FS_n) = \text{Card}(FS_{k+1})$ où $2k + r = n$ et $r = 1$ ou 2 . Donc $\Phi_1 = \Phi_2 = \text{Identité}$ dans E . Pour tout $n \geq 4$, il existe un unique entier k tel que $2^{k+1} < n \leq 2^{k+2}$. Si $2^{k+1} + 2^k < n \leq 2^{k+2}$, nous avons alors $3 < n2^{-k} \leq 4$ donc $4 = n2^{-k}$ si $n2^{-k}$ est entier et $4 = [n2^{-k}] + 1$ sinon et, par suite, $q(n) = q(4) = 2$. Si $2^{k+1} < n \leq 2^{k+1} + 2^k$, nous avons maintenant $2 < n2^{-k} \leq 3$ et le raisonnement précédent donne $q(n) = q(3) = 4$. En remarquant que $3.2^{-2} < 1 \leq 4.2^{-2}$, $3.2^{-1} < 2 \leq 4.2^{-1}$ et $3.2^0 < 4 \leq 4.2^0$, on a en définitive $q(n) = 2$ pour les naturels n pour lesquels il existe un entier relatif k tel que $3.2^{-k} < n \leq 4.2^{-k}$ et $q(n) = 4$ sinon.

Remarque 3. Le calcul automatique ci-dessus permet de retrouver (très rapidement) la formule explicite pour l'énumération des facteurs de la suite de Thue-Morse obtenue par Brlek [3] et par A. De Luca et S. Varricchio [8].

Note : B. Mossé a obtenu récemment des résultats qui généralisent ceux que nous avons exposés ici.

BIBLIOGRAPHIE

- [1] S. Arson, *Démonstration de l'existence de suites asymétriques infinies*, Mat. Sb. **44** (1937), 769–777.
- [2] N. Bleuzen-Guernalec, *Suites points fixes de transductions uniformes*, C. R. Acad. Sci. Paris, Série I **300** (1985), 85–88.
- [3] S. Brlek, *Enumeration of factors in the Thue-Morse word*, Discrete Applied Math. **24** (1989), 83–96.

- [4] G. Christol, T. Kamae, M. Mendès France et G. Rauzy, *Suites algébriques, automates et substitutions*, Bull. Soc. math. France 108 (1980), 401–419.
- [5] A. Cobham, *Uniform tag Sequences*, Math. Systems Theory 6 (1972), 164–192.
- [6] W. H. Gottschalk and G. A. Hedlund, *Topological dynamics*, Am. Math. Soc. Colloq. Publ. 36, Providence R. I. (1968).
- [7] Lothaire, *Combinatorics on words*, Addison Wesley MA (1982), chapter 12.
- [8] A. de Luca and S. Varricchio, *Some combinatorial properties of the Thue-Morse sequence and a problem in semigroups*, Theoret. Comput. Sci. 63 (1989), 333–348.
- [9] M. Morse, *Recurrent geodesic on a surface of negative curvature*, Trans. Amer. Math. Soc. 22 (1921), 84–100.
- [10] M. Queffélec, *Contribution à l'étude spectrale de suites arithmétiques*, Thèse d'État, Paris-Nord, (1984).
- [11] G. Rauzy, *Rotation sur les groupes, nombres algébriques et substitutions*, Séminaire de Théorie des Nombres, Bordeaux, exposé 21 (1987-1988), 21-1–21-12.
- [12] T. Tapsoba, *Complexité de suites automatiques*, Thèse de troisième cycle, Université Aix-Marseille II (1987).
- [13] A. Thue, *Über unendliche Zeichenreihen*, Norske Vid. Skr. I. Math. Kl., Christiana 7 (1906), 1–22.
- [14] A. Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Norske Vid. Selsk. Skr. I. Math. Nat. Kl., Christiana 1 (1912), 1–67.

Théodore Tapsoba
Département de Mathématiques
Faculté des Sciences et Techniques
Université de Ouagadougou
03 B. P. 7021 Ouagadougou 03
Burkina Faso