Systèmes dynamiques symboliques sturmiens et substitutifs

Valérie BERTHÉ

Contents

Une mesure classique du désordre pour une suite à valeurs dans un alphabet fini, consiste à compter le nombre $p(n)$ de facteurs de longueur n. On définit ainsi une fonction p appelée fonction de complexité. Cette fonction permet de caractériser en particulier les suites périodiques (ce sont les suites pour lesquelles il existe un entier n tel que $p(n) \leq n$), ainsi que certaines suites de basse complexité, comme les suites de complexité $n+1$. Celles-ci sont appelées suites sturmiennes et possèdent la particularité remarquable de pouvoir être étudiées tant d'un point de vue combinatoire qu'arithmétique : elles sont en effet obtenues comme des codages de rotation irrationnelle sur le cercle unité. Les suites substitutives constituent une deuxième famille classique de suites de basse complexité. Nous étudions en particulier les substitutions de longueur constante en relation avec les suites automatiques au paragraphe 7. Nous évoquons les propriétés combinatoires de ces suites en introduisant des outils classiques en combinatoire des mots et dans l'étude des systèmes dynamiques symboliques. Notons que ces suites interviennent non seulement dans de nombreux domaines des mathématiques mais aussi en informatique théorique, en biologie, en physique, en cristallographie ...

1 Quelques définitions

Le but de ce paragraphe est d'introduire brièvement quelques notions dynamiques et combinatoires. Pour plus d'informations, on peut par exemple consulter l'incontournable [43]. Voir aussi [33].

Soit $\mathcal{A}$ un ensemble fini appelé *alphabet*. On peut faire agir sur l'ensemble des suites unidirectionnelles $\mathcal{A}^{\mathbb{N}}$ ou bidirectionnelles $\mathcal{A}^{\mathbb{Z}}$ l'application de *décalage* T qui à la suite $(u_n)_n$ associe la suite $(u_{n+1})_n$. Travaillons par exemple avec $\mathcal{A}^{\mathbb{N}}$. Munissons $\mathcal{A}^{\mathbb{N}}$ du produit des topologies discrètes. Cet ensemble est alors compact. Cette topologie est équivalente à la topologie définie par la métrique suivante : si $x, y \in \mathcal{A}^{\mathbb{N}}$ et $x \neq y$

$$d(x,y) = (1 + \inf\{k \geq 0;\ x_k \neq y_k\})^{-1}.$$

Deux suites sont donc d'autant plus proches que leurs premiers termes coïncident longtemps. Le *cylindre* $[w]$, où $w = w_1 \ldots w_n$ appartient à $\mathcal{A}^n$, est l'ensemble des suites de la forme

$$[w] = \{x \in \mathcal{A}^{\mathbb{N}} |\ x_0 = w_1,\ x_1 = w_2, \ldots, x_{n-1} = w_n\}.$$

Les cylindres sont des ensembles ouverts et fermés et engendrent la topologie. L'ensemble $\mathcal{A}^{\mathbb{N}}$ est alors complet car compact et métrique.

Soit $u \in \mathcal{A}^{\mathbb{N}}$. La suite u engendre alors le *système dynamique symbolique* $(\overline{\mathcal{O}(u)}, T)$, où $\overline{\mathcal{O}(u)}$ est l'adhérence de l'orbite $\mathcal{O}(u) = \{T^n(u),\ n \geq 0\}$, sous l'action du décalage T, de la suite u dans $\mathcal{A}^{\mathbb{N}}$. L'ensemble $\overline{\mathcal{O}(u)}$ est à son tour, compact, métrique et complet; il est T*-invariant*: $T(\overline{\mathcal{O}(u)}) \subset \overline{\mathcal{O}(u)}$. En d'autres termes, T agit sur $\overline{\mathcal{O}}(u)$.

Plus généralement, on appelle *système dynamique symbolique* l'action du décalage sur un fermé invariant de $\mathcal{A}^{\mathbb{N}}$. Le décalage T est une application uniformément continue, surjective, bijective sur $\mathcal{A}^{\mathbb{Z}}$, mais pas forcément injective sur $\mathcal{A}^{\mathbb{N}}$.

Rappelons que si T est une application continue agissant sur le compact X, alors le système (X, T) est appelé *système dynamique topologique.*

De nombreuses propriétés combinatoires de la suite u se traduisent en termes dynamiques. On appelle *facteur* de la suite infinie u un bloc fini w de lettres apparaissant de manière consécutive dans la suite : $w = u_{n+1} \cdots u_{n+d}$; d est appelé la longueur de w et est noté $|w|$.

Une suite est dite *récurrente* si chacun de ses facteurs apparaît une infinité de fois.

Exercice 1.1 *1. Montrer que*

$$\overline{\mathcal{O}}(u) = \{x \in \mathcal{A}^{\mathbb{N}},\ L(x) \subset L(u)\},$$

où $L(x)$ est l'ensemble des facteurs de la suite x.

2. Montrer que u est récurrente si et seulement s'il existe une suite strictement croissante (n_k) telle que

$$u = \lim_{k \to +\infty} T^{n_k} u.$$

En déduire que u est récurrente si et seulement si T est surjective sur $\overline{\mathcal{O}}(u)$.

Une suite est dite *uniformément récurrente* si les lacunes entre deux occurrences d'un même facteur sont bornées.

Il n'est alors pas difficile de montrer que le système $(\overline{\mathcal{O}(u)}, T)$ est *minimal* (c'est-à-dire qu'il ne contient pas de fermé invariant sous l'action de T non trivial) si et seulement si la suite u est uniformément récurrente : en effet, si w est un facteur de u, alors

$$\overline{\mathcal{O}(u)} = \bigcup_{n\in\mathbb{N}} T^{-n}[w],$$

et on conclut par un argument de compacité; pour plus de détails, voir par exemple [43]. C'est pour cette raison que les suites "uniformément récurrentes" sont également appelées "suites minimales".

Exercice 1.2 *Montrer que* (X,T) *est minimal si et seulement si* $X = \overline{\mathcal{O}(x)}$*, pour tout élément* x *de* X.

Il est intéressant de munir d'une mesure T-invariante le système dynamique topologique $(\overline{\mathcal{O}(u)}, T)$. (Une mesure de probabilité borélienne μ est dite T-invariante si $\mu(T^{-1}(B)) = \mu(B)$, pour tout borélien B.) La connaissance des fréquences de blocs de la suite u permet ainsi de définir une telle mesure.

La *fréquence* $f(w)$ du facteur w est définie comme la limite (si elle existe) du quotient du nombre d'occurrences de ce facteur dans les n premiers termes de la suite, par n. On suppose que les fréquences de tous les facteurs existent. On définit alors une mesure de probabilité μ sur la famille $\mathcal{B}$ des boréliens de $\overline{\mathcal{O}(u)}$, de la manière suivante : la mesure μ est l'unique mesure de probabilité invariante par T, telle que $\mu([w]) = f(w)$, où $[w]$ est le cylindre des suites de $\overline{\mathcal{O}(u)}$ de préfixe w et où $f(w)$ est la fréquence d'apparition du bloc w dans la suite u. En effet, les cylindres sont à la fois ouverts et fermés et engendrent la topologie; il suffit alors d'appliquer le théorème de compatibilité de Daniell-Kolmogorov (voir par exemple [47]).

La plupart des systèmes que nous considérerons sont uniquement ergodiques : un système dynamique est *uniquement ergodique* s'il existe une unique mesure de probabilité T-invariante. D'où l'intérêt de l'étude des fréquences des facteurs qui permet de définir explicitement cette unique mesure T-invariante.

Notons qu'il existe des suites minimales qui engendrent des systèmes non uniquement ergodiques; dans un tel cas, les fréquences des facteurs peuvent ne pas exister.

2 Fonction de Complexité

La fonction de complexité est un outil très utile à l'étude des suites et des systèmes dynamiques symboliques. Soit $u = (u_n)_n$ une suite à valeurs dans l'alphabet fini $\mathcal{A}$. On appelle *fonction de complexité* de u, et l'on note p, la fonction (définie sur les entiers) qui compte le nombre de facteurs de u de longueur donnée :

$$p(n) = \mathrm{Card}\{w;\ w \text{ est facteur de } u \text{ et } |w| = n\}.$$

Pour plus d'informations sur la fonction de complexité, voir par exemple les survols [3, 34] et le cours [45].

Il est facile de voir que la fonction de complexité est croissante et que pour tout entier n, on a $1 \leq p(n) \leq d^n$, où d est le cardinal de l'alphabet.

Cette fonction peut être considérée comme une *mesure de la prédictibilité* de la suite. La différence première de la fonction de complexité $p(n+1) - p(n)$ compte le nombre d'extensions possibles dans la suite des facteurs de longueur n. On appelle *prolongement à droite* (respectivement *prolongement à gauche*) d'un facteur w une lettre x telle que wx (respectivement xw) est un facteur de la suite. Soit w^+ (respectivement w^-) le nombre de prolongements à droite (respectivement à gauche) de w. (On peut avoir éventuellement $w^- = 0$ si la suite est unidirectionnelle mais l'on a toujours $w^+ \geq 1$; si la suite est bidirectionnelle, alors $w^- \geq 1$.) On a alors

$$p(n+1) = \sum_{|w|=n} w^+ = \sum_{|w|=n} w^-,$$

et

$$p(n+1) - p(n) = \sum_{|w|=n} (w^+ - 1) = \sum_{|w|=n} (w^- - 1).$$

La fonction de complexité permet de caractériser les suites périodiques. On a en effet le résultat classique suivant (voir [40, 24]) :

$$\exists n,\ \ p(n) \leq n \Longleftrightarrow \exists C,\ \forall n\ \ p(n) \leq C \Longleftrightarrow u \text{ est périodique.}$$

Plus précisément, si la suite u est unidirectionnelle, c'est-à-dire indexée par $\mathbb{N}$, la suite est alors *ultimement périodique* ($\exists n_0,\ \exists T > 0,\ \forall n \geq n_0,\ u_{n+T} = u_n$) et si la suite u est bidirectionnelle, c'est-à-dire indexée par $\mathbb{Z}$, la suite est *strictement périodique* ($\exists T > 0,\ \forall n,\ u_{n+T} = u_n$). La preuve de ce résultat [24] repose sur le fait que s'il existe n tel que $p(n) \leq n$, alors il existe n_0 tel que $p(n_0+1) = p(n_0)$, et tout facteur de longueur n_0 admet un unique prolongement à droite. Pour tout entier $k \geq 1$, le "futur" de la suite, c'est-à-dire $(u_m)_{m\geq k}$, est donc déterminé par la donnée de $u_{k-n_0} \ldots u_{k-1}$. Il suffit de considérer un facteur de longueur n_0 qui apparaît deux fois dans la suite pour en déduire la périodicité. Notons que la période est alors inférieure ou égale à n_0.

La fonction de complexité permet d'exprimer simplement l'entropie topologique du système $(\overline{\mathcal{O}(u)}, T)$. Rappelons la définition de *l'entropie topologique* $h(u)$ d'une suite u, ou plus généralement du *système dynamique symbolique* associé $(\overline{\mathcal{O}(u)}, T)$:

$$h(u) = \lim_{n\to+\infty} \frac{\log_d(p(n))}{n},$$

où d est le cardinal de l'alphabet sur lequel la suite est définie. Cette limite existe du fait de la sous-additivité de la fonction $n \mapsto \log(p(n))$. Les suites d'entropie topologique nulle sont dites suites *déterministes*. Les suites que nous étudierons dans ce texte sont déterministes. Nous considérons principalement deux familles de suites : les suites engendrées par substitution (paragraphe 3) et les suites de complexité sturmienne (paragraphe 5).

L'étude de la complexité conduit en particulier aux trois questions suivantes :

- Comment calculer la complexité d'une suite ?
- Quelles fonctions peuvent être réalisées comme des fonctions de complexité pour une suite ?
- Peut-on caractériser des familles de suites par leur complexité; peut-on déduire de la complexité une représentation géométrique de certaines suites ?

Nous allons voir comment résoudre la première question en considérant les facteurs spéciaux d'une suite, au paragraphe 4. En particulier, une question naturelle est de savoir si toutes les fonctions affines peuvent être réalisées (éventuellement ultimement) comme fonctions de complexité. La réponse est affirmative (voir par exemple [1, 19]). Néanmoins, la seconde question est loin d'être résolue, surout dans le cas de l'entropie positive (voir par exemple [19, 34]). Bien que la fonction de complexité soit en général insuffisante pour décrire des suites, nous allons voir que dans le cas des suites sturmiennes (paragraphe 5), beaucoup d'informations peuvent se déduire de la connaissance de la fonction de complexité : les suites sturmiennes sont les suites de complexité $n+1$ indexées par $\mathbb{N}$.

Notons que l'ordre de croissance de la fonction de complexité est une caractéristique fondamentale de la suite. C'est en particulier un invariant topologique (voir par exemple [34]) : une conjugaison *topologique* entre deux systèmes dynamiques topologiques (X, T) et (Y, S) est une bijection bicontinue ϕ telle que $\phi T = S\phi$; or on voit sans trop de difficultés qu'une conjugaison topologique entre deux systèmes dynamiques symboliques est un code fini : il existe un entier r tel que $(\phi(x))_i$ ne dépend que de $(x_i \dots x_{i+r})$. Pour plus de détails, voir le cours [33]. Néanmoins la complexité n'est pas invariante par isomorphisme mesuré. Pour une notion de complexité dont l'ordre de croissance est invariant par isomorphisme mesuré, voir [32]. Voir également [35, 37] pour différentes notions de complexité pour des mots finis, qui permettent par exemple d'obtenir de l'information sur le langage de suites infinies à travers l'étude de la complexité de leurs préfixes.

3 Substitutions

On appelle *substitution* ou *morphisme* un endomorphisme non effaçant pour la concaténation du monoïde libre $\mathcal{A}^*$, formé des mots finis sur $\mathcal{A}$. Un exemple classique est donné par la substitution de Fibonacci : $\sigma(a) = ab$, $\sigma(b) = a$. On peut alors étendre naturellement la définition d'une substitution à l'ensemble des suites $\mathcal{A}^{\mathbb{N}}$.

Soit σ une substitution. Supposons qu'il existe une lettre a telle que $\sigma(a)$ commence par a et $|\sigma(a)| \geq 2$; on note $\overline{a}$ la suite de terme constant a; alors la suite $(\sigma^n(\overline{a}))_{n\in\mathbb{N}}$ converge dans $\mathcal{A}^{\mathbb{N}}$ vers un point fixe de σ. On appelle *suite substitutive* l'image par une projection littérale (c'est-à-dire un morphisme de monoïde libre dont les images des lettres sont des lettres) d'un point fixe de substitution.

De nombreuses propriétés topologiques et ergodiques des suites substitutives sont décrites par leur *matrice d'incidence* : on associe à la substitution σ une matrice $M = [m_{i,j}]_{(i,j)\in\mathcal{A}^2}$ dont le terme général $m_{i,j}$ compte le nombre d'occurrences de la lettre j dans $\sigma(i)$. Une substitution est dite *primitive* si la matrice associée l'est (il existe une puissance de la matrice dont toutes les entrées sont strictement positives, en d'autres termes, il existe un itéré de la substitution tel que l'image de toute lettre contient toutes les lettres de l'alphabet). Le système dynamique engendré par une suite u point fixe d'une substitution σ primitive (c'est-à-dire que $\sigma(u) = u$) est alors minimal et uniquement ergodique. Pour une approche dynamique et spectrale des substitutions, voir également [43].

Exercice 3.1 *Montrer que si σ est une susbtitution primitive alors il existe une suite u et un entier k tel que $\sigma^k(u) = u$.*

4 Facteurs spéciaux et complexité

Nous allons voir que la notion de facteur spécial est un outil efficace pour déterminer la différence première de la fonction de complexité. En effet, dans le cas d'une complexité basse, le nombre de facteurs spéciaux est en général relativement aisé à déterminer. Pour plus de détails, voir [19].

Soit u une suite à valeurs dans l'alphabet $\mathcal{A}$. Soit w^+ (respectivement w^-) le nombre d'extensions à droite (respectivement à gauche) de w. Rappelons que

$$p(n+1) - p(n) = \sum_{|w|=n} (w^+ - 1) = \sum_{|w|=n} (w^- - 1).$$

Un facteur est dit *facteur spécial à droite* (respectivement *facteur spécial à gauche*) s'il admet plus d'une extension à droite (respectivement à gauche).

Exercice 4.1 *Soit u la suite de Thue-Morse définie comme le point fixe commençant par 0 de la substitution suivante : $\sigma(0) = 01$ and $\sigma(1) = 10$.*

1. *Montrer que chaque facteur w peut être décomposé comme $w = r_1\sigma(x)r_2$, où x est un facteur et $r_i \in \{\varepsilon, a, b\}$. Si $|w| \geq 5$, alors cette décomposition est unique.*

2. *Montrer que $p(2n) = p(n) + p(n+1)$ et que $p(2n+1) = 2p(n+1)$, pour $n \geq 1$. En déduire une expression de la fonction de complexité (voir par exemple [15]).*

Exercice 4.2 *La suite de Fibonacci est définie comme le point fixe commençant par a de la substitution suivante : $\sigma(a) = ab$ and $\sigma(b) = a$.*

1. *Montrer que tout facteur w de la suite de Fibonacci peut être décomposé uniquement de la manière suivante :*

$$w = r_1\sigma(v)r_2,$$

où v est un facteur de la suite de Fibonacci, $r_1 \in \{\varepsilon, b\}$, et $r_2 = a$, si la dernière lettre de w est a, et $r_2 = \varepsilon$, sinon.

2. *Montrer que si w un facteur spécial gauche non vide, alors il existe un unique facteur spécial gauche non vide v tel que $w = \sigma(v)r_2$, où $r_2 = a$, si la dernière lettre de W est a, et $s = \varepsilon$, sinon. En déduire la forme générale des facteurs spéciaux gauche.*

3. *Montrer que la suite de Fibonacci n'est pas ultimement périodique.*

4. *En déduire que la fonction de complexité de la suite de Fibonacci satisfait $p(n) = n + 1$ pour tout n.*

5 Suites sturmiennes

L'étude des suites sturmiennes illustre de manière frappante les relations qui existent entre combinatoire des mots et théorie ergodique, ce qui est naturel puisque l'on travaille sur des suites, mais aussi avec l'arithmétique, l'approximation diophantienne et la géométrie. En particulier le développement en fraction continue permet de décrire avec précision de nombreuses propriétés des suites sturmiennes. Pour une approche plus géométrique des suites sturmiennes, voir par exemple [6, 7, 8]. Pour des survols récents, voir [10, 16, 38].

On a vu que si la complexité d'une suite est telle qu'il existe un entier n pour lequel $p(n) \leq n$, alors cette suite u est périodique. Il apparaît alors naturel de s'intéresser aux suites de complexité $n + 1$, c'est-à-dire telles que $p(n) = n + 1$, pour tout n. De telles suites existent sur $\mathbb{N}$ (par exemple, la suite de Fibonacci) et sur $\mathbb{Z}$; la suite suivante a pour complexité $n + 1$

$$\ldots 000010000 \ldots$$

Rappelons, que les suites de complexité $n+1$ indexées par $\mathbb{N}$ sont appelées *suites sturmiennes.* Les suites sturmiennes sont donc les suites de complexité minimale parmi les suites sur $\mathbb{N}$ non ultimement périodiques.

Exercice 5.1 *Montrer que pour toute suite sturmienne, chaque préfixe apparaît au moins deux fois dans la suite. En déduire que tout suite sturmienne est récurrente.*

L'exemple le plus classique de suite sturmienne est la suite de Fibonacci, point fixe de la substitution σ définie par $\sigma(a) = ab$ et $\sigma(b) = a$. Les suites sturmiennes sont donc définies de manière purement combinatoire, mais ce qui est remarquable, c'est qu'elles peuvent être également représentées de manière géométrique (voir [41]) : les suites sturmiennes sont exactement les suites obtenues en codant l'orbite d'un point ρ du cercle unité sous la rotation d'angle irrationnel α, par rapport à des intervalles complémentaires du cercle unité de longueurs α et $1 - \alpha$.

Dans tout ce qui suit R_α désigne la rotation définie sur le tore $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$ de dimension 1, par $R_\alpha x = x + \alpha$ modulo 1. S'il n'y a pas d'ambiguïté, nous omettrons de préciser que nous travaillons modulo 1.

Théorème 5.2 (Morse-Hedlund) *Soit u une suite sturmienne à valeurs dans $\{0, 1\}$. Il existe alors α irrationnel dans $]0, 1[$ et $\rho \in \mathbb{R}$ tels que l'on ait soit*

$$\forall n,\ (u_n = 0 \Longleftrightarrow R_\alpha^n(\rho) = \rho + n\alpha \in [0, 1 - \alpha[),$$

soit

$$\forall n,\ (u_n = 0 \Longleftrightarrow R_\alpha^n(\rho) = \rho + n\alpha \in\ ]0, 1 - \alpha]).$$

On appelle angle *d'une suite sturmienne le réel α qui lui est ainsi associé.*

Un des intérêts de la représentation géométrique des suites sturmiennes indiquée dans le théorème 5.2, est qu'elle fournit une description simple, en termes d'intervalles du cercle unité, des facteurs de longueur donnée. Rappelons la propriété classique suivante (voir par exemple [39]).

Lemme 5.3 *Soit u une suite sturmienne d'angle α. Il existe une bijection entre les facteurs de longueur n de la suite u et les intervalles de la partition du cercle unité par les points $-k\alpha$, $0 \leq k \leq n$.*

En effet, supposons que u code l'orbite du point ρ par rapport à $I_0 = [0, 1-\alpha[$ et $I_1 = [1 - \alpha, 1[$. Un mot fini $w_1 \cdots w_n$ défini sur l'alphabet $\{0, 1\}$ est un facteur de la suite u si et seulement s'il existe un entier k tel que

$$\rho + k\alpha \in I(w_1, \ldots, w_n) = \bigcap_{j=0}^{n-1} R^{-j}(I_{w_{j+1}}).$$

Comme α est irrationnel, la suite $(\rho + n\alpha)_{n\in\mathbb{N}}$ est dense, ce qui implique que $w_1 w_2 \dots w_n$ est un facteur de u si et seulement si $I(w_1, \dots, w_n) \neq \emptyset$. En particulier, l'ensemble des facteurs ne dépend pas du point initial ρ. On vérifie de plus que les ensembles $I(w_1, \dots, w_n)$ sont connexes et bornés par les points $k(1-\alpha)$, pour $0 \leq k \leq n-1$. Il y a $n+1$ tels intervalles (α est irrationnel) et alors $n+1$ facteurs de longueur n : la suite u est bien sturmienne.

Exercice 5.4 *Soit u une suite sturmienne d'angle α. Montrer que $\overline{\mathcal{O}(u)}$ est l'ensemble de suites sturmiennes de même angle α. En déduire la minimalité du système dynamique $(\overline{\mathcal{O}(u)}, T)$.*

Notons que les suites sturmiennes ont de nombreuses autres caractérisations, tant géométriques que combinatoires (voir, par exemple, [38]). En particulier, les suites sturmiennes sont les suites équilibrées sur un alphabet à deux lettres qui sont non ultimement périodiques. Une suite *équilibrée* est telle que la différence entre le nombre d'occurrences d'une lettre dans deux de ses facteurs de même longueur, est bornée par 1 en valeur absolue.

6 Graphe des mots

Un outil très utile pour l'étude des suite sturmiennes (et en particulier pour l'étude des fréquences des facteurs) est le *graphe des mots* de Rauzy. Soit u une suite définie sur l'alphabet fini $\mathcal{A}$ (de cardinal d). Le graphe des mots Γ_n des facteurs de longueur n de la suite u est un graphe orienté (voir, par exemple, [44]), qui est un sous-graphe du graphe des mots de de Bruijn[1] (voir [17]).

Le graphe Γ_n a pour sommets les facteurs de longueur n de la suite, avec une arête de U vers V, s'il existe un mot W de longueur $n-1$ tel que :

$$U = xW \text{ et } V = Wy, \text{ avec } x, y \in \mathcal{A},$$

et tel que xWy soit un facteur de la suite. Considérons une suite sturmienne. De la complexité ($\forall n,\ p(n) = n+1$), on déduit l'existence d'un unique

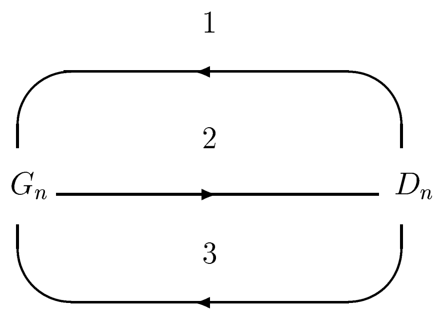
[1] Le graphe des mots de de Bruijn correspond au graphe des mots d'une suite de complexité maximale ($\forall n,\ p(n) = d^n$,) et a été introduit par de Bruijn dans le but de construire des suites finies circulaires de longueur d^n à valeurs dans $\{0, 1, \dots, d-1\}$, telles que tout facteur de longueur n apparaît une fois et une seule : une telle suite correspond à un chemin Hamiltonien fermé dans le graphe de de Bruijn.

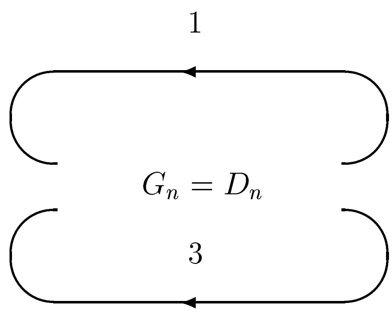
facteur D_n de longueur n *biprolongeable* à droite, c'est-à-dire ayant deux prolongements à droite dans la suite. Soit, de même, G_n l'unique facteur de longueur n biprolongeable à gauche. Une suite sturmienne présente deux types de graphes selon que $G_n = D_n$ ou que $G_n \neq D_n$.

Cette forme simple du graphe des mots permet de déduire des informations sur les fréquences des facteurs des suites sturmiennes. Cette idée a été introduite par Dekking dans le cas de la suite de Fibonacci (voir [25]). Rappelons que la connaissance des fréquences de blocs d'une suite sturmienne u permet une description précise de la mesure associée au système dynamique $(\overline{\mathcal{O}(u)}, T)$. En effet le système dynamique associé à une suite sturmienne est *uniquement ergodique* (ceci est une conséquence de la description des suites sturmiennes donnée par le théorème 5.2, la rotation sous-jacente étant d'angle irrationnel).

Soit U un sommet de Γ_n. On note U^+ le nombre d'arêtes de Γ_n d'origine U et U^- le nombre d'arêtes d'extrémité U. On a le lemme suivant.

Lemme 6.1 *Supposons que $U \to V$ et que $U^+ = 1 = V^- = 1$, alors les facteurs U et V ont même fréquence*





On déduit alors de ce lemme que tous les mots du chemin (1) (voir la figure ci-dessous), privé de D_n et G_n ont même fréquence, que, de même, tous les mots du chemin (3), privé de D_n et G_n ont même fréquence et enfin, que tous les mots du chemin (2), D_n et G_n inclus, ont même fréquence. On en déduit donc que les fréquences des facteurs de même longueur d'une suite sturmienne prennent au plus 3 valeurs. De plus, si $G_{n-1} = D_{n-1}$ alors l'un des deux chemins (1) ou (3) est vide, c'est-à-dire que l'on a une arête de D_n vers G_n. On en déduit donc la proposition suivante.

Proposition 6.2 *Les fréquences des facteurs de même longueur d'une suite sturmienne prennent au plus* 3 *valeurs. Si* $G_{n-1} = D_{n-1}$*, les fréquences des facteurs de longueur* n *prennent au plus* 2 *valeurs.*

Plus généralement, supposons que toutes les fréquences des facteurs d'une suite u existent. Notons que la fonction qui associe à une flèche étiquetée par xWy la fréquence du facteur xWy est un *flot* satisfaisant à la loi de Kirchhoff. Une *branche* du graphe Γ_n est une suite de longueur maximale $(U_1, \ldots, U_m)$ de flèches connectées de Γ_n, éventuellement vide, satisfaisant

$$U_i^+ = 1, \quad \text{pour} \quad i < m, \quad U_i^- = 1, \quad \text{pour} \quad i > 1.$$

Par conséquent, les flèches d'une même branche ont la même fréquence et le nombre de fréquences des facteurs de longueur donnée est majoré par le nombre de branches. On montre facilement le résultat suivant [13].

Théorème 6.3 *On considère une suite récurrente de complexité* $p(n)$*, les fréquences des facteurs de longueur* n *prennent au plus* $3(p(n+1) - p(n))$ *valeurs.*

On en déduit que si $p(n+1) - p(n)$ est uniformément majoré en n, les fréquences des facteurs de longueur donnée prennent un nombre fini de valeurs. En fait, en utilisant le résultat suivant de Cassaigne dont la preuve repose aussi sur une utilisation du graphe des mots (voir [19]), on obtient le résultat suivant.

Théorème 6.4 (Cassaigne) *Si la complexité* $p(n)$ *d'une suite est au plus linéaire, alors* $p(n+1) - p(n)$ *est borné.*

Corollaire 6.5 *Si la complexité* $p(n)$ *d'une suite est au plus linéaire, alors les fréquences des facteurs de longueur donnée ne prennent qu'un nombre fini de valeurs.*

7 Suites automatiques

Les suites *automatiques* (c'est-à-dire les suites engendrées par automates finis) forment une classe de suites largement étudiées, notamment puisque munies d'une forte structure algorithmique. En effet, ce sont des projections littérales de points fixes de *substitutions de longueur constante*. Les suites point fixe de substitutions de longueur constante ont de nombreuses propriétés. En particulier, elles peuvent être engendrées par un processus algorithmique simple : un automate fini [23]. Rappelons les définitions d'automate et de suite automatique. Pour plus d'informations sur le sujet, voir les survols [2, 5, 26]. Un *q-automate* est défini par la donnée :

- d'un ensemble fini d'états $S = \{i = a_1, a_2, ..., a_d\}$, dont on a privilégié un élément i, appelé état initial,
- de q applications ou "flèches" de l'ensemble des états S dans lui-même, notées $0, 1, ..., q-1$,
- d'une application φ de S dans un ensemble Y, dite fonction de sortie.

Une suite $(u(n))_{n\in\mathbb{N}}$ à valeurs dans Y est dite *q-automatique* si elle est engendrée par un q-automate de la façon suivante : considérons un entier n, écrivons-le en base q et identifions les chiffres de son développement aux applications $0, 1, ..., q-1$; en lisant les chiffres de n de droite à gauche, on obtient une application composée de S dans lui-même. Soit a_f l'image de l'état initial i par cette application, on pose alors : $u(n) = \varphi(a_f)$.

La notion de q-automaticité que nous considérons ici est basée sur la numération en base q. Notons que l'on peut étendre la notion d'automaticité à des systèmes de numération plus généraux, comme la numération de Fibonacci. Dans tout ce qui suit, nous entendrons par automaticité la notion de q-automaticité.

Le théorème suivant dû à Christol, Kamae, Mendès France et Rauzy [20, 21] relie l'automaticité aux propriétés suivantes combinatoires et algébriques.

Théorème 7.1 (Christol, Kamae, Mendès France et Rauzy)
Soit $u = (u(n))_{n\in\mathbb{N}}$ une suite à valeurs dans le corps $\mathbb{F}_q$. Les conditions suivantes sont équivalentes :

1. *la série formelle $\sum_{n\geq 0} u(n)X^n$ est algébrique sur le corps $\mathbb{F}_q(X)$,*

2. *le q-noyau $N_q(u)$ de la suite u est fini, où $N_q(u)$ est défini comme l'ensemble des sous-suites de la forme*

$$N_q(u) = \{(u(q^k n + r))_{n \in \mathbb{N}};\ k \geq 0;\ 0 \leq r \leq q^k - 1\},$$

3. *la suite u est q-automatique,*

4. *la suite u est l'image par une projection lettre à lettre d'un point fixe d'une substitution de longueur q.*

La dernière équivalence est due à Cobham [23] et l'équivalence entre les conditions 2 et 3 est due à Eilenberg [28].

Durand généralise dans [27] l'équivalence entre la condition 3 et l'automaticité, aux suites substitutives uniformément récurrentes. Cette caractérisation est basée sur la notion de mots de retour et de suite dérivée. Un mot de retour sur le facteur w est un mot séparant deux occurrences successives de w; on appelle *suite dérivée* d'une suite minimale u fixée, une suite obtenue en codant u par les mots de retour sur un préfixe non vide de la suite u. La caractérisation est alors la suivante.

Théorème 7.2 (Durand) *Une suite uniformément récurrente est substitutive si et seulement si l'ensemble de ses suites dérivées est fini.*

Ce résultat a été prouvé indépendamment par Holton et Zamboni [36].

L'équivalence entre les conditions 1 et 2 a de nombreuses applications en transcendance (pour des séries formelles à coefficients dans un corps fini), voir par exemple les survols [4, 46].

8 Complexité au plus linéaire

En conclusion, nsous allons d'abord rappeler quelques propriétés générales de la fonction de complexité et des fréquences de facteurs des suites sturmiennes et substitutives. Puis nous nous intéresserons aux suites de complexité au plus linéaire.

Théorème 8.1 1. *La complexité d'un point fixe d'une substitution primitive (voir par exemple [43]) ou d'une suite automatique [23] satisfait*

$$\forall n,\ p(n) \leq Cn.$$

2. *Plus généralement la complexité d'un point fixe de substitution satisfait [29, 42]*
$$\forall n,\ p(n) \leq Cn^2.$$

Cette propriété permet de montrer qu'une suite de "haute complexité" n'est pas automatique. Notons que des suites automatiques ont des fonctions de complexité similaires mais des propriétés très dissemblables pour ce qui concerne leurs propriétés spectrales (voir par exemple [43]).

Considérons quelques résultats sur les fréquences des facteurs d'un point fixe d'une substitution de longueur constante. Pour des résultats généraux sur les suites substitutives primitives, voir [43]. Rappelons que pour de telles suites les fréquences existent et sont strictement positives, d'après le théorème de Perron-Frobénius. Le résultat suivant d'équirépartition s'applique pour les fréquences de facteurs de substitutions primitives de longueur constante.

Théorème 8.2 (Dekking) *Soit σ une substitution primitive de longueur constante. On note $f(w)$ la fréquence du facteur w. Il existe $C_2 > C_1 > 0$ tels que*
$$C_1 \leq nf(w) \leq C_2,$$
pour tout $n \geq 1$ et pour tous les facteurs w de longueur n.

Le résultat suivant se prouve en appliquant les propriétés des matrices stochastiques à la puissance n-ième d'une matrice de substitution divisée par l^n, où l est la longueur de la substitution (voir [23]).

Théorème 8.3 (Cobham) *Si les fréquences des facteurs d'une suite automatique existent, elles sont rationnelles. De plus, si la substitution correspondante est primitive, alors les fréquences existent.*

En particulier, une suite sturmienne ne peut être automatique, puisque les fréquences des lettres sont irrationnelles. Pour des résultats plus précis sur les suites automatiques dont une lettre a une fréquence nulle, voir [23].

Considérons plus généralement le cas des suites de complexité au plus linéaire : $\exists C > 0,\ \forall n \in \mathbb{N},\ p(n) \leq Cn$. De nombreuses propriétés tant combinatoires, ergodiques, qu'arithmétiques peuvent se déduire de cette simple indication sur l'ordre de croissance de la fonction de complexité.

Rappelons le théorème suivant :

Théorème 8.4 (Cassaigne) *Si la complexité $p(n)$ d'une suite est au plus linéaire, alors $p(n+1)-p(n)$ est borné.*

Notons que ce résultat est faux pour la complexité sous-quadratique : Ferenczi donne dans [31] un exemple de suite de complexité au plus quadratique dont les différences secondes $p(n+2)+p(n)-2p(n+1)$ ne sont pas bornées. De plus, Ferenczi déduit du théorème 6.4 le résultat suivant [31] : les systèmes minimaux de complexité au plus linéaire sont engendrés par un nombre fini de substitutions (un tel système est appelé *S-adique* selon la terminologie due à Vershik). De même, une suite engendrée par l'itération d'un nombre fini de substitutions est appelée *suite S-adique*.

Théorème 8.5 (Ferenczi) *Soit u une suite uniformément récurrente de complexité au plus linéaire définie sur l'alphabet $\mathcal{A}$; il existe alors un nombre fini c de substitutions σ_i, $1 \leq i \leq c$, définies sur un alphabet $\mathcal{B}$, une application φ de $\mathcal{B}$ dans $\mathcal{A}$ et une suite infinie $(i_n)_{n\in\mathbb{N}}$ à valeurs dans $\{1,\ldots,c\}$ tels que d'une part*

$$\lim_{n\to+\infty} \inf_{b\in\mathcal{B}} |\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_n}(b)| = +\infty$$

et d'autre part il existe une lettre b de $\mathcal{B}$ pour laquelle tout facteur de la suite u est facteur de $\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_n}(b)$, pour un certain entier n.

Le nombre c de ces substitutions peut être de plus majoré explicitement dans le cas où $\forall n,\ p(n+1)-p(n) \leq 2$.

On connaît ainsi explicitement un développement S-adique pour les systèmes dynamiques engendrés par les suites sturmiennes [9].

La réciproque du théorème 8.5 est fausse. Il suffit en effet de considérer une substitution de complexité quadratique pour produire un contre-exemple, comme le point fixe commençant par a de la substitution $a \mapsto ab$, $b \mapsto bc$, $c \mapsto c$ [42]. Il reste à définir une notion plus forte de S-adicité qui permettrait de caractériser les suites de complexité au plus linéaire : considérons donc une suite u engendrée par l'itération d'un nombre fini de substitutions; quelles restrictions faut-il imposer aux substitutions pour que la suite u soit de complexité au plus linéaire ?

La complexité au plus linéaire implique de plus les propriétés ergodiques suivantes dans le cas des suites minimales. Rappelons que cette propriété est invariante par isomorphisme topologique.

Soit u une suite primitive de complexité au plus linéaire.

- Une première conséquence est que le décalage unilatère T sur $\overline{\mathcal{O}}(u)$ peut être rendu bijectif sauf sur un nombre fini d'orbites (voir par exemple le cours [33]).

- Boshernitzan a montré que l'on obtient une majoration explicite (en fonction de l'ordre de croissance de la fonction de complexité) du nombre n de mesures ergodiques [12, 14].

- En particulier les conditions suivantes impliquent l'unique ergodicité [12, 14] :
$$\liminf_{n\to+\infty} \frac{p(n)}{n} < 2,$$
ou
$$\limsup_{n\to+\infty} \frac{p(n)}{n} < 3.$$

- Enfin Ferenczi a montré dans [31] l'absence de mélange fort.

References

[1] P. Alessandri, *Codages de Rotations et Basses Complexités*, Université Aix-Marseille II, Thèse (1996).

[2] J.-P. Allouche, *Automates Finis en Théorie des Nombres*, Expo. Math., **5** (1987), 239–266.

[3] J.-P. Allouche, *Sur la Complexité des Suites infinies*, Bull. Belg. Math. Soc. **1** (1994), 133–143.

[4] J.-P. Allouche *Finite automata and arithmetic,* Séminaire Lotharingien de Combinatoire (Gerolfingen, 1993), Prépubl. Inst. Rech. Math. Av., Univ. Louis Pasteur, Strasbourg, 1993 **34** (1993), 1–18.

[5] J.-P. Allouche, M. Mendès France, *Automata and Automatic Sequences*, Actes de l'École de Physique Théorique des Houches: "Beyond quasicrystals", Les Éditions de Physique, Springer (1995), 293–367.

[6] P. Arnoux *Le codage du flot géodésique sur la surface modulaire*, Enseig. Math. **40** (1994), 29–48.

[7] P. Arnoux *Recoding Sturmian sequences on a subshift of finite type: chaos from order, a worked out example*, Complex Systems, E. Goles and S. Martinez (éd.), Kluwer Academic Publ. (2001), 1–67.

[8] P. Arnoux, A. Fisher *The scenery flow for geometric structures on the torus: the linear setting*, Chin. Ann. of Math. **2B** (2001), 1–44.

[9] P. Arnoux, G. Rauzy, *Représentation Géométrique de Suites de Complexité* $2n+1$, Bull. Soc. Math. France **199** (1991), 199–215.

[10] J. Berstel, *Recent Results in Sturmian Words*, Developments in Language Theory II (DLT'95) (Dassow, Rozenberg, Salomaa eds) World Scientific (1996), 13–24.

[11] V. Berthé, *Fréquences des Facteurs des Suites Sturmiennes*, Theoret. Comput. Sci. **165** (1996), 295–309.

[12] M. Boshernitzan, *A Unique Ergodicity of Minimal Symbolic Flows with Linear Block Growth*, J. Anal. Math. **44** (1984), 77–96.

[13] M. Boshernitzan, *A Condition for Minimal Interval Exchange Maps to be Uniquely Ergodic*, Duke Math. J. **52** (1985), 723–752.

[14] M. Boshernitzan, *A condition for unique ergodicity of minimal symbolic flows*, Ergodic Theory Dynam. Sys. **12** (1992), 425–428.

[15] S. Brlek, *Enumeration of Factors in the Thue-Morse Word*, Discr. Appl. Math. **24** (1989), 83–96.

[16] T. C. Brown, *Descriptions of the Characteristic Sequence of an Irrational*, Canad. Math. Bull **36** (1993), 15–21.

[17] N. G. de Bruijn *A combinatorial problem*, Rominklijke Netherland Academic Van Wetenschepen Proc. **49** Part. 20 (1946), 758–764.

[18] J. Cassaigne, *Special Factors of Sequences with Linear Subword Complexity*, Developments in Language Theory II (DLT'95) (Dassow, Rozenberg, Salomaa eds) World Scientific (1996), 25–34.

[19] J. Cassaigne, *Complexité et Facteurs Spéciaux*, Bull. Belg. Math. Soc. **4** (1997), 67–88.

[20] G. Christol, *Ensembles presque Périodiques k-Reconnaissables*, Theoret. Comp. Sci. **9** (1979), 141–145.

[21] G. Christol, T. Kamae, M. Mendès France, G. Rauzy, *Suites Algébriques, Automates et Substitutions*, Bull. Soc. Math. France **108** (1980), 401-419.

[22] A. Cobham, *On the Base Dependence of Sets of Numbers Recognisable by Finite Automata*, Math. Systems Theory **3** (1969), 186–192.

[23] A. Cobham, *Uniform Tag Sequences*, Math. Systems Theory **6** (1972), 164–192.

[24] E. M. Coven, G. A. Hedlund, *Sequences with Minimal Block Growth*, Math. Systems Theory **7** (1973), 138–153.

[25] F. M. Dekking, *On the Prouhet-Thue-Morse Measure*, Acta Universitatis Carolinae, Mathematica et Physica **33**, pp. 35–40 (1992).

[26] M. Dekking, M. Mendès France, A. van der Poorten *Folds!*, Math. Intell. **4** (1982), 130–138; *Folds!II: symmetry disturbed* Math. Intell. **4** (1982), 173–181; *Folds!III: more morphisms* Math. Intell. **4** (1982), 190–196.

[27] F. Durand, *A Characterization of Substitutive Sequences using Return Words*, Discrete Math. **179** (1998), 89–101.

[28] S. Eilenberg, *Automata, Languages, and Machines*, Vol. A Academic Press (1974).

[29] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, *Subwords of various classes of deterministic developmental languages without intercations*, Theoret. Comput. Sci. **1** (1975), 59–75.

[30] S. Ferenczi, *Les Transformations de Chacon : Combinatoire, Structure Géométrique, Lien avec les Systèmes de Complexité* $2n+1$, Bull. Soc. math. France **123** (1995), 271–292.

[31] S. Ferenczi, *Rank and Symbolic Complexity*, Erg. Theory Dynam. Sys. **16** (1996), 663–682.

[32] S. Ferenczi *Measure-theoretic complexity of ergodic systems*, Israel J. Math. **100** (1997), 189–207.

[33] S. Ferenczi *Substitutions and symbolic dynamical systems*, Substitutions in Dynamics, Arithmetics and Combinatorics, (V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel éd.) Lecture Notes Math., Springer Verlag, à paraître.

[34] S. Ferenczi, *Complexity of Sequences and Dynamical Systems*, Discrete Math. **206** (1999), 663–682.

[35] S. Ferenczi, Z. Kása *Complexity for finite factors and infinite sequences*, Theoret. Comput. Sci. (actes de la confŕence *Words*, Rouen, 1997) **218** (1999), 177–195.

[36] C. Holton, L. Q. Zamboni *Descendants of primitive substitutions*, Theory Comput. Systems **32** (1999), 133–157.

[37] A. Iványi *On the d-complexity of words*, Ann. Univ. Sci. Budapest Sect. Comput. **8** (1987), 69–90.

[38] M. Lothaire *Algebraic Combinatorics on Words*, Chapitre 2: Sturmian words, par J. Berstel et P. Séébold, à paraître.

[39] F. Mignosi *On the number of factors of Sturmian words*, Theoret. Comput. Sci. **82** (1991), 71–84.

[40] M. Morse, G. A. Hedlund, *Symbolic Dynamics*, Amer. J. Math. **60**, pp. 815–866 (1938).

[41] M. Morse, G. A. Hedlund, *Symbolic Dynamics II: Sturmian Trajectories*, Amer. J. Math. **62**, pp. 1–42 (1940).

[42] J.-J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Lecture Notes in Comput. Sci. **172** (1984), 380–389.

[43] M. Queffélec, *Substitution Dynamical Systems. Spectral Analysis*, Lecture Notes in Math. **1294**, Springer-Verlag (1987).

[44] G. Rauzy, *Suites à Termes dans un Alphabet Fini*, Sém. de Théorie des Nombres de Bordeaux, 25-01–25-16 (1983).

[45] G. Rauzy, *Low Complexity and Geometry*, Dynamics of Complex Interacting Systems (Santiago, 1994), 147–177, Nonlinear Phenom. Complex Systems, 2, Kluwer Acad. Publ., Dordrecht (1996).

[46] D. S. Thakur, *Automata and Transcendence*, Contemporary Mathematics, **210** (1998), 387–399.

[47] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics **79**.