

# Complexité de suites à valeurs dans un ensemble fini : quelques exemples

PIERRE ARNOUX

## 1. Qu'est-ce que la complexité d'une suite ?

Nous considérons ici des suites à valeurs dans un ensemble fini  $\mathcal{A}$ , appelé alphabet. On en rencontre dans des situations diverses : bien entendu, dans tout ce qui relève du codage de l'information (texte, son...), mais aussi en biologie (génomique), et en physique (étude de l'évolution d'un système au cours du temps, quand on ne peut faire que des mesures discrètes). Une méthode assez générale pour obtenir de telles suites est de prendre un système dynamique  $T : X \rightarrow X$ , et une fonction de codage  $f : X \rightarrow \mathcal{A}$  (ou, ce qui revient au même, une partition finie de  $X$  indexée par  $\mathcal{A}$ ) ; on considère alors les suites de terme général  $f(T^n(x))$ , codage symbolique de l'orbite du point  $x$  sous l'action de  $T$ . On peut considérer suivant les cas des suites infinies (indexées par  $\mathbf{N}$ ) ou bi-infinies (indexées par  $\mathbf{Z}$ ) ; nous nous restreindrons ici au cas des suites infinies.

On s'intéresse particulièrement à évaluer le caractère plus ou moins aléatoire d'une telle suite ; si l'on connaît  $n$  symboles de la suite, que peut-on dire des suivants ? Une méthode possible consiste à compter le nombre de combinaisons possibles de  $n$  symboles consécutifs. C'est ce que l'on appelle la *complexité* de la suite ; donnons d'abord quelques définitions :

On appelle *mot* sur l'alphabet  $\mathcal{A}$  une suite finie  $W = w_1 \dots w_n$  à valeurs dans  $\mathcal{A}$  ; on notera  $|W| = n$  la longueur du mot, et  $|W|_{\mathbf{a}}$  le nombre d'occurrences de la lettre  $\mathbf{a}$  dans  $W$ . On dit qu'un mot  $W$  est *facteur* de la suite  $u = (u_n)_{n \in \mathbf{N}}$  s'il existe un entier  $k$  tel que  $W = u_k \dots u_{k+n-1}$ .

**DÉFINITION.** — *On appelle complexité de la suite  $u$  la fonction  $p$  qui, à tout entier  $n$ , associe le nombre de facteurs de longueur  $n$  de  $u$ .*

Il est clair que, si l'alphabet  $\mathcal{A}$  est de cardinal  $k$ ,  $p(n)$  est borné par  $k^n$  ; il est facile de montrer que, puisqu'un mot de longueur  $n + m$  peut se décomposer en deux mots de longueur respective  $n$  et  $m$ , on a  $p(n + m) \leq p(n)p(m)$ . On en déduit que  $\log(p(n))/n$  admet une limite finie quand  $n$  tend vers l'infini ; c'est ce qu'on appelle *l'entropie* de la suite. Une suite d'entropie  $\log k$  contient tous les mots possibles ; une suite d'entropie strictement positive peut être considérée comme très aléatoire.

Dans cet exposé, nous allons montrer comment, dans certains cas, on peut calculer explicitement la complexité d'une suite en considérant un système dynamique. Nous nous intéresserons à des suites peu aléatoires, d'entropie nulle.

Si la complexité d'une suite est majorée par une exponentielle, peut-on donner des minoration? on a un lemme évident :

LEMME. — *la complexité est une fonction croissante.*

En effet, tout facteur de longueur  $n$  de  $u$  est le début d'au moins un facteur de longueur  $n + 1$ , et deux facteurs distincts sont les débuts de deux facteurs distincts, il y a donc au moins autant de facteurs de longueur  $n + 1$  que de facteurs de longueur  $n$ .

On ne peut pas faire mieux en général : les suites les moins aléatoires sont les suites périodiques, et pour une telle suite, le nombre de mots de longueur  $n$  est borné par la période de la suite ; la complexité est alors constante à partir d'un certain rang. Plus généralement, rappelons qu'une suite est dite *ultimement périodique* si elle est périodique à partir d'un certain rang. On a en fait une caractérisation complète des suites de complexité bornée :

LEMME. — *Les trois propriétés suivantes sont équivalentes :*

- (i) *La suite  $u$  est ultimement périodique.*
- (ii) *La complexité est bornée*
- (iii) *il existe un entier  $n$  tel que  $p(n) = p(n + 1)$ .*

*Démonstration.* — Si la suite est périodique de période  $t$  à partir du rang  $n_0$ , on a toujours  $p(n) \leq n_0 + t$ , donc *i* implique *ii* ; il est clair que *ii* implique *iii*, il suffit donc de montrer que *iii* implique *i*.

Mais s'il y a autant de facteurs de longueur  $n$  et  $n + 1$ , chaque facteur de longueur  $n$  est le début d'un et d'un seul facteur de longueur  $n + 1$  ; autrement dit, quand on connaît  $n$  lettres, on connaît la suivante. De plus, comme il n'y a qu'un nombre fini de facteurs, on retrouve en au plus  $p(n)$  fois un facteur déjà vu, et la suite est alors périodique à partir de l'apparition de ce facteur.  $\square$

On en déduit qu'*a contrario*, la complexité d'une suite non ultimement périodique est strictement croissante ; comme  $p(1)$  vaut au moins 2, sinon il n'y aurait qu'un symbole et la suite serait constante, la complexité d'une suite non ultimement périodique vérifie  $p(n) \geq n + 1$ .

DÉFINITION. — *On dit qu'une suite est sturmiennne si elle est non ultimement périodique et de complexité minimale  $p(n) = n + 1$ .*

Remarquons que, puisque  $p(1) = 2$ , une suite sturmiennne est définie sur un alphabet à 2 lettres, que l'on peut nommer  $\mathbf{0}$  et  $\mathbf{1}$ . Les suites sturmiennnes sont en quelque sorte les plus ordonnées des suites non périodiques ; elles ont été étudiées par de nombreux auteurs, nous citerons [HM], [He], [CH].

## 2. L'exemple le plus simple : les suites sturmiennes

Les suites sturmiennes possèdent de nombreuses caractérisations, combinatoires, dynamiques, arithmétiques, voire géométriques, que l'on peut résumer dans le théorème suivant :

**THÉORÈME.** — *Pour une suite non ultimement périodique sur l'alphabet à deux lettres  $\{0, 1\}$ , les propriétés suivantes sont équivalentes :*

(i) (complexité) *La suite est de complexité minimale :  $p(n) = n + 1$ .*

(ii) (équilibre) *Les nombres de 0 contenus dans deux facteurs de même longueur diffèrent au plus de 1 : si  $U$  et  $V$  sont deux facteurs de  $u$  tels que  $|U| = |V|$ , alors  $|U|_0 - |V|_0 \leq 1$ .*

(iii) (engendrement par substitutions) *Il existe deux suites  $a_n$  et  $b_n$  d'entiers strictement positifs, vérifiant  $a_n \geq b_n$ , et une lettre  $\mathbf{a} = 0$  ou  $1$  telles que, pour tout  $n$ , la suite  $u$  débute par le mot  $S^{b_1} \sigma_1^{a_1-1} S^{b_2} \sigma_0^{a_2} S^{b_3} \sigma_1^{a_3} \dots S^{b_{2n}} \sigma_0^{a_{2n}}(\mathbf{a})$ , où  $S$  est l'application qui, à un mot, associe ce mot privé de sa première lettre et  $\sigma_0$  (resp.  $\sigma_1$ ) est l'application (substitution, ou encore morphisme du monoïde libre) qui à un mot associe le mot obtenu en remplaçant  $0$  par  $0$  et  $1$  par  $10$  (resp.  $0$  par  $01$  et  $1$  par  $1$ )*

(iv) (arithmétique) *Il existe deux nombres  $\alpha, \beta$  dans l'intervalle  $[0, 1[$ , avec  $\alpha$  irrationnel, tels que la suite  $u_n$  soit définie par  $u_n = [(n+1)\alpha + \beta] - [n\alpha + \beta]$  ou  $u_n = \lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil$ , où  $[x]$  désigne le plus grand entier inférieur ou égal à  $x$  (partie entière de  $x$ ), et  $\lceil x \rceil$  le plus petit entier supérieur ou égal à  $x$ .*

(v) (rotation) *Il existe deux réels  $\alpha$  et  $\beta$  dans  $[0, 1[$ , avec  $\alpha$  irrationnel, tels que la suite  $u$  soit donnée par le codage de l'orbite de  $\beta$  pour le système dynamique*

$$R_\alpha : [0, 1[ \rightarrow [0, 1[ \quad x \mapsto x + \alpha \bmod 1$$

*par rapport à la partition  $[0, 1 - \alpha[, [1 - \alpha, 1[$  (ou par le codage obtenu en prenant les intervalles ouverts à gauche et fermés à droite)*

*Remarque.* — Les propriétés *iv* et *v* sont susceptibles d'une interprétation simple : si l'on trace dans le plan, quadrillé par les horizontales et les verticales à coordonnées entières, la droite  $y = \alpha x + \beta$ , et que l'on compte le nombre d'horizontales que cette droite coupe entre 2 verticales successives (ce nombre vaut 0 ou 1 si  $\alpha < 1$ ), on obtient la suite  $u$  (cf. Figure 1); sous cette forme, le sens de la propriété d'équilibre est assez clair.

Il peut être plus naturel de regarder le codage obtenu en notant  $\mathbf{H}$  chaque fois que la droite considérée coupe une horizontale, et  $\mathbf{V}$  à chaque fois qu'elle coupe une verticale. Ce codage s'obtient à partir du précédent en remplaçant  $0$  par  $\mathbf{V}$  et  $1$  par  $\mathbf{HV}$ ; on peut montrer que la suite obtenue est encore sturmiennne, comme image d'une suite sturmiennne par une "bonne" substitution.

On peut aussi montrer que cette suite est directement associée à la rotation d'angle  $\alpha/(1 + \alpha)$ ; le raisonnement qui suit peut paraître peu naturel, mais il fait

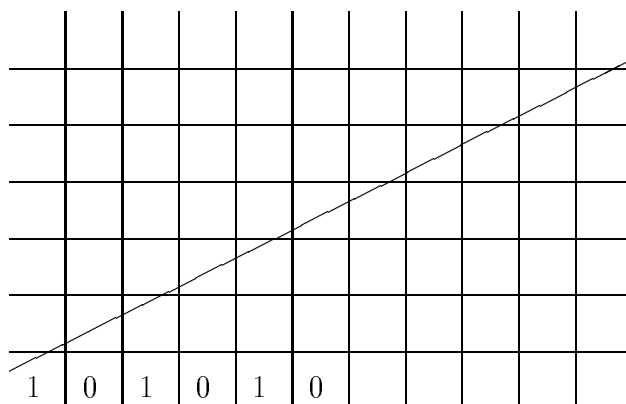


Figure 1

jouer le même rôle aux deux coordonnées et il est donc plus facile à généraliser, comme nous le verrons plus loin.

Notons  $L_n$  la ligne brisée (“escalier”) qui joint les points de la forme  $(k, n - k)$  et les points de la forme  $(k, n + 1 - k)$ , pour  $k \in \mathbf{Z}$ , par des segments horizontaux et verticaux; une droite de pente  $\alpha$  coupe une fois et une seule chaque ligne brisée  $L_n$ , et le codage est déterminé par le fait qu’elle coupe  $L_n$  en un segment horizontal ou vertical. Notons  $D$  la droite  $x + y = 0$ , et cherchons le codage associé à la droite issue d’un point  $p$  de  $D$ ; pour trouver le  $n$ -ième terme de la suite, on projette  $L_n$  sur  $D$  parallèlement à la direction  $(1, \alpha)$ , ce qui donne un pavage périodique de  $D$  en deux types d’intervalles, de longueurs respectives  $1/(1 + \alpha)$  et  $\alpha/(1 + \alpha)$  (si l’on prend le vecteur  $(-1, 1)$  comme base); le type d’intervalle auquel appartient  $p$  donne le terme cherché de la suite. Si l’on décale le point  $p$  de  $(-1, 1)$ , on retrouve le même codage; on peut donc quotienter  $D$  par un groupe de translation pour se ramener à un cercle partitionné en deux intervalles, et le dessin montre que la partition obtenue à partir de  $L_{n+1}$  se déduit de la partition obtenue de  $L_n$  par une rotation de  $\alpha/(1 + \alpha)$ , d’où le résultat cherché (voir la figure 2).

Une autre façon d’obtenir cette suite est de jouer au billard sur le carré, en partant dans la direction  $(1, \alpha)$  et en notant **H** (resp. **V**) chaque fois que l’on touche un côté horizontal (resp. vertical).  $\square$

*Preuve du théorème.* — La preuve complète est assez longue; nous ne pouvons ici donner que des indications. On trouvera une preuve complète, sous forme de problème, dans l’épreuve optionnelle d’informatique de l’agrégation de mathématiques 1994.

Les propriétés  $i$  et  $ii$  sont combinatoires, et la preuve de leur équivalence est aussi purement combinatoire (et non triviale!).

Les propriétés  $iv$  et  $v$  sont évidemment équivalentes:  $iv$  n’est rien d’autre que l’écriture explicite du codage donné par  $v$ . On peut remarquer en particulier qu’utiliser la fonction  $[x]$  (resp.  $\lceil x \rceil$ ) pour  $iv$  revient à prendre les intervalles fermés à gauche

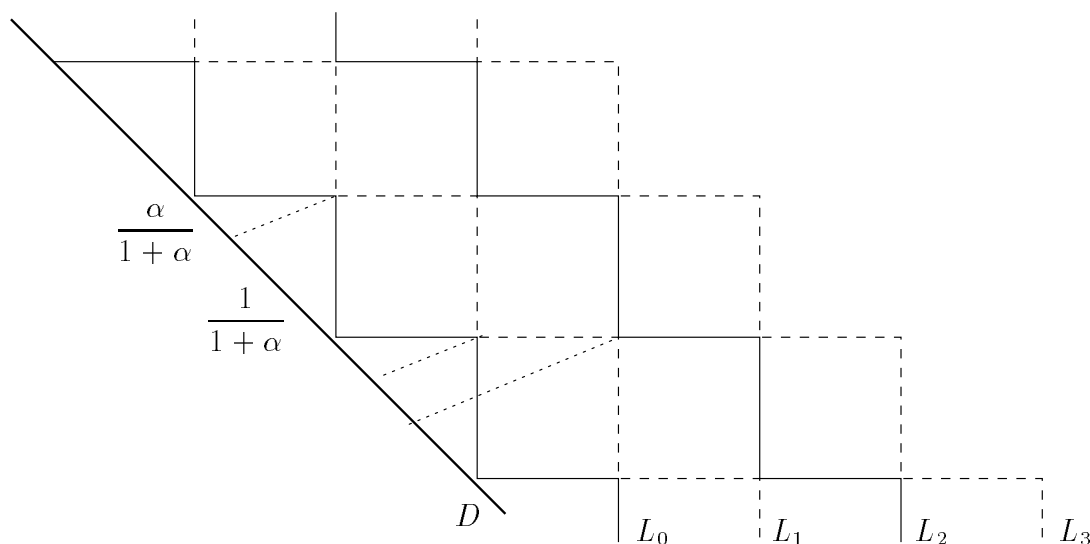


Figure 2

(resp. à droite) dans  $v$  ; en fait, les suites obtenues par l'un ou l'autre codage sont les mêmes sauf pour un nombre dénombrables de valeurs de  $\beta$  (de la forme  $p - n\alpha$ ) où elles diffèrent en deux positions.

Il est facile de montrer que toute suite obtenue par codage d'une rotation est de complexité minimale, et nous le prouverons à la fin de cette section. C'est la réciproque qui est difficile, et c'est ici qu'intervient la propriété *iii*, qui est la plus compliquée à interpréter.

Ce qui est simple à montrer, c'est que toute suite sturmienne peut se réécrire par  $\sigma_0$  ou  $\sigma_1$  ; en effet, puisque  $p(2) = 3$ , il n'y a que trois facteurs de longueur 2 ; or, puisque la suite n'est pas périodique, elle n'est pas constante, donc les deux lettres **0** et **1** apparaissent une infinité de fois, donc les deux mots **01** et **10** aussi ; donc l'un des deux facteurs **00** et **11** n'apparaît pas. Supposons que ce soit **11** ; alors, tout **1** est suivi par **0**, et l'on peut recoder la suite en utilisant **0** et **10** ; autrement dit, il existe une unique suite  $v$  telle que  $u = \sigma_0(v)$ . Il suffirait alors que  $v$  soit sturmienne pour pouvoir itérer, et obtenir  $u$  comme issue d'une suite infinie de recodage.

Il y a malheureusement un problème technique : en général, ce n'est pas la suite  $v$  qui est sturmienne, mais la suite  $Sv$ , c'est-à-dire  $v$  privée de son premier terme ; en effet, si  $u$  commence par un **0**, on ne peut savoir directement si celui-ci doit être considéré, pour le recodage, comme le mot **0** ou comme la deuxième lettre du mot **10**. On peut cependant montrer que l'on peut toujours écrire  $u$  sous la forme  $\sigma_0(v)$  ou  $S\sigma_0(v)$ , où  $v$  est une suite sturmienne ; c'est ici qu'intervient l'application  $S$  de décalage.

Puisque toute suite de codage pour une rotation est sturmienne, elle peut se recoder de cette façon, et on peut alors interpréter la suite  $(a_n)$  : ce n'est autre que le développement en fraction continue du nombre  $\alpha$  ; la suite  $b_n$  est associée, de

façon plus complexe, au nombre  $\beta$ . Il est alors assez simple de prouver que toute suite sturmienne est une suite de rotation, puisqu'il suffit alors de prendre les suites  $(a_n)$  et  $(b_n)$  données par la propriété *iii*, de construire les réels  $\alpha$  et  $\beta$ , et de montrer que le codage associé donne bien la suite cherchée. La propriété *iii* donne en fait une interprétation combinatoire du développement en fraction continue du nombre  $\alpha$ ; dans la cas où ce développement est périodique, la suite peut prendre une forme très simple.

En particulier, si l'on pose  $a_n = 1$ ,  $b_n = 0$ ,  $\mathbf{a} = \mathbf{0}$  ou  $\mathbf{1}$ , on obtient les deux codages possibles pour l'orbite, par la rotation d'angle  $(\sqrt{5} - 1)/2$ , du point  $(3 - \sqrt{5})/2$ . On les trouve donc en partant d'une des deux lettres, en remplaçant  $\mathbf{0}$  (resp.  $\mathbf{1}$ ) par  $\mathbf{01}$  (resp.  $\mathbf{101}$ ) et en itérant. On peut montrer que l'orbite de  $(\sqrt{5} - 1)/2$  pour la même rotation est, elle, donnée en itérant la substitution dite de *Fibonacci*,  $\mathbf{1} \mapsto \mathbf{10}$ ,  $\mathbf{0} \mapsto \mathbf{1}$ ; le mot infini obtenu, dit *mot de Fibonacci*, est la suite sturmienne la plus facile à exhiber.  $\square$

Il reste à montrer que la suite donnée par le codage de l'orbite d'une rotation est sturmienne; nous allons en donner deux preuves. Pour montrer que la suite  $[(n + 1)\alpha + \beta] - [n\alpha + \beta]$  est sturmienne, on peut utiliser la propriété d'équilibre: le nombre de  $\mathbf{1}$  compris entre  $u_k$  et  $u_{k+n-1}$  est égal par construction à  $[(k + n)\alpha + \beta] - [k\alpha + \beta]$ , et il est clair que, pour  $n$  fixé, ce nombre ne peut prendre au plus que deux valeurs distinctes suivant  $n$ . On peut aussi étudier directement la complexité, et la preuve qui suit est plus susceptible de généralisation.

Fixons quelques notations. Pour  $\alpha$  irrationnel fixé, on note  $R_\alpha$  la rotation d'angle  $\alpha$  sur le cercle identifié à l'intervalle  $[0, 1[$ ; on note  $\mathcal{P}$  la partition en les deux intervalles  $I_0 = [0, 1 - \alpha[$  et  $I_1 = [1 - \alpha, 1[$ . L'application  $f$  de codage est donnée par  $f(x) = \mathbf{0}$  (resp.  $\mathbf{1}$ ) si  $x \in I_0$  (resp.  $I_1$ ), et on cherche la complexité de la suite  $u(\beta) = (f(R_\alpha^n \beta))_{n \in \mathbf{N}}$ .

Mais un facteur de  $u(\beta)$  qui apparaît en position  $k$  est un facteur initial de  $u(b + k\alpha)$ : plutôt que de chercher tous les facteurs de  $\beta$  de longueur  $n$ , avec  $\beta$  fixé, on peut chercher le facteur initial de toutes les suites  $u(x)$ .

La première lettre de  $u(x)$  est déterminée par la position de  $x$  par rapport à la partition  $\mathcal{P}$ ; la deuxième lettre est déterminée par la position de  $R_\alpha x$  par rapport à  $\mathcal{P}$ , ou encore par la position de  $x$  par rapport à  $R_\alpha^{-1}\mathcal{P}$ . En effet on a  $u_1(x) = \mathbf{0}$  si et seulement si  $R_\alpha x \in I_0$ , ou  $x \in R_\alpha^{-1}I_0$ . Pour connaître le facteur initial de  $u(x)$ , il faut placer  $x$  par rapport aux partition  $\mathcal{P}, \mathcal{R}_\alpha^{-\infty}\mathcal{P}, \dots, \mathcal{R}_\alpha^{-\lambda+\infty}\mathcal{P}$ , et il y a donc autant de facteurs initiaux possibles que d'ensembles dans l'intersection  $\mathcal{P} \wedge \mathcal{R}_\alpha^{-\infty}\mathcal{P} \wedge \dots \wedge \mathcal{R}_\alpha^{-\lambda+\infty}\mathcal{P}$  de ces partitions.

Un calcul immédiat montre que la partition  $R_\alpha^{-k}\mathcal{P}$  est constituée des deux intervalles  $[-k\alpha, 1 - (k + 1)\alpha[$  et  $[1 - (k + 1)\alpha, 1 - k\alpha[$  (attention, tous ces calculs se font modulo 1! voir figure 3); l'intersection des  $n$  premiers itérés inverses de la partition  $\mathcal{P}$  est donc donnée par  $n + 1$  points, et la partition obtenue a donc  $n + 1$  éléments. On a bien prouvé qu'il y a  $n + 1$  facteurs initiaux de longueur  $n$  possibles, donc au plus  $n + 1$  facteurs possibles pour  $u(\beta)$ . Compte tenu du fait que la suite  $u(\beta)$  n'est pas ultimement périodique, elle est de complexité  $n + 1$ . On pourrait en

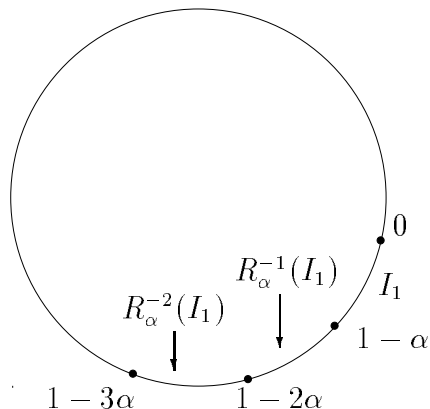


Figure 3

fait, en utilisant le fait que l'orbite d'un point pour une rotation irrationnelle sur le cercle est dense, montrer directement que tous les facteurs initiaux possibles sont des facteurs de  $u(\beta)$ .

### 3. Suites engendrées par des systèmes dynamiques

On peut en fait généraliser la démonstration précédente. Considérons, comme dans l'introduction, un système dynamique  $T : X \rightarrow X$ , une partition  $\mathcal{P}$  indicée par  $\mathcal{A}$ , et la fonction  $f$  de codage associée. Notons  $u(x) = (f(T^n x))_{n \in \mathbf{N}}$  la suite donnée par le codage de l'orbite de  $x$ ; comme ci-dessus, il est clair qu'un facteur de  $u(x)$  qui apparaît en position  $k$  est facteur initial de  $u(T^k(x))$ .

Par ailleurs, la lettre d'ordre  $k$  de  $u(x)$  donne la position de  $x$  par rapport à la partition  $T^{-k}\mathcal{P}$ , et comme ci-dessus, il y a donc autant de facteurs initiaux de longueur  $n$  que d'atomes dans l'intersection  $\mathcal{P} \wedge T^{-1}\mathcal{P} \wedge \dots \wedge T^{-n+1}\mathcal{P}$ ; on a donc :

**PROPOSITION.** — *Soit  $u(x)$  la suite donnée par le codage de l'orbite de  $x$  pour le système  $T : X \rightarrow X$  par rapport à la partition  $\mathcal{P}$ ; la complexité d'ordre  $n$  de  $u$  est majorée par le nombre d'atomes de la partition  $\mathcal{P} \wedge T^{-1}\mathcal{P} \wedge \dots \wedge T^{-n+1}\mathcal{P}$ .*

Il est en général impossible d'aller plus loin, pour deux raisons : d'une part, l'orbite de  $x$  peut éviter de larges régions de  $X$  (par exemple, si l'orbite de  $x$  est périodique), et les mots initiaux correspondants n'apparaîtront pas comme facteurs de  $u(x)$ ; d'autre part, la région associée à un mot initial peut être très petite, voire réduite à un nombre fini de points, et dans ce cas le mot correspondant n'apparaîtra pas dans la plupart des orbites.

Dans la cas d'un système dynamique topologique, on peut donner des conditions qui évitent ces phénomènes, et l'on a :

**PROPOSITION.** — *Soit  $X$  un espace métrique compact, et  $T$  un homéomorphisme de  $X$ . Soit  $\mathcal{P} = \{\mathcal{P}_a, a \in \mathcal{A}\}$  une partition de  $X$  qui vérifie la condition suivante :*

(\*) Pour toute suite finie  $\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{n-1}$ , l'ensemble  $P_{\mathbf{a}_0} \cap T^{-1}P_{\mathbf{a}_1} \cap \dots \cap T^{-n+1}P_{\mathbf{a}_{n-1}}$  est vide ou d'intérieur non vide.

Alors, si  $x$  est un point d'orbite dense pour  $T$ , la complexité de la suite  $u(x)$  est égale au nombre d'atomes des partitions  $\mathcal{P} \wedge T^{-1}\mathcal{P} \wedge \dots \wedge T^{-n+1}\mathcal{P}$

*Démonstration.* — En effet, tout mot initial correspond à un ensemble qui contient un ouvert; par densité de l'orbite de  $x$ , ce mot initial est un facteur de  $u(x)$ .  $\square$

Il y a un cas où l'on vérifie facilement la condition sur  $x$  : on dit qu'un système dynamique topologique est **minimal** s'il n'admet pas de sous-ensemble fermé invariant, ou, de manière équivalente, si toute orbite est dense; dans ce cas, on a le théorème suivant :

**THÉORÈME.** — Soit  $T : X \rightarrow X$  un système dynamique minimal, et soit  $\mathcal{P}$  une partition de  $X$  vérifiant la condition (\*) de la proposition précédente; alors, pour tout point  $x$ , la complexité de la suite  $u(x)$  est égale au nombre d'atomes des partitions  $\mathcal{P} \wedge T^{-1}\mathcal{P} \wedge \dots \wedge T^{-n+1}\mathcal{P}$ ; en particulier, toutes les suites de codages ont la même complexité (et en fait, les mêmes facteurs).

La suite de l'exposé est consacrée à quelques applications de ce lemme. On va, pour quelques types de suites, exhiber un système dynamique associé, montrer qu'il est minimal, et calculer, par des méthodes géométriques, le nombre d'atomes de la partition associée.

#### 4. Un exemple : les différences secondes de la suite $[n^2\alpha]$

On a vu qu'une suite sturmiennne peut être obtenu comme suite des différences de termes consécutifs de la suite à valeurs entières  $[n\alpha + \beta]$ . On peut essayer de généraliser en considérant la suite  $w$  définie par  $w_n = [n^2\alpha]$ , avec  $\alpha$  irrationnel. Pour obtenir une suite à valeurs dans un alphabet fini, il faut cette fois prendre la suite  $u$  des différences secondes, donnée par  $u_n = w_{n+2} - 2w_{n+1} + w_n$ .

Cette suite n'est bien sûr plus associée à une rotation, mais on peut aussi l'obtenir comme suite de codage d'un système dynamique. Plus précisément, nous allons donner une transformation  $T$  du tore  $\mathbf{T}^2$  et une fonction  $f$  sur  $\mathbf{T}^2$ , ne prenant qu'un nombre fini de valeurs, et telles que l'on ait  $u_n = f(T^n(0, 0))$ .

Il est facile de trouver  $f$  et  $T$  en écrivant  $n^2\alpha$  comme une suite récurrente; en effet, si l'on considère l'application affine  $A$  définie par  $A(x, y) = (x + y + \alpha, y + 2\alpha)$ , on vérifie immédiatement que  $A^n(0, 0) = (x_n, y_n) = (n^2\alpha, n\alpha)$ . On peut alors écrire  $u_n$  en fonction de  $x_n$  et  $y_n$ , puisqu'on a par définition :

$$\begin{aligned} u_n &= w_{n+2} - 2w_{n+1} + w_n \\ &= [x_{n+2}] - 2[x_{n+1}] + [x_n] \\ &= [x_n + 2y_n + 4\alpha] - 2[x_n + y_n + \alpha] + [x_n] \end{aligned}$$

On peut donc écrire  $u_n = f(x_n, y_n)$ , avec  $f(x, y) = [x + 2y + 4\alpha] - 2[x + y + \alpha] + [x]$ ; on vérifie immédiatement que  $f$  est  $\mathbf{Z}^2$ -périodique, donc peut être considéré comme une application définie sur le tore  $\mathbf{T}^2 = \mathbf{R}^2/\mathbf{Z}^2$ , et que de même  $A$ , étant associée à une matrice de  $SL(2, \mathbf{Z})$ , passe au quotient en une transformation affine  $T$  de  $\mathbf{T}^2$ .

On sait que  $T$ , qui est ce qu'on appelle un produit croisé au-dessus d'une rotation, est minimal dès que  $\alpha$  est irrationnel; cela découle de travaux de Furstenberg. En fait, on ne montre pas que toute orbite est dense, mais une propriété bien plus forte : toute orbite est équirépartie sur le tore  $\mathbf{T}^2$  (cf [F], ou [CFS], p. 100-104).

On est donc en position d'appliquer le théorème précédent; il reste à montrer que  $\mathcal{P}$  satisfait la propriété (\*), et à calculer le nombre d'atomes.

Mais d'après l'écriture de  $f$ , les intérieurs des atomes de la partition  $\mathcal{P}$  sont les cellules de la décomposition cellulaire de  $\mathbf{T}^2$  engendrée par les trois cercles  $x = 0$ ,  $x + y + \alpha = 0$  et  $x + 2y + 4\alpha = 0$  (cf. figure 4); un calcul simple montre que la partition  $T^{-n}\mathcal{P}$  est engendrée de même par les cercles  $H_k = T^{-k}H_0$  d'équation  $x + ky + k^2\alpha = 0$ , pour  $n \leq k \leq n + 2$ .

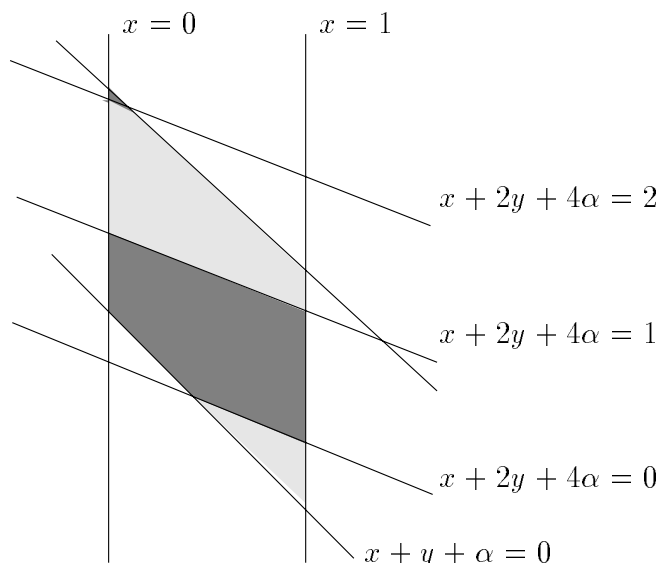


Figure 4 : Les 4 atomes de  $\mathcal{P}$

Ces cellules sont des polygones avec un nombre fini de côté; pour que la condition (\*) ne soit pas satisfaite, il faudrait qu'il y ait une coïncidence de sommets entre certaines de ces cellules, et qu'au moins 3 des cercles  $H_K$  aient une intersection non vide, ce qui est impossible par irrationalité de  $\alpha$ .

On peut alors calculer effectivement la complexité; nous avons une décomposition cellulaire du tore, donnée par un nombre fini de cercles, dont nous voulons connaître le nombre  $F_n$  de faces. Puisque toute arête a deux sommets, et que tout sommet, appartenant à deux cercles, délimite 4 arêtes, les nombres  $S_n$  de sommets et  $A_n$  d'arêtes vérifient  $A_n = 2S_n$ . Compte tenu de la relation d'Euler sur le tore,  $S - A + F = 0$ , on en déduit qu'il y a autant de faces que de sommets. On se ramène donc

à calculer le nombre de points d'intersections de  $H_j$  et  $H_k$ , ou encore, en faisant agir  $T^j$ , de  $H_0$  et  $H_{k-j}$ , qui vaut  $|k-j|$ . En faisant la somme sur tous les couples d'hyperplans, on obtient :

$$p(n) = \sum_{0 \leq j < k \leq n+1} (k-j) = \frac{(n+1)(n+2)(n+3)}{6}$$

D'où le résultat :

**THÉORÈME.** — *la complexité de la suite  $u$  des différences secondes de  $[n^2\alpha]$  est donnée par :*

$$p(n) = \frac{(n+1)(n+2)(n+3)}{6}$$

En particulier, comme pour les suites sturmiennes, la complexité ne dépend pas de la valeur de  $\alpha$ .

*Remarque.* — En toute rigueur, nous avons compté, non pas le nombre d'atomes de l'intersection  $\mathcal{P} \wedge T^{-1}\mathcal{P} \wedge \dots \wedge T^{-n+1}\mathcal{P}$ , mais le nombre de composantes connexes de ces atomes ; il faudrait donc, ce qui peut être fait moyennant un peu de travail supplémentaire, montrer que ces atomes sont connexes ; voir [AMM] pour plus de détails.

Ces arguments se généralisent à des suites obtenues à partir de polynômes de degré quelconque ; rappelons que, si  $w$  est une suite à valeurs réelles, la suite  $v = \Delta w$  des différences est définie par :  $v_n = w_{n+1} - w_n$ . Le résultat énoncé ci-dessus est un cas particulier du théorème suivant :

**THÉORÈME.** — *Soit  $Q$  un polynôme de degré  $d$  à coefficients réels dont le coefficient dominant est irrationnel ; la suite  $(\Delta^d([Q(n)]))_{n \in \mathbf{N}}$  ne prend qu'un nombre fini de valeurs, et sa complexité, qui ne dépend que de  $d$ , est donnée par la formule :*

$$p(n) = \frac{1}{V(0, 1, \dots, d-1)} \sum_{0 \leq k_1 < k_2 < \dots < k_d \leq n+d-1} V(k_d, \dots, k_1)$$

où  $V(k_d, \dots, k_1) = \prod_{1 \leq i < j \leq d} (k_j - k_i)$  est le déterminant de Vandermonde associé à  $(k_d, \dots, k_1)$ .

Il est remarquable que la complexité trouvée ne dépende que du degré du polynôme.

## 5. Un autre exemple : les suites de billard cubique

Au lieu de généraliser la définition arithmétique des suites sturmiennes, on peut regarder la définition géométrique : billard carré, ou droite dans le plan ; nous considérerons donc le billard cubique, c'est-à-dire le système formé par un point mobile dans un cube, sans forces extérieures, avec réflexions élastiques sur les parois.

On code chacune des trajectoires par la suite des faces du cube qu'elle rencontre (on note  $\mathbf{0}$ ,  $\mathbf{1}$  et  $\mathbf{2}$  les trois types de faces, on ne fait pas de différence entre les faces parallèles). A toute trajectoire est ainsi associée un mot infini sur l'alphabet  $\{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$ . On a le résultat suivant :

**THÉORÈME.** — *La complexité d'une trajectoire de pente  $(\alpha, \beta, \gamma)$ , où  $\alpha, \beta$  et  $\gamma$  sont rationnellement indépendants, est égale à  $n^2 + n + 1$ .*

On trouvera une démonstration détaillée dans [AMST]; nous donnons ci-dessous les grandes lignes de la preuve. On supposera, pour simplifier les calculs, que  $\alpha, \beta, \gamma$  sont trois réels rationnellement indépendants dans leur ensemble qui satisfont :  $\alpha + \beta + \gamma = 1$ .

La première étape est de remarquer que, de façon analogue aux suites sturmiennes, il est équivalent d'étudier le billard cubique ou d'étudier une droite de pente irrationnelle dans  $\mathbf{R}^3$  et ses intersections avec les plans  $x = n, y = n, z = n$  pour tout entier  $n$ .

On procède alors de façon analogue à l'étude faite pour la droite de pente irrationnelle dans le plan : Si l'on appelle hauteur du point  $(a, b, c)$  le nombre  $a + b + c$ , on appelle  $\Sigma_n$  la "surface plissée" qui joint les sommets de hauteur  $n, n + 1, n + 2$  (voir figure 5). La droite considérée coupe chaque surface plissée une fois et une seule, dans un de ses trois types de faces. Si l'on projette la surface plissée sur le plan diagonal  $x + y + z = 0$ , on obtient un pavage périodique de ce plan; un calcul simple montre que l'on passe du pavage projeté de  $\Sigma_n$  au pavage projeté de  $\Sigma_{n+1}$  par une translation de vecteur  $(1 - \alpha, -\beta, -\gamma)$ , projection sur le plan diagonal du vecteur  $(1, 0, 0)$  (on pourrait projeter un autre des trois vecteurs de base, car les différences entre ces trois projections appartiennent au groupe de translation du pavage); en quotientant par le groupe du pavage, on se ramène donc à étudier une translation du tore  $\mathbf{T}^2$ , codée par rapport à une partition en trois quadrilatères. Il est bien connu (théorème de Kronecker) qu'une translation irrationnelle du tore est minimale.

Il faut alors, comme précédemment, montrer que la partition considérée satisfait à la condition (\*), ce qui vient de l'hypothèse d'irrationalité, puis montrer que les atomes de l'intersection des partitions itérées sont connexes, et on se ramène à un problème de comptage.

La partition intersectée d'ordre  $n$  vient d'une triangulation du tore avec  $S_n$  sommets,  $A_n$  arêtes et  $F_n$  faces, qui satisfait la formule d'Euler  $S_n - A_n + F_n = 0$ . La partition d'ordre 1, qui est la projection d'un cube en perspective, a 3 sommets, 6 arêtes et 3 faces, qui correspondent aux trois lettres du codage. On passe de la partition d'ordre  $n$  à la partition d'ordre  $n + 1$  en rajoutant 3 segments et un point (projection d'un cube de hauteur plus grande que les précédents); mais ces trois segments recourent les arêtes déjà existantes. On montre que les segments ajoutés à l'étape  $n$  recourent les segments ajoutés à l'étape  $i$  en 2 points si  $i < n - 1$ , et en 0 points si  $i = n - 1$  (plus exactement, dans ce cas ils recourent ces segments en leur extrémité, ce qui ne rajoute pas de nouveau sommet). On a donc  $S_{n+1} = S_n + 2n + 1$ , et donc  $S_n = n^2 + 2$ . On montre de même que  $A_{n+1} = A_n + 4n + 3$ , donc  $A_n = 2n^2 + n + 3$ , d'où l'on déduit le résultat cherché.

Figure 5

## 6. Suites de billard : la conjecture de Tamura

Une fois que l'on a réussi à calculer la complexité du billard carré et du billard cubique, il est bien sûr tentant, comme on l'a fait pour les parties entières de polynômes, de généraliser en dimension quelconque.

**DÉFINITION.** — *Nous noterons  $p(n, s)$  le nombre de facteurs de longueur  $n$  dans une suite engendrée par le billard dans le cube de dimension  $s + 1$*

Le cas  $s = 1$  correspond au billard carré, c'est-à-dire aux suites sturmiennes,  $s = 2$  est le billard cubique de la section précédente,  $s = 0$  est un cas dégénéré, billard sur un segment, correspondant à une suite constante de complexité 1.

Il est à noter que nous ne savons pas si  $P(n, s)$  est bien défini dans le cas général : la définition que nous avons donnée suppose implicitement que la complexité d'une suite de billard cubique irrationnel ne dépend que de la dimension, et pas de la direction initiale, ce qui n'a pas de raison d'être vrai *a priori*.

Cependant, il est facile de vérifier que  $p(n, s)$  est bien défini pour  $n < 3$ , et que l'on a  $p(0, s) = 1$  (cas du mot vide),  $p(1, s) = s + 1$  (nombre de lettres, c'est-à-dire d'hyperfaces du cube de dimension  $s + 1$ ), et  $p(2, s) = s^2 + s + 1$  (car tous les mots  $\mathbf{ij}$ , avec  $\mathbf{i} \neq \mathbf{j}$ , sont possibles, mais un seul des mots  $\mathbf{ii}$  est possible. D'autre part, J.I. Tamura a calculé numériquement la complexité pour un certain nombre d'exemples de dimension supérieure à 3.

Si l'on résume les résultats connus, on trouve le tableau de la page suivante, où les chiffres en italique résultent de simulations numériques

$n \setminus s$	0	1	2	3	4	5	...	$s$
0	1	1	1	1	1	1		1
1	1	2	3	4	5	6		$s + 1$
2	1	3	7	13	21	31		$s^2 + s + 1$
3	1	4	13	<i>34</i>	<i>73</i>	<i>136</i>		
4	1	5	21	<i>73</i>	<i>209</i>			
5	1	6	31	<i>136</i>				
...								
$n$	1	$n + 1$	$n^2 + n + 1$					

Au vu de ces résultats, J.I. Tamura a conjecturé que la fonction  $p(n, s)$  est définie et symétrique :  $p(n, s) = p(s, n)$ .

On peut pousser la conjecture plus loin : il est clair que l'on a  $p(n, s) < (s + 1)^n$ , car c'est le nombre de mots de longueur  $n$  que l'on peut former avec  $s + 1$  lettres ; au vu des résultats obtenus, il est donc naturel de supposer que, pour  $n$  fixé,  $p(n, s)$  est une polynôme unitaire de degré  $n$  en  $s$ .

Mais on peut alors calculer de proche en proche chaque ligne, puisque, par symétrie, les  $n$  premières lignes déterminent les  $n$  premières colonnes, et que l'on connaît donc  $n$  valeurs du polynôme unitaire de degré  $n + 1$  associé à la ligne suivante, ce qui le détermine complètement ; le calcul peut être fait explicitement, et conduit à la conjecture suivante :

CONJECTURE. — *La fonction  $p(n, s)$  est donnée par :*

$$p(n, s) = \sum_{i=0}^{\inf(n,s)} \frac{n!s!}{(n-i)!i!(s-i)!}$$

Cette conjecture est en parfait accord avec les simulations numériques, et a été vérifiée dans un grand nombre de cas particuliers ; elle n'est cependant pas démontrée dès que  $n$  et  $s$  sont plus grands que 2 : les calculs de comptage qui généralisent ceux de la section précédente deviennent alors inextricable. On n'a par ailleurs à l'heure actuelle aucune preuve que cette fonction soit définie dans le cas général ; en fait, même dans les cas  $n = 1$  ou  $n = 2$ , la démonstration la plus simple consiste à calculer explicitement  $p(n, s)$  et à montrer que le résultat ne dépend pas de la direction. Il existe dans ce cas des preuves directes, sans calculer la valeur mais en montrant directement que le nombre de mots ne varie pas quand on change la direction, mais elles sont plus compliquées.

On n'a également aucune idée, même heuristique, de la raison pour laquelle cette fonction serait symétrique en  $n$  et  $s$ .

## 7. Quelques autres résultats

Citons pour terminer quelques autres cas où l'on sait calculer explicitement la complexité :

Une autre généralisation du billard carré est le billard dans un polygone rationnel (c'est-à-dire dont les angles sont des multiples rationnels de  $\pi$ ); dans ce cas, P. Hubert a montré (cf. [Hu]) que, si, dans un polygone à  $q$  côtés, les angles sont de la forme  $k_1\pi/r, k_2\pi/r, \dots, k_q\pi/r$ , où les entiers  $k_1, k_2, \dots, k_q, r$  sont premiers entre eux dans leur ensemble, alors la suite obtenue en codant une trajectoire non périodique par les côtés qu'elle rencontre est de complexité  $p(n) = n(q - 2)r + 2r$ ; si l'on veut appliquer la formule au carré, il faut faire attention que ce codage distingue les côtés parallèles, contrairement à celui que nous avons étudié plus haut, d'où le résultat  $4n + 4$  qu'on obtient ici. On ne sait pas par contre quelle est la complexité du billard dans un polygone irrationnel, et encore moins ce qui se passe pour un polyèdre autre que le cube ou un cylindre sur un polygone rationnel.

On peut également s'intéresser aux généralisations de la propriété *iii*, c'est-à-dire aux suites engendrées par substitution; dans ce cas, B. Mossé (cf. [Mo]) a montré que la suite  $p(n + 1) - p(n)$  est bornée, donc que  $p(n)$  est sous-linéaire, et a donné un algorithme effectif de calcul de  $p(n)$ ; mais la complexité n'a pas en général de forme simple, même pour une suite substitutive très simple telle que la suite de Morse, obtenue en partant de **1**, en remplaçant **1** par **10** et **0** par **01** et en itérant. Pour d'autres résultats sur ce sujet, on peut aussi consulter [A].

### Bibliographie

- [A] J.P. ALLOUCHE, *Sur la complexité des suites infinies*, Prépublication.
- [AMST] P. ARNOUX, C. MAUDUIT, I. SHIOKAWA, J.I. TAMURA, *Complexity of sequences defined by billiards in the cube*, Bull. Soc. Math. France **122** (1994), 1–12.
- [AMM] P. ARNOUX, C. MAUDUIT, G. MEIGNIEZ, *Complexité de suites engendrées par des récurrences unipotentes*, En cours de rédaction.
- [CFS] I. P. CORNFELD, S.V. FOMIN, YA. G. SINAI, *Ergodic theory*, Springer Verlag, 1982.
- [CH] E.M. COVEN, G.A. HEDLUND, *Sequences with minimal block growth*, Mathematical Systems Theory **7** (1973), 138–153.
- [F] H. FÜRSTENBERG, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.
- [He] G. A. HEDLUND, *Sturmian minimal sets*, Amer. J. Math. **66** (1944), 605–620.
- [HM] G.A. HEDLUND, M. MORSE, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42.
- [Hu] P. HUBERT, *Complexité des suites définies par des trajectoires de billard dans un polygone rationnel*, Bull. Soc. Math. France (1993), à paraître.

- [Mo] B. MOSSÉ, *Notion de reconnaissabilité pour les substitutions et complexité des suites automatiques*, Prépublication N. 93-21 du Laboratoire de Mathématiques Discrètes (1993).

Laboratoire de Mathématiques discrètes

UPR CNRS 9016

Faculté des Sciences de Luminy

Case 930

163 avenue de Luminy

13288 Marseille Cedex 9

France

*Adresse électronique* : [arnoux@lmd.univ-mrs.fr](mailto:arnoux@lmd.univ-mrs.fr)