La suite de Morse-Thue

Définition et propriétés

Il s'agit d'un mot infini en binaire construit étape par étape à partir du mot initial $m_0 = 0$, A chaque étape, on lit le mot obtenu en pratiquant les substitutions $0 \rightarrow 01$ et $1 \rightarrow 10$ pour avoir le mot suivant. Cela donne :

$0 \rightarrow 01 \rightarrow 0110 \rightarrow 01101001 \rightarrow 0110100110010110 \rightarrow \ldots$ où à chaque fois la longueur du mot double.

Les deux formules de passage peuvent aussi s'écrire : $f(0) = 01$, et $f(1) = 10$ ou encore $f(\underline{0}) = \underline{01}$, le fait de souligner un mot signifiant que l'on a remplacé les 0 par des 1, et les 1 par des 0, ou encore que l'on ajoute 1 modulo 2. Cette fonction f permet de passer d'un mot au suivant par $f(m_n) = m_{n+1}$. Avec $m_1 = f(0) = 01$, on obtient $m_2 = f(01) = f(0)\,f(1) = 0110$, puis $m_3 = f(m_2) = f(0110) = f(0)f(1)f(1)f(0) = 01101001$, ou encore $m_3 = f(0110) = f(01)\,f(10) = m_2\,\underline{m_2}$. La fonction f a la particularité d'obéir à la relation $f(ab) = f(a)\,f(b)$ où a et b sont deux sous-mots dont la concaténation donne le mot ab. On dit qu'il s'agit d'un morphisme.

Ces mots ont les propriétés suivantes :

- $m_{n+1} = m_n\,\underline{m_n}$, Cette propriété exprime que l'on passe au mot suivant par un simple allongement du mot déjà obtenu, en rajoutant derrière lui le mot obtenu en lui ajoutant 1 modulo 2..

Cela se démontre par récurrence : c'est vrai au départ, puisque $m_1 = 01 = m_0\,\underline{m_0}$ .Supposons la formule vraie au rang n : $m_n = m_{n-1}\,\underline{m_{n-1}}$, alors $m_{n+1} = f(m_n) = f(m_{n-1}\,\underline{m_{n-1}}) = f(m_{n-1})\,f(\underline{m_{n-1}}) = m_n\,\underline{m_n}$, et la formule reste vraie au rang $n+1$.

- En appelant a_n le chiffre en position n dans la suite de Morse-Thue, la succession de ces chiffres s'écrit par blocs de deux chiffres qui sont soit 01 soit 10. Plus précisément on a les relations de récurrence : $a_{2n} = a_n$, et $a_{2n+1} = \underline{a_{2n}}$, avec au départ $a_0 = 0$. Et a_n est aussi le nombre de 1 dans l'écriture binaire de n ramené modulo 2.

C'est une conséquence de la définition de la suite. Lorsque l'on passe du mot m_n au suivant, on écarte les chiffres de m_n en laissant une place entre les chiffres successifs, les termes de m_n occupant les positions paires, d'où $a_{2n} = a_n$, puis on remplit les places vides en position impaire en mettant $\underline{a_{2n}}$ derrière a_{2n}. Les relations de récurrence obtenues avec la condition initiale supplémentaire permettent de construire la suite chiffre après chiffre. On vérifie aisément que le nombre de 1 ramené modulo 2 dans l'écriture binaire de n vérifie exactement les mêmes relations.

Remarque : Si l'on a vu comment passer du mot m_n au mot suivant m_{n+1}, on sait aussi passer d'un mot au mot précédent, il suffit de ne garder que les chiffres en position paire.

La régularité de la construction de la suite de Morse-Thue pourrait faire penser que cette suite est périodique, c'est-à-dire qu'elle serait formée de blocs identiques répétés, leur longueur T étant la période de la suite. On va constater qu'il n'en est rien. Pour cela nous avons besoin de quelques notions nouvelles.

Notions de carré, de cube, et d'entrelacement

- carré : On dit qu'un mot contient un carré s'il possède deux blocs (ou facteurs ou sous-mots) successifs identiques. Par exemple le mot 011011001 contient les carrés 11, puis 11 puis 00, ainsi que les carrés 011011 et 110110. Il est aisé de constater que les seuls mots en binaire n'ayant pas de carré sont 0, 1, 01, 10, 010, 101.

- cube : un mot contient un cube lorsqu'il possède trois blocs successifs identiques, comme par exemple le mot 01110, qui a comme cube 111, ou le mot 11010100 qui contient le cube 101010.

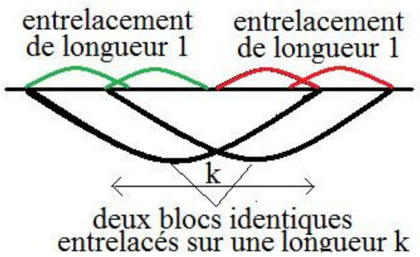
- entrelacement : un mot contient un entrelacement s'il contient deux blocs identiques ayant une partie en commun, une fin du premier bloc étant aussi un début du deuxième. Par exemple le mot 11010100 contient l'entrelacement 10101 où le premier bloc 101 a sa dernière lettre 1 qui est le début du deuxième bloc 101. Ce même mot a aussi l'entrelacement 101010 où les deux blocs 1010 ont en commun 10.

On a les propriétés suivantes :

1) Si un mot présente un entrelacement avec une partie commune de longueur k supérieure à 1, il existe aussi un entrelacement avec une partie commune de longueur 1.

2) Un mot (en binaire) possède un entrelacement si et seulement s'il contient un sous-mot de la forme $avava$, a étant soit 0 soit 1, v et w étant des sous-mots (éventuellement vides). Ainsi la présence ou non de $0v0v0$ ou $1w1w1$ signifie l'existence ou non d'un entrelacement.

Démonstration de la propriété 1 :



A partir d'un entrelacement avec une partie commune de longueur k, il est possible de fabriquer deux entrelacements avec une partie commune de longueur 1, comme on le voit sur le dessin.

Démonstration de la propriété 2 :

D'abord si l'on a la présence de $avava$ dans le mot, il y a bien entrelacement, puisque le bloc ava se trouve répété deux fois avec en commun la lettre a dans $avava$.

Inversement, si l'on a un entrelacement, il existe aussi un entrelacement de longueur 1. En appelant a la première lettre de chacun des deux blocs entrelacés, a est aussi la dernière lettre du premier bloc puisque c'est la première du deuxième bloc. Chacun de ces blocs s'écrit ava, et l'entrelacement s'écrit $avava$.

3) Conséquence : si un mot possède un cube, il a aussi un entrelacement. Ou bien, s'il n'a aucun entrelacement, il n'a aucun cube.

En effet, si le mot contient un cube, chacun des trois facteurs du cube s'écrit *av* en notant *a* la première lettre du facteur, d'où la présence de *avavav*, et à fortiori de *avava*, ce qui crée un entrelacement.

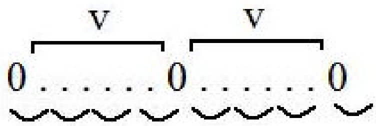
Nous allons maintenant montrer que la suite de Morse-Thue ne présente aucun entrelacement. Elle n'aura donc aucun cube, et par suite, elle ne sera sûrement pas périodique.

Propriété : la suite de Morse-Thue ne contient jamais un sous-mot de la forme 0*v*0*v*0 ni 1*v*1*v*1

On distingue deux cas :

- le bloc v a une longueur impaire. Prenons $0v0v0$, les trois 0 étant tous en position paire, ou en position impaire. S'ils sont en position paire dans le mot m_n, on prend le mot précédent (qui est aussi présent dans le mot m_n), il va contenir $0v'0v'0$, avec v' de longueur moitié. S'ils sont en position impaire, on a aussi la présence de $10v_1 10v_1 10$ avec $v = v_1 1$. Les 1 étant en position paire, le mot précédent contient $1v'1v'1$ avec v' de longueur moitié. Dans tous les cas, on en arrive à la présence d'un bloc $av'av'a$ plus court (environ de moitié) de celui que l'on avait déjà. Au cas où les v' sont de longueur impaire, on recommence, sinon on est dans le cas où v' est de longueur paire. Dans le meilleur des cas, les blocs v restant de longueur impaire, on tombe finalement sur le facteur 01010, et ce facteur provient à son tour de 000 ou 111. Or la suite de Morse-Thue ne contient jamais 000 ni 111.

- le bloc v est de longueur paire. Si sa longueur est nulle, on tombe sur l'impossibilité 000 ou 111. Sinon on est dans le contexte suivant :



On s'aperçoit, en groupant les chiffres par deux, que ces blocs successifs sont décalés d'un cran dans le deuxième mot v, ce qui les enchaîne les uns aux autres. Cela permet de remplir au coup par coup, par la gauche et par la droite ce qui se trouve dans v, où l'on a une succession 0101… des deux côtés, et l'on aboutit toujours à un bloc central 00 ou 11, ce qui est impossible.

Une application : construction d'un mot infini à base de trois lettres sans la présence d'aucun carré [1]

La suite infinie de Morse-Thue est ponctuée de 0. Entre deux de ces 0 qui se succèdent, il y a soit rien, soit un 1, soit deux 1. Il ne peut pas y en avoir trois puisque la suite n'a aucun cube. Procédons aux substitutions suivantes : $0 \to 0 \quad 01 \to 1 \quad 011 \to 2$. Cela donne un mot B infini à base des lettres 0, 1, 2. Inversement à partir de ce mot B on obtient la suite de Morse-Thue en faisant $0 \to 0 \quad 1 \to 01 \quad 2 \to 011$. Nous allons montrer que ce mot B ne présente aucun carré. En effet, si B possédait un carré mm, il aurait aussi le facteur mma, où a est la lettre qui suit (0 ou 1 ou 2). En procédant aux

[1] On a vu qu'il n'existe pas de mot infini à base de deux lettres sans carré.

substitutions menant à la suite de Morse-Thue, *m* devient un mot commençant par 0, soit 0*v*, et *a* devient un mot commençant par *0*. On obtiendrait alors le facteur 0*v*0*v*0, ce qui est impossible dans la suite de Morse-Thue. On a donc trouvé une suite *B* infinie sans carré, soit *B* = 210201210120210…

Références bibliographiques :

M. Lothaire, *Combinatorics on words*, Addison-Wesley 1983.

J.P.Allouche, J. Shallit, *Automatic Sequences* , Cambridge University Press, 2003.