Définition

Soit Σ un alphabet. On appelle *mot infini* sur Σ une suite $u = (a_i)_{i\in\mathbb{N}}$ telle que pour $i \in \mathbb{N}, a_i \in \Sigma$. On note Σ^ω l'ensemble des mots infinis.

Un mot infini $u = a_0a_1\dots$ est dit *périodique* s'il existe $p \in \mathbb{N}^*$ tel que pour $i \in \mathbb{N}$, $a_{i+p} = a_i$. Il est dit *ultimement périodique* s'il existe $v \in \Sigma^*$ et $w \in \Sigma^\omega$, w périodique, tel que $u = vw$.

Si $u \in \Sigma^*$, on note u^ω le mot infini constitué des répétitions de u.

Définition: Topologie

Soient $u, v \in (\Sigma^* \cup \Sigma^\omega)^2$. On note $\ell(u, v)$ la longueur du plus long préfixe commun à u et v. C'est une valeur de $\mathbb{N} \cup \{+\infty\}$. On note $d(u, v) = 2^{-\ell(u,v)}$. La fonction d est une distance sur Σ^ω.

On dit qu'une suite $(u_n)_{n\in\mathbb{N}} \in (\Sigma^*)^{\mathbb{N}}$ converge vers $u \in \Sigma^\omega$ si et seulement si $d(u_n, u) \underset{n\to+\infty}{\longrightarrow} 0$.

Définition

Soit $\mu : \Sigma^* \to \Sigma^*$. On dit que μ est un *morphisme de mots* si et seulement si pour $u, v \in \Sigma^*$, on a $\mu(uv) = \mu(u)\mu(v)$. Si pour tout $a \in \Sigma$, $\mu(a) \neq \varepsilon$, on étend naturellement μ aux mots infinis.

Si une suite de mots est définie par $u_0 \in \Sigma^*$ et pour $n \in \mathbb{N}$, $u_{n+1} = \mu(u_n)$ et que (u_n) converge vers un mot infini, on note $\mu^\omega(u_0)$ sa limite. C'est un point fixe pour μ.

Exercice 1

Soit μ un morphisme de mots et $a \in \Sigma$ tel qu'il existe $u \in \Sigma^*$ vérifiant $\mu(a) = au$. Soit (u_n) la suite définie par $u_n = \mu^n(a)$.

1. Montrer que pour tout n, u_n est un préfixe de u_{n+1}.
2. On suppose $u \neq \varepsilon$. Montrer que si pour tout $b \in \Sigma$, $\mu(b) \neq \varepsilon$, alors (u_n) converge vers un mot infini.
3. On pose $\mu : \begin{cases} 0 & \mapsto & 01 \\ 1 & \mapsto & 221 \\ 2 & \mapsto & 2 \end{cases}$ pour $\Sigma = \{0, 1, 2\}$. Déterminer $\mu^\omega(0)$.
4. On pose $\mu : \begin{cases} 0 & \mapsto & 1 \\ 1 & \mapsto & 10 \end{cases}$. Quelle est la suite de mots formée par $(\mu^n(0))_{n\in\mathbb{N}}$?

Corrigé

1. On montre ce résultat par récurrence sur n en montrant que $u_{n+1} = u_n\mu^n(u)$:
   - Si $n = 0$, alors $u_0 = a$ et $u_1 = au$. La propriété est vérifiée.
   - Supposons le résultat vrai pour $n \in \mathbb{N}$ fixé. Alors $u_{n+2} = \mu(u_{n+1})$.

     Or $\mu(u_{n+1}) = \mu(u_n\mu^n(u)) = \mu(u_n)\mu^{n+1}(u) = u_{n+1}\mu^{n+1}(u)$ donc la propriété est vraie pour $n + 1$.
2. Le morphisme n'est pas effaçant, donc $|u_{n+1}| > |u_n|$ pour tout $n \in \mathbb{N}$. Grâce à la question précédente, on en déduit que (u_n) converge vers $u = a \bigodot_{n=0}^{+\infty} \mu^n(u)$ où le point désigne ici la concaténation. En effet, u_n est toujours un préfixe de u, dont la taille tend vers $+\infty$.
3. On a $\mu(0) = 01$, $\mu^2(0) = 01221$, $\mu^3(0) = 0122122221$, ... Montrons par récurrence sur n, en étudiant $\mu^n(0)$, que $\mu^\omega(0) = a_0a_1a_2\dots$ où $a_0 = 0$, $a_i = 1$ si i est un carré et $a_i = 2$ sinon. On montrera entre autre que $|\mu^n(0)| = 1 + n^2$.
   - La propriété est vraie pour $n = 0$, car $\mu^0(0) = 0$.
   - Supposons la propriété vraie pour $n \in \mathbb{N}$ fixé. Alors $\mu^{n+1}(0) = \mu^n(0)\mu^n(1)$. Or, $\mu^n(1) = 2^{2n}1$. Dès lors, $|\mu^{n+1}(0)| = |\mu^n(0)| + 2n + 1 = 1 + (n + 1)^2$. De plus, le seul 1 rajouté se trouve à

l'indice $(n+1)^2$. Les autres lettres rajoutées sont des 2, et aucun indice n'est un carré entre l'indice n^2 et $(n+1)^2$.

4. On a $\mu(0) = 1$, $\mu^2(0) = 10$, $\mu^3(0) = 101$, $\mu^4(0) = 10110$, ... On reconnaît ici la suite des mots de Fibonacci.

Proposition

Soit $F \subseteq \Sigma^*$. Alors les deux propositions suivantes sont équivalentes :

1. Il existe une infinité de mots sans facteur dans F.
2. Il existe un mot infini n'ayant aucun facteur dans F.

Exercice 2

Montrer la proposition précédente. On commencera par montrer le *lemme de König* : tout arbre infini d'arité finie possède une branche infinie.

Corrigé

On commence par montrer le lemme de König. On construit la branche de la façon suivante : on choisit la racine de l'arbre. Cette racine a un nombre fini de fils. L'arbre étant infini, il existe un des fils qui est infini. On itère alors le procédé précédent pour continuer la branche infinie.

Dès lors, pour montrer $1 \Rightarrow 2$, créons un arbre d'arité $|\Sigma|$, ayant pour racine le mot vide, et où chaque niveau de profondeur k représente l'ensemble Σ^k. De la même façon que le lemme de König, on peut créer une suite infinie de mots en choisissant à chaque étape le fils qui contient une infinité de mots sans facteur dans F. On sait qu'il existe toujours un mot de Σ^k qui est sans facteur dans F, car il suffit de prendre un préfixe d'un mot plus long. On crée ainsi un mot infini n'ayant aucun facteur dans F.
Réciproquement, si on suppose l'existence d'un mot infini n'ayant aucun facteur dans F, l'ensemble des préfixes de ce mots est un ensemble infini de mots sans facteur dans F.

Définition

On définit le morphisme de *Thue-Morse* par $\tau : \left\{ \begin{array}{ccc} 0 & \mapsto & 01 \\ 1 & \mapsto & 10 \end{array} \right.$. Le mot de Thue-Morse est le mot $\tau^\omega(0)$.

Exercice 3

1. Déterminer les 32 premières lettres de $\tau^\omega(0)$.
2. Montrer que pour $n \in \mathbb{N}^*$, $\tau^n(0)10$ est préfixe de $\tau^{n+1}(0)$.
3. Montrer que la n-ème lettre (en commençant à 0) de $\tau^\omega(0)$ vaut 0 si l'écriture de n en binaire comporte un nombre pair de 1, et vaut 1 sinon.

Corrigé

1. Le mot cherché est $\tau^5(0) = 011010011001011010010110011001$.
2. Par récurrence sur $n \in \mathbb{N}^*$:
   - on a $\tau(0) = 01$ et $\tau^2(0) = 0110$, donc le résultat est établi pour $n = 1$;
   - supposons le résultat vrai pour $n \in \mathbb{N}^*$ fixé. Il existe donc $u \in \{0,1\}^*$ tel que $\tau^{n+1}(0) = \tau^n(0)10u$. Alors $\tau^{n+2}(0) = \tau(\tau^{n+1}(0)) = \tau(\tau^n(0)10u) = \tau^{n+1}(0)\tau(10)\tau(u) = \tau^{n+1}(0)1001\tau(u)$ ce qui permet de conclure.
3. Notons $\tau^\omega(0) = a_0a_1a_2\ldots$. Montrons par récurrence sur k que le résultat est vrai pour tout

$n < 2^k$:

- On a $\tau(0) = 01$, donc le résultat est vrai pour $k = 1$.
- Supposons le résultat vrai pour $k \in \mathbb{N}^*$ fixé. Remarquons que pour $j \in [\![0; 2^{k-1}-1]\!]$, $a_{2^k+2j} = a_{2^{k-1}+j}$. et $a_{2^k+2j+1} = 1 - a_{2^{k-1}+j}$. Or, $2j$ et j ont la même parité de 1, et comme $j < 2^{k-1}$, les nombres $2^k + 2j$ et $2^{k-1} + j$ également. On en déduit que le résultat est vrai pour les indices pairs entre 2^k compris et 2^{k+1} non compris. De plus, les nombres $2j + 1$ et j n'ont pas la même parité de 1, et $2^k + 2j + 1$ et $2^{k-1} + j$ non plus. Comme les lettres à ces indices ne sont jamais les mêmes, on en déduit que le résultat est également vrai pour les indices impairs.

Définition: Chevauchement

Un chevauchement est un mot de la forme $uvuvu$ où $u \in \Sigma^+$ et $v \in \Sigma^*$. Un mot sans chevauchement est un mot dont aucun facteur n'est un chevauchement.

Exercice 4

On cherche à montrer que le mot de Thue-Morse est sans chevauchement. Pour ce faire, on note $L = (01|10)^*$.

1. Montrer que si $u \in L$, alors $0u0$ et $1u1$ ne sont pas dans L.

On suppose qu'il existe $n \in \mathbb{N}$ tel que $\tau^n(0)$ ne contient pas de chevauchement et $w = \tau^{n+1}(0)$ contient un chevauchement $uvuvu$, $u \in \Sigma^+$, $v \in \Sigma^*$.

2. Montrer qu'on peut choisir u tel que $|u| = 1$.

On suppose donc qu'il existe $v, x, y \in \Sigma^*$ et $a \in \Sigma$ tels que $w = xavavay$.

3. Montrer que $|x|$ et $|y|$ ont une parité différente.
4. En supposant que $|x|$ est pair, montrer qu'on arrive à une contradiction.
5. Conclure.

Corrigé

1. On remarque que pour $u \in L$, $|u|_0 = |u|_1$, ce qui permet de conclure.
2. Supposons que $|u| > 1$. Alors $u = au'$, avec $u' \in \Sigma^+$. Dès lors, $uvuvu = au'vau'vau'$. En posant $v' = u'v$, on a donc $av'av'a$ est un facteur de w.
3. Par définition de τ, $w \in L$. On en déduit que $|w|$ est pair. Or, $|w| = |x| + |y| + 2|v| + 3$. On en déduit que $|x|$ et $|y|$ sont de parité différente.
4. Supposons $|x|$ de longueur paire. Comme $w \in L$, on en déduit que :
   - Si $|v|$ est paire, alors ava et v sont des mots de L. C'est absurde d'après la première question.
   - Si $|v|$ est impaire, alors $v = bv_1$ où $b = 1 - a$. On remarque que de plus, $y = by_1$. On a donc $w = xabv_1abv_1aby_1$. Comme x est de longueur paire, alors abv_1abv_1ab est l'image d'un facteur bv_2bv_2b de $\tau^n(0)$. C'est absurde par hypothèse.
5. On raisonne sur le mot w^R si $|y|$ est paire. On conclut par récurrence en remarquant que 0 ne contient pas de chevauchement.

Définition: Carré

Un carré est un mot de la forme uu où $u \in \Sigma^*$. Un mot sans carré est un mot qui n'a aucun facteur qui est un carré autre que ε.

Exercice 5

Déterminer les mots sans carré sur $\Sigma = \{0, 1\}$.

Corrigé

Les seuls mots sans carrés sur cet alphabet sont ε, 0, 1, 10, 01, 101 et 010. En effet, les autres mots de taille inférieure à 3 comportent une répétition de 2 lettres identiques. Les mots de taille supérieure à 4 contiendront soit une répétition de 2 lettres identiques, soit 1010, soit 0101.

Exercice 6

On cherche à montrer le théorème de Thue (1906) : il existe un mot infini sans carré sur $\Sigma = \{0, 1, 2\}$.

On pose pour cela $\sigma : \begin{cases} 0 & \mapsto & 0 \\ 1 & \mapsto & 01 \\ 2 & \mapsto & 011 \end{cases}$

1. Montrer que $\sigma(\{0, 1, 2\}^+)$ est l'ensemble des mots commençant par 0 et ne contenant pas 111 comme facteur.
2. Montrer que pour $n \in \mathbb{N}$, il existe un unique $u_n \in \{0, 1, 2\}^*$ tel que $\sigma(u_n) = \tau^n(0)$.

Pour $n \in \mathbb{N}$, on note v_n le mot u_n privé de sa dernière lettre.

3. Montrer que pour tout $n \in \mathbb{N}$, v_n est préfixe de v_{n+1}. En déduire la convergence de (v_n) et celle de (u_n).

On note $\sigma^{-1}(\tau^\omega(0))$ cette limite.

4. Montrer que $\sigma^{-1}(\tau^\omega(0))$ est sans carré.
5. Déterminer les 15 premières lettres de $\sigma^{-1}(\tau^\omega(0))$.

Corrigé

1. Pour la première inclusion, on remarque que si $u = a_1 \ldots a_n \in \{0, 1, 2\}^+$, alors $\sigma(u)$ possède $\sigma(a_1)$ comme préfixe, donc commence par 0. Par ailleurs, montrons par induction sur u que $\sigma(u)$ ne contient pas 111 comme facteur :
   - c'est vrai pour $n = 1$ d'après la forme des images des lettres de $\{0, 1, 2\}$;
   - supposons le résultat vrai pour $u \in \{0, 1, 2\}^+$ et soit $a \in \{0, 1, 2\}$. Alors $\sigma(ua) = \sigma(u)\sigma(a)$. Comme $\sigma(u)$ ne contient pas 111 comme facteur, termine par au plus deux 1 et que $\sigma(a)$ ne contient pas 111 comme facteur et commence par 0, on en déduit que $\sigma(ua)$ ne contient pas 111 comme facteur.

   On conclut par récurrence.

   Réciproquement, soit $v \in \{0, 1\}^+$ commençant par 0 et ne contenant pas 111 comme facteur. Alors $v = 0^{i_1}1^{j_1}0^{i_2}1^{j_2} \ldots 0^{i_n}1^{j_n}$ où pour tout $k \in [\![1, n]\!]$:
   - $i_k \in \mathbb{N}^*$;
   - $j_k \in \{1, 2\}$ et $j_n \in \{0, 1, 2\}$.

   En effet, il y a au plus deux 1 consécutifs entre deux suites non vides de 0. Posons $u = 0^{i_1-1}j_1 0^{i_2-1}j_2 \ldots 0^{i_n-1}j_n$. Alors $u \in \{0, 1, 2\}^+$ et $\sigma(u) = v$.
2. On a montré l'existence dans la deuxième partie de la réponse précédente. Pour l'unicité, supposons qu'il existe $u = a_1 \ldots a_n \neq u' = b_1 \ldots b_m$ deux mots de $\{0, 1, 2\}^*$ tels que $\sigma(u) = \sigma(u')$. Remarquons que u ne peut pas être préfixe de u' (ou réciproquement), car sinon $\sigma(u) \neq \sigma(u')$ (car σ n'est pas effaçant). Soit alors $i \in [\![1, \min(n, m)]\!]$ minimal tel que $a_i \neq b_i$. Alors $\sigma(a_i) \neq \sigma(b_i)$, et σ étant un morphisme, $\sigma(a_1 \ldots a_i) \neq \sigma(b_1 \ldots b_i)$. On conclut que $\sigma(u) \neq \sigma(u')$, d'où l'injectivité de σ.
3. Remarquons que la dernière lettre de u_n ne peut pas être un 2, car sinon $\tau^n(0)$ terminerait par 11, ce qui n'est pas possible. Remarquons que $u_0 = 0$, $u_1 = 1$ et $u_2 = 20$, soit $v_0 = v_1 = \varepsilon$ et $v_2 = 2$. v_n est bien préfixe de v_{n+1} pour $n = 0$ et $n = 1$. Par ailleurs, pour $n \geqslant 2$, comme $\tau^n(0)$ est préfixe de $\tau^{n+1}(0)$, $\tau^{n+1}(0) = \tau^n(0)abw$ où $ab \in \{01, 10\}$ et $w \in \{0, 1\}^*$. Si $a = 0$, alors u_n est préfixe de u_{n+1}, donc v_n est préfixe de v_{n+1}. Sinon, $ab = 10$ et alors si $u_n = v_n c$, avec $c \in \{0, 1\}$, on aura u_{n+1} qui commence par $v_n(c+1)$, donc v_n est préfixe de v_{n+1}. De plus, comme $|v_{n+1}| > |v_n|$ pour $n \geqslant 1$, on en déduit que (v_n) converge, puis que (u_n) converge.

4. Supposons que v_n contienne un carré uu. Quitte à considérer v_{n+1}, on peut supposer que ce carré n'est pas à la fin de v_n. Il est donc suivi par une lettre $a \in \{0, 1, 2\}$, et uua est donc un facteur de v_n. On en déduit que $\sigma(u)\sigma(u)\sigma(a)$ est un facteur de $\tau^n(0)$. Or, $\sigma(u)$ et $\sigma(a)$ commencent par un zéro. On en déduit que $\tau^n(0)$ a un chevauchement de la forme $0v0v0$, ce qui est absurde.
5. On trouve l'antécédent de $\tau^5(0)$, qui est 210201210120201. On remarque qu'il faut transformer le dernier 1 en 2, car $\tau^6(0)$ a $\tau^5(0)10$ comme préfixe.