

## DS 3 : Mots de Christoffel, mots sturmiens

### • L'algorithme de Christoffel

#### Question 1.

1.a Considérons la division euclidienne de  $n$  par  $q$  :  $n = bq + r$  avec  $r \in \llbracket 0, q - 1 \rrbracket$ . Alors :

$$\begin{cases} \lfloor n\alpha \rfloor = \lfloor bp + r\alpha \rfloor = bp + \lfloor r\alpha \rfloor \\ \lfloor (n+1)\alpha \rfloor = \lfloor bp + (r+1)\alpha \rfloor = bp + \lfloor (r+1)\alpha \rfloor \end{cases}$$

donc  $c_n = c_r$ . De ceci il résulte que  $c$  est périodique de période  $q$ .

1.b Supposons  $c$  ultimement périodique : il existe deux entiers  $n_0$  et  $q$  tels que  $c_{n+q} = c_n$  pour  $n \geq n_0$ . Alors :

$$\forall n \geq n_0, \quad \lfloor (n+q+1)\alpha \rfloor - \lfloor (n+q)\alpha \rfloor = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor$$

soit encore :  $\lfloor (n+q+1)\alpha \rfloor - \lfloor (n+1)\alpha \rfloor = \lfloor (n+q)\alpha \rfloor - \lfloor n\alpha \rfloor$ . Ceci montre que la suite  $(\lfloor (n+q)\alpha \rfloor - \lfloor n\alpha \rfloor)_{n \in \mathbb{N}}$  est stationnaire ; notons  $p$  la valeur du stationnement. Il s'agit d'un entier puisque la suite est à valeurs entières.

Pour tout  $n \geq n_0$ ,  $\lfloor n\alpha + q\alpha \rfloor = \lfloor n\alpha \rfloor + p$  donc pour tout  $k \in \mathbb{N}$ ,  $\lfloor n_0\alpha + kq\alpha \rfloor = \lfloor n_0\alpha \rfloor + kp$  ; ainsi :

$$\forall k \in \mathbb{N}, \quad \frac{\lfloor n_0\alpha + kq\alpha \rfloor}{k} = \frac{\lfloor n_0\alpha \rfloor}{k} + p.$$

En faisant tendre  $k$  vers  $+\infty$  on obtient :  $q\alpha = 0 + p$ , soit  $\alpha = \frac{p}{q}$ . Il s'agit bien d'un nombre rationnel.

#### Question 2.

$aa$  est un facteur de  $c$  lorsqu'il existe un entier  $n$  tel que  $\lfloor (n+2)\alpha \rfloor = \lfloor (n+1)\alpha \rfloor = \lfloor n\alpha \rfloor$ .

L'égalité :  $\lfloor (n+2)\alpha \rfloor = \lfloor n\alpha \rfloor$  impose :  $n\alpha < (n+2)\alpha + 1$ , soit  $\alpha < \frac{1}{2}$ .

$bb$  est un facteur de  $c$  lorsqu'il existe un entier  $n$  tel que  $\lfloor (n+2)\alpha \rfloor = \lfloor (n+1)\alpha \rfloor + 1 = \lfloor n\alpha \rfloor + 2$ .

L'égalité :  $\lfloor (n+2)\alpha \rfloor = \lfloor n\alpha \rfloor + 2$  impose :  $n\alpha + 1 < (n+2)\alpha$ , soit  $\alpha > \frac{1}{2}$ .

Ainsi, dans tout mot de Christoffel, l'un des mots  $aa$  (lorsque  $\alpha \geq \frac{1}{2}$ ) ou  $bb$  (lorsque  $\alpha \leq \frac{1}{2}$ ) n'apparaît pas.

On constate que lorsque  $\alpha = \frac{1}{2}$ , aucun de ces deux mots n'est facteur de  $c$  ; en effet,  $c = ababababa \dots$ .

#### Question 3.

3.a Notons  $m = c_k \dots c_{k+n-1}$ . Par définition de  $p$ ,  $\lfloor (k+n)\alpha \rfloor = \lfloor k\alpha \rfloor + (n-p)$  (car  $n-p$  est le nombre de  $b$  dans le mot  $m$ ). On en déduit :

$$k\alpha + n - p - 1 < (k+n)\alpha < k\alpha + n - p + 1 \quad \text{soit} \quad n(1-\alpha) - 1 < p < n(1-\alpha) + 1.$$

3.b Considérons un autre facteur  $m'$  de longueur  $n$  de  $c$ , et  $p' = |m'|_a$ . D'après la question précédente,  $p$  et  $p'$  sont deux entiers de l'intervalle  $]n(1-\alpha) - 1, n(1-\alpha) + 1[$ . Or cet intervalle, d'amplitude 2, contient au plus deux entiers ; ainsi,  $|p - p'| \leq 1$ .

### • Mots sturmiens

#### Question 4.

4.a Les seuls mots vérifiant :  $P_u(n) = 1$  pour tout  $n \in \mathbb{N}$  sont les mots  $aaaaaaa \dots$  et  $bbbbbbb \dots$  (il n'y en a pas d'autre car  $u$  ne doit posséder qu'un seul facteur de longueur 1, c'est à dire qu'une seule des deux lettres).

- 4.b** Lorsque  $u = abababa \dots$ , nous avons pour tout  $p \in \mathbb{N}^*$ ,  $\mathcal{L}_{2p}(u) = \{(ab)^p, (ba)^p\}$  et  $\mathcal{L}_{2p+1}(u) = \{a(ba)^p, b(ab)^p\}$   
 donc  $P_u(n) = \begin{cases} 1 & \text{si } n = 0 \\ 2 & \text{si } n \geq 1 \end{cases}$ .
- 4.c** Il s'agit ici de construire un mot de complexité maximale, puisque  $\text{card } A^n = 2^n$ ; ainsi, tout mot fini doit être facteur de ce mot infini  $u$ .  
 On peut par exemple d'ordonner  $A^+$  par ordre croissant de longueur, et à longueur égale suivant l'ordre lexicographique :  $u_0 = a$ ,  $u_1 = b$ ,  $u_2 = aa$ ,  $u_3 = ab$ ,  $u_4 = ba$ ,  $u_5 = bb$ ,  $u_6 = aaa$ ,  $u_7 = aab \dots$  et de poser  $u = u_0 u_1 u_2 \dots = abaaabbabbaaaaaab \dots$  pour obtenir un exemple d'un tel mot.

### Question 5.

- 5.a** Considérons l'application  $\varphi : \mathcal{L}_{n+1}(u) \rightarrow \mathcal{L}_n(u)$  qui à un mot de longueur  $n+1$  associe son préfixe de longueur  $n$ . Cette application est à l'évidence surjective, donc  $\text{card } \mathcal{L}_{n+1}(u) \geq \text{card } \mathcal{L}_n(u)$ , soit :  $P_u(n) \leq P_u(n+1)$ .
- 5.b** Considérons l'application  $\psi : \mathcal{L}_{n_1+n_2} \rightarrow \mathcal{L}_{n_1}(u) \times \mathcal{L}_{n_2}(u)$  qui à un mot de longueur  $n_1+n_2$  associe le couple formé de son préfixe de longueur  $n_1$  et son suffixe de longueur  $n_2$ . Cette application est à l'évidence injective, donc  $P_u(n_1+n_2) \leq P_u(n_1) + P_u(n_2)$ .
- 5.c** Considérons la division euclidienne de  $n$  par  $n_0$  :  $n = n_0q + r$  avec  $r \in \llbracket 0, n_0 - 1 \rrbracket$ . D'après la question précédente,  $P_u(n) \leq P_u(n_0)^q P_u(r)$  donc  $2^{-n} P_u(n) \leq 2^{-r} P_u(r) \times \left(\frac{P_u(n_0)}{2^{n_0}}\right)^q \leq \left(\frac{P_u(n_0)}{2^{n_0}}\right)^q$ . Sachant que  $\frac{P_u(n_0)}{2^{n_0}} < 1$ , on en déduit :  $\lim_{n \rightarrow +\infty} 2^{-n} P_u(n) = 0$ .

### Question 6.

- 6.a** Soit  $u$  un mot ultimement périodique; on peut donc l'écrire  $u = vw$ , où  $v$  est un mot fini et  $w$  un mot infini périodique.  
 Un facteur  $x$  de longueur  $n$  de  $u$  peut toujours s'écrire  $x = yz$ , où  $y$  est un facteur de  $v$  de longueur  $k$  et  $z$  un facteur de  $w$  de longueur  $n-k$  (avec  $k \in \llbracket 0, n \rrbracket$ ).  $v$  est fini donc possède au plus  $2^{|v|}$  facteurs.  $w$  est périodique de période  $p$ , donc le nombre de facteurs de longueur  $n-k$  de  $w$  est majoré par  $p$ . Ainsi,  $\text{card } \mathcal{L}_n(u) \leq p 2^{|v|}$ . La suite  $(P_u(n))_{n \in \mathbb{N}}$  est donc majorée. Sachant qu'il s'agit d'une suite croissante d'entiers, on peut même dire qu'elle est stationnaire.
- 6.b** Considérons de nouveau l'application  $\varphi : \left( \begin{array}{ccc} \mathcal{L}_{n_0+1}(u) & \longrightarrow & \mathcal{L}_{n_0}(u) \\ u_k \dots u_{k+n_0} & \longmapsto & u_k \dots u_{k+n_0-1} \end{array} \right)$  qui à un facteur de longueur  $n_0+1$  associe son préfixe de longueur  $n_0$ . Il s'agit d'une application surjective, et l'hypothèse faite sur  $n_0$  montre qu'elle est en fait bijective.  
 Il en est de même de l'application  $\psi : \left( \begin{array}{ccc} \mathcal{L}_{n_0+1}(u) & \longrightarrow & \mathcal{L}_{n_0}(u) \\ u_k \dots u_{k+n_0} & \longmapsto & u_{k+1} \dots u_{k+n_0} \end{array} \right)$  qui à un facteur de longueur  $n_0+1$  associe son suffixe de longueur  $n_0$ .  
 L'application  $\sigma = \psi \circ \varphi^{-1} : \left( \begin{array}{ccc} \mathcal{L}_{n_0}(u) & \longrightarrow & \mathcal{L}_{n_0}(u) \\ u_k \dots u_{k+n_0-1} & \longmapsto & u_{k+1} \dots u_{k+n_0} \end{array} \right)$  est donc une permutation de  $\mathcal{L}_{n_0}(u)$ . Comme toute permutation, elle est d'ordre fini  $p \leq \text{card } \mathcal{L}_{n_0}(u) = P_u(n_0)$  : on a  $\sigma^p = \text{Id}$ . Or  $\sigma^p(u_k \dots u_{k+n_0-1}) = u_{k+p} \dots u_{k+n_0-1+p}$ ; on a donc  $u_{k+p} = u_k$  pour tout entier  $k \in \mathbb{N}$ , ce qui prouve que  $u$  est ultimement périodique, de période inférieure ou égale à  $P_u(n_0)$ .
- 6.c** D'après ce qui précède, si  $u$  n'est pas ultimement périodique, alors pour tout  $n \in \mathbb{N}$ ,  $P_u(n+1) \geq P_u(n) + 1$ . On en déduit par récurrence :  $P_u(n) \geq P_u(0) + n = n + 1$ .

### Question 7.

- 7.a** Considérons de nouveau l'application surjective  $\varphi : \mathcal{L}_{n+1}(u) \rightarrow \mathcal{L}_n(u)$  qui à un mot de longueur  $n+1$  associe son préfixe de longueur  $n$ . Sachant que  $\text{card } \mathcal{L}_{n+1}(u) = \text{card } \mathcal{L}_n(u) + 1$ , un seul facteur  $m$  de longueur  $n$  possède deux antécédents (qui sont donc  $ma$  et  $mb$ ), les autres n'en ayant qu'un; ils se prolongent à droite de manière unique.
- 7.b** Nous allons montrer qu'un suffixe  $v$  d'un mot sturmien  $u$  est aussi sturmien.

Puisque  $u$  n'est pas ultimement périodique,  $v$  non plus, et donc :  $\forall n \in \mathbb{N}, P_v(n) \geq n + 1$ .

Mais tout facteur de  $v$  est facteur de  $u$ , donc  $P_v(n) \leq P_u(n) = n + 1$ . On en déduit que pour tout  $n \in \mathbb{N}$ ,  $P_v(n) = n + 1$  ;  $v$  est sturmien.

Supposons maintenant qu'un facteur  $x$  de  $u$  n'apparaisse qu'un nombre fini de fois. Il existerait alors un suffixe  $v$  de  $u$  n'admettant pas  $x$  comme facteur. Mais tout facteur de  $v$  est facteur de  $u$ , donc  $P_u(|x|) > P_v(|x|)$ , ce qui contredit le fait que  $u$  et  $v$  sont sturmiens.

Ainsi, tout facteur de  $u$  apparaît une infinité de fois dans  $u$ .

**Question 8.**

- 8.a** Le caractère minimal de  $n_0$  impose :  $P_u(n_0) = n_0 + 1$ . L'application surjective  $\varphi : \mathcal{L}_{n_0+1}(u) \rightarrow \mathcal{L}_{n_0}(u)$  qui à un mot de longueur  $n + 1$  associe son préfixe de longueur  $n$  relie deux ensembles de cardinaux respectifs  $n_0 + 3$  et  $n_0 + 1$ , donc deux éléments  $v$  et  $w$  de  $\mathcal{L}_{n_0}(u)$  possèdent deux antécédents (en effet, un élément ne peut en avoir trois, puisque l'alphabet  $A$  ne comporte que deux lettres). Forcément,  $va, vb, wa, wb$  sont ces quatre antécédents.
- 8.b** Puisque  $v$  et  $w$  sont distincts, ils possèdent chacun un suffixe  $v'$  et  $w'$  de même longueur vérifiant :  $|v'|_a \neq |w'|_a$ . Sans perte de généralité, on peut supposer  $|v'|_a - |w'|_a \geq 1$ . Alors  $v'a$  et  $w'b$  sont facteurs de  $u$  et vérifient :  $|v'a|_a - |w'b|_a \geq 2$ .
- 8.c** D'après la question 1., un mot de Christoffel associé à un nombre irrationnel  $\alpha$  n'est pas ultimement périodique, et d'après la question 3.b, deux facteurs de même longueur ont, à une unité près, le même nombre de  $a$ . D'après le théorème que nous venons de démontrer, il s'agit d'un mot sturmien.

**Question 9.**

- 9.a** Posons  $v = u_i \cdots u_{i+n-1}$ ,  $w = u_j \cdots u_{j+n-1}$ , et considérons l'application :

$$f : \begin{pmatrix} \llbracket 0, n \rrbracket & \longrightarrow & \mathbb{Z} \\ k & \longmapsto & |u_i \cdots u_{i+k-1}|_a - |u_j \cdots u_{j+k-1}|_a \end{pmatrix}$$

Nous avons  $f(0) = 0$ ,  $f(n) \geq 2$  et pour tout  $k \in \llbracket 0, n-1 \rrbracket$ ,  $f(k+1) - f(k) \in \{-1, 0, 1\}$ . Il existe nécessairement  $k \in \llbracket 0, n \rrbracket$  tel que  $f(k) = 2$ . En posant  $v' = u_i \cdots u_{i+k-1}$  et  $w' = u_j \cdots u_{j+k-1}$ , on a bien  $|v'|_a - |w'|_a = 2$ .

- 9.b** Choisissons  $v'$  et  $w'$  de longueur minimale parmi les facteurs vérifiant  $|v'|_a - |w'|_a = 2$ . S'ils commençaient par la même lettre, on pourrait la supprimer, ce qui contredirait le caractère minimal de leur longueur. De même, ils ne se terminent pas par la même lettre.

Quatre cas sont à envisager :

$$(v' = av''b, w' = bw''a) \quad (v' = bv''a, w' = aw''b) \quad (v' = av''b, w' = aw''a) \quad (v' = av''a, w' = bw''b)$$

Les deux premiers cas sont à exclure car on aurait alors  $|v''|_a - |w''|_a = 2$ , ce qui contredirait le caractère minimal de la longueur de  $v'$  et  $w'$ . Le troisième cas est aussi à exclure car alors  $|v''|_a - |w''|_a = 4$ , et d'après la question 9.a on pourrait extraire de  $v''$  et  $w''$  deux facteurs de même longueur contredisant le caractère minimal de  $v'$  et  $w'$ .

On en déduit :  $v' = av''a$  et  $w' = bw''b$ , avec  $|v''|_a = |w''|_a$ .

Supposons maintenant  $v'' \neq w''$ , et notons alors  $x$  le plus long des préfixes qu'ils ont en commun : on peut écrire  $v'' = x\alpha y$  et  $w'' = x\beta y'$ , où  $\alpha$  et  $\beta$  sont deux lettres différentes.

Si on avait  $\alpha = a, \beta = b$  on aurait  $v' = axaya$  et  $w' = bxy'by$  et les facteurs  $axa$  et  $bxby$  contrediraient le caractère minimal de  $v'$  et  $w'$ . Donc  $\alpha = b, \beta = a$  et  $v'' = xby, w'' = xay'$ .

Puisque  $|v''|_a = |w''|_a$ , nécessairement  $|y|_a = |y'|_a + 1$ . Mais  $ya$  et  $y'b$  sont facteurs de même longueur de  $u$  et vérifient alors  $|ya|_a - |y'b|_a = 2$ , ce qui là encore contredit le caractère minimal de  $v'$  et  $w'$ .

Ainsi, nous avons prouvé que  $v'' = w''$  ; en notant  $z$  cette valeur commune, nous avons prouvé que  $aza$  et  $bzb$  sont facteurs de  $u$ .

- 9.c** Si  $z$  n'est pas un palindrome, on peut écrire :  $z = z_0 \cdots z_k \alpha x \beta z_k \cdots z_0$ , où  $\alpha$  et  $\beta$  sont deux lettres distinctes. Dans ces conditions,

$$v' = az_0 \cdots z_k \alpha x \beta z_k \cdots z_0 a \quad \text{et} \quad w' = bz_0 \cdots z_k \alpha x \beta z_k \cdots z_0 b.$$

Si  $\alpha = a, \beta = b$ , les mots  $az_0 \cdots z_k a$  et  $bz_0 \cdots z_k b$  sont facteurs de  $u$  et contredisent le caractère minimal de  $|z|$ ; Si  $\alpha = b, \beta = a$ , les mots  $az_k \cdots z_0 a$  et  $bz_k \cdots z_0 b$  sont facteurs de  $u$  et contredisent le caractère minimal de  $|z|$ .

On en déduit que  $z$  est un palindrome.

**9.d** Supposons maintenant que  $u$  soit un mot sturmien.

Posons  $n = |z| + 1$ . Les mots  $za$  et  $zb$  sont facteurs de  $u$ , donc d'après la question 7.a,  $z$  est le seul facteur de longueur  $n - 1$  ayant cette propriété. On en déduit que  $az$  ou  $bz$  est le seul mot de longueur  $n$  se prolongeant de deux manières à droite. Supposons par exemple que ce soit  $az$ .

Considérons le facteur  $bzb$ , et surtout les  $n - 1$  lettres suivantes, formant le mot  $z' : bzbz'$  est un facteur de  $u$  de longueur  $2n$ , il possède donc  $n + 1$  facteurs de longueur  $n$ . Nous allons montrer que parmi ces facteurs ne se trouve pas  $az$ .

Supposons que ce soit le cas. Alors on peut décomposer  $z$  et  $z'$  sous la forme :

$$z = ras = sbr' \quad \text{et} \quad z' = r't \quad \text{car} \quad bzbz' = b \underbrace{(ras)}_z b(r't)$$

Mais  $z$  est un palindrome donc s'écrit aussi  $z = \tilde{s}a\tilde{r}$ , où  $\tilde{r}$  et  $\tilde{s}$  sont les images miroir de  $r$  et  $s$ . Les deux égalités  $z = sbr'$  et  $z = \tilde{s}a\tilde{r}$  sont contradictoires.

Les  $n + 1$  facteurs de longueur  $n$  de  $bzbz'$  ne peuvent donc se prolonger que d'une seule manière à droite. De plus,  $\text{card } \mathcal{L}_n(u) \setminus \{az\} = n$ , donc deux de ces facteurs sont identiques; notons-les  $x = u_i \cdots u_{i+n-1}$  et  $y = u_j \cdots u_{j+n-1}$ , avec  $i < j \leq i + n$ .

Mais alors  $u_{j+n} = u_{i+n}, u_{j+n+1} = u_{i+n+1} \dots$  et on contredit ainsi le fait que  $u$  n'est pas ultimement périodique.

$u$  n'est donc pas un mot sturmien.

