

# Mots sturmiens et mots équilibrés

(d'après le sujet d'agrégation de mathématiques de 1994 et son corrigé)

## Définitions

Soit  $x \in \{a, b\}^{\mathbb{N}}$  un mot infini non ultimement périodique. On note  $L(x)$  l'ensemble de ses facteurs.

La *complexité* du mot  $x$  est la suite définie par  $p_n(x) = \text{Card } L(x) \cap \{a, b\}^n$  ( $n \in \mathbb{N}$ ).

Si  $u$  est un mot fini, on note  $|u|$  sa longueur et  $|u|_a$  le nombre d'occurrences de  $a$  dans  $u$ .

On dit que le mot  $x$  est *sturmien* si  $\forall n \in \mathbb{N} \quad p_n(x) = n + 1$ .

On dit que le mot  $x$  est *équilibré* si  $\forall (u, v) \in L(x)^2 \quad |u| = |v| \Rightarrow ||u|_a - |v|_a| \leq 1$ .

## Énoncé

### Équilibré $\Rightarrow$ sturmien

Soit  $x$  un mot non ultimement périodique et non sturmien. Soit  $n_0$  le plus petit nombre entier tel que  $P_{n_0+1}(x) \geq n_0 + 3$ . Montrer qu'il existe deux mots distincts  $u$  et  $v$  de longueur  $n_0$  tels que  $ua, ub, va, vb$  appartiennent à  $L(x)$ . En déduire que  $L(x)$  contient deux mots  $u'$  et  $v'$  tels que  $|u'|_a - |v'|_a \geq 2$ .

### Sturmien $\Rightarrow$ équilibré

On suppose que  $L(x)$  contient deux mots  $u$  et  $v$  de même longueur tels que  $|u|_a - |v|_a \geq 2$ .

1. Montrer que  $L(x)$  contient deux mots  $u'$  et  $v'$  de même longueur qui vérifient  $|u'|_a - |v'|_a = 2$ .
2. Montrer que dans ce cas on peut supposer, quitte à prendre des mots plus courts, qu'il existe un mot  $w$  (éventuellement vide) tel que  $u' = awa$  et  $v' = bwb$ .
3. Soit  $w$  un mot de longueur minimale satisfaisant à la condition précédente. Montrer que  $w$  est un palindrome (c'est-à-dire que, si  $w = w_0w_1 \dots w_n \in \{a, b\}^{n+1}$ , on a  $w_i = w_{n-i}$ , pour tout nombre entier  $i$  compris entre 0 et  $n$ ).
4. En déduire que le mot  $x$  n'est pas sturmien.

## Corrigé

### Équilibré $\Rightarrow$ sturmien

Remarquons d'abord que, puisque  $P_0(x) = 1$  et  $P_1(x) = 2$ , on a  $n_0 \geq 1$ . Puisqu'il y a  $n_0 + 1$  facteurs de longueur  $n_0$ , par définition de  $n_0$ , et au moins  $n_0 + 3$  facteurs de longueur  $n_0 + 1$ , et qu'un facteur de longueur  $n_0$  se prolonge d'au moins une façon et d'au plus de deux façons, il existe au moins 2 facteurs  $u$  et  $v$  de longueur  $n_0$  qui se prolongent de deux façons, c'est-à-dire deux mots  $u, v$  de longueur  $n_0$  tels que les 4 mots  $ua, ub, va, vb$  appartiennent à  $L(x)$ .

Montrons que, par définition de  $n_0$ ,  $u$  et  $v$  ont même suffixe de longueur  $n_0 - 1$ ; en effet, sinon, il y aurait deux facteurs de longueur  $n_0 - 1$  qui se prolongent de deux façons différentes, et on aurait donc  $P_{n_0-1}(x) \leq P_{n_0}(x) - 2 \leq n_0 - 1$ , ce qui est impossible puisque cela impliquerait que  $x$  est ultimement périodique.

On peut donc écrire, en échangeant au besoin  $u$  et  $v$ ,  $u = aM$  et  $V = bM$ , où  $M$  est un facteur de  $x$  de longueur  $n_0 - 1$ . En posant  $u' = ua$  et  $v' = vb$ , on voit que  $L(x)$  contient deux mots de même longueur qui vérifient  $|u'|_a - |v'|_a = 2$ .

### Sturmien $\Rightarrow$ équilibré

1. Notons  $u_i$  (resp.  $v_i$ ) le préfixe de longueur  $i$  de  $u$  (resp.  $v$ ), et posons  $n = |v|$ . Si l'on note  $d_i$  la différence  $|u_i|_a - |v_i|_a$ , il est clair que  $d_0 = 0$ , que deux termes consécutifs de la suite varient au plus de 1 ( $d_i$  est égal à  $d_{i+1}$  si  $u_{i+1}$  et  $v_{i+1}$  finissent par la même lettre, il y a une variation de 1 sinon), et que par hypothèse  $d_n \geq 2$ ; donc il existe  $k \leq n$  tel que  $d_k = 2$ ;  $u_k$  et  $v_k$  sont les deux mots  $u'$  et  $v'$  cherchés.

On a utilisé le “lemme des valeurs intermédiaires” suivant:

**Lemme**

soit  $(d_i)_{i \in \mathbb{N}}$  une suite à valeurs entières, vérifiant  $|d_{i+1} - d_i| \leq 1$ , et telle que  $d_0 = a$  et  $d_n = b \geq a$ ; alors, pour tout entier  $c$  tel que  $a \leq c \leq b$ , il existe un entier  $i$  compris entre 0 et  $n$  tel que  $d_i = c$ .

Preuve : on peut supposer  $c > a$ , sinon on prend  $i = 0$ . Soit  $A = \{i \in \{0, 1, \dots, n\} | d_j < c\}$ ;  $A$  est fini, et non vide puisqu’il contient 0. Donc  $A$  possède un plus grand élément  $j < n$ . On doit avoir par définition  $d_j < c \leq d_{j+1}$ ; puisque  $d_j$  et  $d_{j+1}$  sont deux entiers qui diffèrent au plus de 1, on a donc  $d_{j+1} = c$ .

2. On peut supposer que les mots  $u'$  et  $v'$  trouvés à la question précédente sont minimaux avec cette propriété; alors  $u'$  doit commencer par  $a$  et  $v'$  par  $b$ , car sinon, soit ils commencent par la même lettre et on peut supprimer cette première lettre, soit on peut écrire  $u' = bu''$  et  $v' = av''$ , donc  $|u''|_a - |v''|_a = 3$ , et la question précédente montre que le couple  $(u', v')$  n’est pas minimal. De même  $u'$  finit par  $a$  et  $v'$  par  $b$ . Si l’on définit la suite  $d'_i$  comme dans la question précédente, avec  $m = |u'| = |v'|$ , on voit que l’on a  $d'_1 = 1$  et  $d'_m = 2$ . S’il existe  $1 < k < m$  tel que  $d_k = 2$ , les préfixes de longueur  $k$  de  $u'$  et  $v'$  répondent encore à l’hypothèse, ce qui contredit la minimalité; de même si  $d_k = 0$ , on peut considérer les suffixes de longueur  $m - k$ . On doit donc avoir  $d_k = 1 = d_{k-1}$  si  $1 < k < m$  : ce qui prouve, comme on l’a vu à la question précédente, que  $u'$  et  $v'$  ont même lettre de rang  $k$ . Autrement dit, il existe un mot  $w$  tel que  $u' = awa$  et  $v' = bwb$ .
3. Soit  $w$  un mot de longueur minimale satisfaisant à la condition précédente. Si  $w$  est de longueur 0 ou 1, il n’y a rien à prouver. On suppose donc que  $w = w_0 \dots w_n$ , avec  $n \geq 1$ . Par hypothèse, les mots  $aw_0 \dots w_n a$  et  $bw_0 \dots w_n b$  appartiennent à  $L(x)$ . Supposons  $w_0$  et  $w_n$  distincts, par exemple  $w_0 = a$  et  $w_n = b$ ; alors  $L(x)$  contient  $aa$  et  $bb$ , et  $w$  n’est pas minimal. Supposons démontrées les égalités  $w_k = w_{n-k}$ , pour  $0 \leq k < i$ . Si  $w_i$  est différent de  $w_{n-i}$ , par exemple  $w_i = a$  et  $w_{n-i} = b$ , le langage  $L(x)$  contient les mots  $u'' = aw_0 w_1 \dots w_{i-1} a$  et  $v'' = bw_{n-i+1} \dots w_{n-1} w_n b$ . Or les mots  $w_0 w_1 \dots w_{i-1}$  et  $w_{n-i+1} \dots w_{n-1} w_n$  sont miroirs l’un de l’autre, donc ont le même nombre de  $a$  et de  $b$ ; donc  $u''$  et  $v''$  vérifient  $|u''|_a - |v''|_b = 2$ , ce qui contredit la minimalité de  $w$ , donc  $w_i = w_{n-i}$  et  $w$  est un palindrome.
4. Nous allons raisonner par l’absurde; supposons le mot  $x$  sturmien, il possède donc  $n + 2$  facteurs de longueur  $n + 1$ , qui sont  $w$  et  $n + 1$  autres mots. Comme  $awa$  et  $bwb$  sont dans  $L(x)$ , le mot  $w$  peut être prolongé à droite de deux façons, et donc les autres mots ne peuvent être prolongés à droite que d’une façon (sinon il y aurait plus de  $n + 3$  mots de longueur  $n + 2$ ), et de même  $w$  peut être prolongé à gauche de deux façons, et les autres mots d’une seule.

Au moins l’un des deux mots  $aw$  et  $bw$  peut être prolongé à droite de deux façons, sinon tout mot de longueur  $n + 2$  se prolongerait d’une seule façon (car un suffixe d’un mot ayant deux prolongements à droite a aussi deux prolongements à droite; donc  $aw$  et  $bw$  sont les seuls mots de longueur  $n + 2$  susceptibles d’avoir deux prolongements) et la suite serait ultimement périodique. Supposons que c’est  $aw$ ; alors les mots  $awa$ ,  $awb$  et  $bwb$  apparaissent dans le mot  $x$ , mais pas  $bwa$  sinon on aurait  $n + 5$  mots de longueur  $n + 3$  et  $x$  ne serait pas sturmien.

Soit  $i$  un rang d’apparition de  $bw$  dans  $x$ ; la remarque essentielle (et non triviale!) est que le mot  $aw$  n’est pas facteur de  $x_i x_{i+1} \dots x_{i+2n+3}$ . Sinon on aurait  $aw = x_{i+k} \dots x_{i+k+n+1}$ , avec  $0 \leq k \leq n + 2$ . Mais on ne peut avoir  $k = 0$ , car  $x_i = b$  par hypothèse, ni  $k = n + 2$  car  $x_{i+n+2} = b$  puisque  $bwa$  n’apparaît pas. Puisque le mot  $bwb$  apparaît au rang  $i$  et  $aw$  apparaît au rang  $i + k$ , on a  $w_{k-1} = a$  et  $w_{n+1-k} = b$  (regarder les deux façons de placer  $w$  dans ce bloc), ce qui contredit le fait que  $w$  est un palindrome.

Or il y a  $n + 3$  facteurs de longueur  $n + 2$  dans le mot  $x_i x_{i+1} \dots x_{i+2n+3}$ , de longueur  $2n + 4$ . Mais le mot infini  $x$  admet en tout  $n + 3$  facteurs de longueur  $n + 2$ , tous prolongeables à droite de façon unique, sauf  $aw$ ; on en déduit qu’un même facteur de longueur  $n + 2$  apparaît 2 fois dans  $x_i x_{i+1} \dots x_{i+2n+3}$ , ce qui entraîne que le mot est ultimement périodique : contradiction.