Reconnaissabilité de langages de numeration généralisés.

Nathalie Loraud

Nous nous intéressons à des systèmes de numération généralisés dans une base $(d_n)_n$, où $(d_{r\}})_n$ est une suite strictement croissante d'entiers de premier terme 1, appelée aussi échelle de numération. Tout entier naturel n admet un développement du type :

$$n = \sum_{i=0}^{k} n_i d_i$$

où les n_i sont des entiers positifs ou nuls. Le mot $\tilde{n} = n_k \dots n_0$ associé à cette écriture est unique si les inégalités $n_0 d_0 + \cdots + n_j d_j < d_{j+1}$, sont vérifiées pour tout $j = 0, \dots, k$; ce qui signifie que nous utilisons l'algorithme "glouton" pour écrire les entiers en base $(d_n)_n$. Nous renvoyons à [?] pour plus de détails sur ces systèmes de numération dits "standards"
L'ensemble $\mathcal{L}(d)$ de tous les mots $\tilde{n}$ est appelé le langage de la numération :

$$\mathcal{L}(d) := \left\{ n_k \dots n_0 \in \mathbb{N}^{k+1}; \forall j \leq k, n_0 d_0 + \cdots + n_j d_j < d_{j+1} \right\}$$

La question principale qui nous préoccupe est de caractériser les échelles de numération dont le langage est régulier, *i.e.* reconnaissable par un automate fini ; c'est une question posée par J. Shallit dans [8]. Nous rappelons, à cet égard, les définitions d'automate (fini) sur un alphabet A, de langage sur A et de reconnaissabilité :

Un *automate fini* sur A est un quadruplet $\mathcal{A} = (S, \Phi, I, T)$ où S est un ensemble fini appelé ensemble des états, $\Phi = (\phi_a)_{a \in A}$ est la famille des flèches de l'automate, représentée par un graphe, I est un état distingué de S appelé état initial et $T \subset S$ est l'ensemble des états terminaux. Un *langage* sur A est un sous ensemble $\mathcal{L}$ du monoide libre A^* engendré par A. Le langage $\mathcal{L}$ est reconnu par un automate $\mathcal{A}$ si l'ensemble des mots de $\mathcal{L}$ coïncide avec l'ensemble des mots obtenus comme inversion de la concaténation des étiquettes d'un chemin dans l'automate, partant de I et aboutissant à un état terminal, ce qui veut dire que le mot $n_k \dots n_0$ est dans le langage L si et seulement si l'on peut définir successivement les états $A_1 = \Phi_{n_0}(A_0), \dots, A_k = \Phi_{n_k}(A_{k-1})$. On dira que $\mathcal{L}$ est *reconnaissable par automate* ou *régulier* s'il existe un automate fini $\mathcal{A}$ qui le reconnait.

Nous allons étudier des classes particulières de suites, et déterminer dans chaque classe, celles qui sont base d'un système de numération dont le langage est reconnaissable par automate.

J. Shallit a démontré qu'une condition nécessaire est que la suite $(d_n)_n$ soit récurrente linéaire à coefficients dans $\mathbb{Z}$, mais que ce n'est pas une condition suffisante.

Nous obtenons les résultats suivants

1. Suites arithmético-géométriques

Théorème 1 *Soit* $(d_n)_n$ *une suite d'entiers naturels arithmético-géométrique de coefficients* a *et* b *dans* $\mathbb{Z}$(*i.e.* $\forall n \geq 0,\ d_{n+1} = ad_n + b$) *et de premier terme* $d_0 = 1.(d_n)_n$ *est la base d'un système de numération dont le langage est régulier si et seulement si* ($a > 1$ et $b \geq 0$) *ou* ($a = 1$ et $b \geq 1$) .

Ce corollaire permet de voir 1'importance du choix de d_1 pour l'obtention d'un langage régulier (la récurrence étant fixée), ce qui répond à une question de G. Hansel.

Corollaire 1 *Soit* $(d_n)_n$ *une suite d'entiers naturels vérifiant une relation de récurrence d'ordre 2 du type* : $\forall n \geq 1, d_{n+1} = (a+1)d_n - ad_{n-1}$, *où* $a \in \mathbb{Z}$, *et de premier terme* $d_0 = 1$. *Alors* $\mathcal{L}(d)$ *est régutier si et seulement si* $(1 < a \leq d_1)$ *ou* ($a = 1$ *et* $d_1 \geq 2$) .

2. Bases de Cantor

Soit $(q_n)_n$ une suite d'entiers telle que : $q_0 = 1$ et $q_n \geq 2$, pour tout $n \geq 1$. Soit $(d_n)_n$ la suite définie par : $d_n = q_0 \dots q_n$, $n \geq$ O. C'est la base d'un système de numération appelé *système de Cantor associé à la suite* $(q_n)_n$ et le langage de cette numération est donné par le lemme classique suivant :

Lemme 1 $\mathcal{L}(d) = \left\{ n_{k \cdots n_0} \in \mathbb{N}^{k+1}; \forall j \leq k, n_j < q_{j+1} \right\}$.

Dans ce qui suit, on caractérise les bases de Cantor qui sont aussi récurrentes linéaires à coefficients dans $\mathbb{Z}$.

Théorème 2 *Soit* $(d_n)_n$ *la base du système de Cantor associé à la suite d'entiers naturels* $(q_n)_n$. *Les assertions suivantes sont équivalentes* :

i) la suite $(d_n)_n$ *est récurrente linéaire à coefficients dans* $\mathbb{Z}$;

ii) la suite $(q_n)_n$ *est ultimement périodique*;

iii) la série $\Sigma_{n=0}^{+\infty} d_n X^n$ *est* $\mathbb{N}$*-rationnelle ;*

i) le langage $\mathcal{L}(d)$ *est reconnaissable par automate.*

Exemples : quand on développe les entiers en base k, on obtient un langage reconnaissable car dans ce cas $d_{n+1} = kd_n$ i.e $(q_n)_n$ est la suite constante $(k)_n$.
La représentation factorielle ne fournit pas, quant à elle, un langage régulier;
en effet, $d_{n+1} = (n+1)d_n$ et $(q_n)_n = (n+1)_n$ n'est pas une suite ultimement périodique.

3. Systèmes de numeration d'Ostrowski

Soit un nombre irrationnel $\alpha > 1$. On note $\Delta(\alpha) = [a_0; a_1, a_2, \ldots]$ son développement en fraction continue.

Pour tout $n \geq 0, \dfrac{p_n(\alpha)}{q_n(\alpha)}$ désigne le n$^{\text{ieme}}$ convergent de α : $\dfrac{p_n}{q_n} = [a_0; a_1, \ldots, a_n]$ et l'on a pour les deux suites $(q_n)_{n\geq 0}$ et $(p_n)_{n\geq 0}$ les relations de récurrence classiques :

$$\begin{array}{c} q_0 = 1; q_1 = a_1 \text{ et } \forall n \geq 2, q_n = a_n q_{n-1} + q_{n-2}; \\ p_0 = a_0; p_1 = a_1 a_0 + 1 \text{ et } \forall n \geq 2, p_n = a_n p_{n-1} + p_{n-2} \end{array}.$$

Si $\alpha > 1$, la suite $Q_\alpha = (q_n(\alpha))_n$ est strictement croissante, de premier terme 1, pourvu que $a_1 > 1$; sinon on considère la suite $Q_\alpha = (q_{n+1}(\alpha))_n$.Pour le système de numération de base Q_α le langage associé est donné par le lemme suivant :

Lemme 2 *([7]) Si $\alpha > 1$ est un nombre réel de développement $\Delta(\alpha) = [a_0; a_1, a_2, \ldots]$, alors*

$$\mathcal{L}(Q_\alpha) = \{n_k \ldots n_0; \forall j \leq k,\ (0 \leq n_{\mathrm{j}} < a_{j+1}) \textit{ ou } (n_j = a_{\mathrm{i}+1} \textit{ et } n_{j-1} = 0)\}.$$

Le système de numération de base Q_α est appelé *système de numération d'Ostrowski relatif au nombre réel* α, [7].
Le problème de la reconnaissabilité est totalement résolu pour ces langages. J. Shallit a obtenu une caractérisation qu'il est possible de retrouver de façon simple. Il en résulte que, ici aussi, pour qu'une échelle de numérotation donne un langage régulier, il suffit qu'elle soit récurrente linéaire à coefficients dans $\mathbb{Z}$:

Théorème 3 *Soit Q_α la suite des dénominateurs des convergents d'un nombre réel* $\alpha > 1$. *Les assertions suivantes sont équivalentes* :
i) $\mathcal{L}(Q_\alpha)$ est reconnaissable par automate;
ii) $\Delta(\alpha)$ est ultimement périodique;
iii) α est un nombre quadratique;
iv) Q_α est récurrente linéaire à coefficients dans $\mathbb{Z}$.

Bibliographie

[1] *A. Bertrand*, Comment écrire les nombres entiers dans une base qui n'est pas entière. à paraître dans Acta. Math. Acad. Sci. Hungar.

[2] *A. S. Fraenkel*, Systems of numeration. Amer. Math. Monthly **92** (1985), 105-114.

[3] *C. Frougny*, Representations of numbers and finite automata. Math. Syst. Theory **25** (1992), 37-60.

[4] *C. Frougny*, Linear Numeration Systems of Order Two. Inform. & Comput. 77 (1988) 233-259.

[5] *H. W. Lenstra, J. Shallit*, Continued fractions and linear recurrences. Math. Comp. **61** (1993), 351-354.

[6] *N. Loraud* β-shift, systhmes de numération et automates. J.T.N. B **7** (1995) 473-498

[7] *A. Ostrowski* : Bemerkungen zur Theorie der Diophantishen Approximationene. Abh. Math. Sem. Hamburg **1** (1922), 77-98.

[8] *J. Shallit2* Numeration Systems, Linear Recurrences, and Regular Sets. Inform. and comput. to appear

CMI Université de Provence
39, rue Joliot Curie, 13453 Marseille FRANCE
loraud@gyptis.univ-mrs.fr