Multiplication by a Constant and Recognizability in an Abstract Numeration System

Emilie CHARLIER
joint work with Michel RIGO
University of Liège

echarlier@ulg.ac.be

Introduction

One of the main interesting concerns in work about numeration systems is studying the relationships that exist between arithmetic properties of numbers and syntactical properties of their representation. More precisely, we prefer to manipulate set of numbers whose representations are characterized by very simple syntactical rules, that is, forming a regular language. Such sets are called *S-recognizable sets*, where S is the numeration system we are working with. In particular, we would like to obtain numeration systems in which the whole set $\mathbb{N}$ is recognizable, partly due to the fact that, if the set of representations of all integers is regular, there are very simple algorithms making it possible to decide if a word stands or does not stand for a number.

Some general questions in this area are the following.

- Characterization of the recognizable parts in a fixed numeration system.
- Determination of the numeration systems for which a given set of numbers is recognizable.
- To examine the stability of recognizability under arithmetic operations.

In this poster we present our research about that last problem. First, we will define the abstract numeration systems based on a regular language. Then we will study the relationship between multiplication by a constant and the S-recognizability for languages of the type $a_1^* \dots a_\ell^*$.

Numeration Systems

Definition 1. A *numeration system* is simply a bijection from $\mathbb{N}$ onto a language, called the *numeration language.* In other words, it is a way of representing numbers by words.

Example 2. The entire base-k systems and the Fibonacci system are particular cases of the positional numeration systems which are defined as a strictly increasing sequence $U = (U_n)_{n\in\mathbb{N}}$ of integers such that $U_0 = 1$ and the ratio $\frac{U_{n+1}}{U_n}$ is bounded. The representation bijection is noted r_U. Such numeration systems are based on the greedy algorithm.

Numeration Systems on a Regular Language

As we prefer to work with regular numeration languages, the idea of P. Lecomte and M. Rigo was to define new numeration systems that would be based on a regular language. More clearly, instead of searching for a classical numeration system with a regular numeration language, the idea was to start immediately with a regular language and then construct the numeration system on this language.

L regular $\longrightarrow$ ABSTRACT NUMERATION SYSTEM

Such a construction is possible thanks to the following remark. If the greedy algorithm is useful for defining a numeration system, it is not necessary. Indeed, we have the following result.

Proposition 3. *Let L be the numeration language associated with the numeration system based on a strictly increasing sequence $U = (U_n)_{n\in\mathbb{N}}$ of integers such that $U_0 = 1$ and consider the genealogical order on L. Then*

$$r_U : \mathbb{N} \to L$$

is a strictly increasing function.

This proposition shows that positional numeration systems are entirely characterized by the numeration language and the order on the alphabet.

The following definition was introduced by P. Lecomte and M. Rigo in [1].

Definition 4. An *abstract numeration system* or *numeration system on a regular language* is a triple $S = (L, \Sigma, <)$ where L is an infinite language over the totally ordered alphabet $(\Sigma, <)$. Enumerating the elements of L genealogically with respect to $<$ leads to a one-to-one map r_S from $\mathbb{N}$ onto L. To any natural number n, it assigns the $(n+1)$th word of L, its *S-representation,* while the inverse map val_S sends any word belonging to L onto its numerical value.

Definition 5. A subset X is said to be *S-recognizable* if $r_S(X)$ is a regular subset of L.

Multiplication by a Constant and S-recognizability

Having those generalized numeration systems at our disposal, we can consider the effect of addition on S-recognizable sets. If addition preserves S-recognizability then multiplication by 2 also preserves the S-recognizability. So, a natural question about the stability of S-recognizability arises.

Q : When does the multiplication by an integer λ preserve the S-recognizability?

$S = (a_1^* \dots a_\ell^*, \{a_1, \dots, a_\ell\}, a_1 < \dots < a_\ell)$

Let us define some useful notations. We denote by $\mathcal{B}_\ell = a_1^* \cdots a_\ell^*$ the bounded language over the alphabet $\Sigma_\ell = \{a_1, \dots, a_\ell\}$ of size ℓ. We set

$$\mathbf{u}_\ell(n) = \#(\mathcal{B}_\ell \cap \Sigma_\ell^n) \quad \text{and} \quad \mathbf{v}_\ell(n) = \#(\mathcal{B}_\ell \cap \Sigma_\ell^{\le n}) = \sum_{i=0}^{n} \mathbf{u}_\ell(i).$$

In what follows, we assume that $(\Sigma_\ell, <)$ is totally ordered by $a_1 < \cdots < a_\ell$. So we can enumerate the words of $\mathcal{B}_\ell$ by the increasing genealogical ordering induced by the ordering of Σ_ℓ. Let $n \ge 0$, the n-th word of $\mathcal{B}_\ell$ is denoted $\mathrm{rep}_\ell(n)$. The reciprocal map $\mathrm{rep}_\ell^{-1} =: \mathrm{val}_\ell$ maps the n-th word of $\mathcal{B}_\ell$ onto n. We recall that the binomial coefficient $\binom{i}{j}$ vanishes if $i < j$.

Remark 6. We have

$$\mathbf{u}_{\ell+1}(n) = \mathbf{v}_\ell(n),\ \forall \ell \ge 1,\ \forall n \ge 0. \tag{1}$$

From this observation, it follows that

$$\mathbf{u}_\ell(n) = \binom{n+\ell-1}{\ell-1},\ \forall \ell \ge 1,\ \forall n \ge 0. \tag{2}$$

Lemma 7. *Let $S = (a_1^* \cdots a_\ell^*,\ \Sigma_\ell,\ \{a_1 < \cdots < a_\ell\})$. We have*

$$\mathrm{val}_\ell(a_1^{n_1} \cdots a_\ell^{n_\ell}) = \sum_{i=1}^{\ell} \binom{n_i + \cdots + n_\ell + \ell - i}{\ell - i + 1}.$$

Example 8. Consider the words of length 3 in the language $a^*b^*c^*$,

$$aaa < aab < aac < abb < abc < acc < bbb < bbc < bcc < ccc.$$

We have $\mathrm{val}_3(aaa) = \binom{5}{3} = 10$ and $\mathrm{val}_3(acc) = 15$. If we apply the erasing morphism $\varphi : \{a, b, c\} \to \{a, b, c\}^*$ defined by $\varphi(a) = \varepsilon$, $\varphi(b) = b$ and $\varphi(c) = c$ on the words of length 3, we get

$$\varepsilon < b < c < bb < bc < cc < bbb < bbc < bcc < ccc.$$

So the ordered list of words of length 3 in $a^*b^*c^*$ contains an ordered copy of the words of length at most 2 in the language b^*c^* and to obtain $\mathrm{val}_3(acc)$, we just add to $\mathrm{val}_3(aaa)$ the position of the word cc in the ordered language b^*c^*.

Corollary 9. *Let $\ell \in \mathbb{N} \setminus \{0\}$. Any integer n can be uniquely written as*

$$n = \binom{z_\ell}{\ell} + \binom{z_{\ell-1}}{\ell-1} + \cdots + \binom{z_1}{1}$$

with $z_\ell > z_{\ell-1} > \cdots > z_1 \ge 0$.

Example 10. For $\ell = 3$, one gets for instance

$$278 = \binom{12}{3} + \binom{11}{2} + \binom{3}{1}$$

and solving the system

$$\left.\begin{aligned} n_1 + n_2 + n_3 &= 12 - 2 \\ n_2 + n_3 &= 11 - 1 \\ n_3 &= 3 \end{aligned}\right\} \Leftrightarrow (n_1, n_2, n_3) = (0, 7, 3),$$

we have $\mathrm{rep}_3(278) = b^7c^3$. In the same way,

$$12345678901234567890 = \binom{4199737}{3} + \binom{3803913}{2} + \binom{1580642}{1}$$

and

$$\mathrm{rep}_3(12345678901234567890) = a^{395823}b^{2223270}c^{1580642}.$$

The Case of $S = (a^*b^*, \{a, b\}, a < b)$

In their paper [1], P. Lecomte and M. Rigo have studied that question Q on the particular abstract numeration system $S = (a^*b^*, \{a, b\}, a < b)$. Here follows one of the main results of that paper.

Theorem 11. *Let S be the system $(a^*b^*, \{a, b\}, a < b)$ and $\alpha \in \mathbb{N}$. Multiplication by a α transforms the S-recognizable sets into S-recognizable sets if and only if α is an odd perfect square.*

Multiplication for Polynomial Languages

Another interesting result of [1] concerns the relationship between the complexity of the numeration language, multiplication by a constant and S-recognizability.

Theorem 12. *Let $L \subset \Sigma^*$ be a regular language such that its complexity function $\mathbf{u_n}(\mathrm{L})$ is $\Theta(n^l)$ for some integer l. If $\lambda \in \mathbb{N} \setminus \{n^{l+1} : n \in \mathbb{N}\}$, then there exists a subset X of $\mathbb{N}$ such that $rep_S(X)$ is regular and that $rep_S(\lambda X)$ is not. In other words, multiplication by a constant λ conserves recognizability only if λ is of the form n^{l+1}, for some $n \in \mathbb{N}$.*

The Case of $S = (a^*b^*c^*, \{a, b, c\}, a < b < c)$

Thanks to the theorem 12, we already know that only the multiplication by constants of the form β^3, with β being an integer, can preserve the S-recognizability. The following result shows that the condition on the constant is even more restrictive.

Theorem 13. *For the abstract numeration system $S = (a^*b^*c^*, \{a < b < c\})$, if $\beta \in \mathbb{N} \setminus \{0, 1\}$ is such that $\beta \not\equiv \pm 1 \pmod 6$ then the multiplication by β^3 does not preserve the S-recognizability.*

Unfortunately, even in the "simple" case of a three-letter alphabet, we still do not know if that condition is sufficient or not.

Conjecture for the General Case

Thanks to computer experiments, we conjecture the following result.

Theorem 14. *Multiplication by β^ℓ preserves S-recognizability for the abstract numeration system $S = (a_1^* \cdots a_\ell^*, \{a_1 < \cdots < a_\ell\})$ built on the bounded language $\mathcal{B}_\ell$ over ℓ letters if and only if*

$$\beta = \prod_{i=1}^{k} p_i^{\theta_i}$$

where $p_1, \dots, p_k$ are prime numbers strictly greater than ℓ.

References

[1] P.B.A. Lecomte, M. Rigo, Numeration Systems on a Regular Language, *Theory Comput. Syst.* **34** (2001), 27-44.

Presented at the first BMS' Ph.D. day on 12th september 2005