A note on syndeticity, recognizable sets and Cobham’s theorem

Michel Rigo, Laurent Waxweiler

Abstract

In this note, we give an alternative proof of the following result. Let $p, q \geq 2$ be two multiplicatively independent integers. If an infinite set of integers is both $p$- and $q$-recognizable, then it is syndetic. Notice that this result is needed in the classical proof of the celebrated Cobham’s theorem. Therefore the aim of this paper is to complete [13] and [1] to obtain an accessible proof of Cobham’s theorem.

1 Introduction

Cobham’s theorem is related to numeration systems and can be considered as a classical result in formal languages theory. It is formulated as follows. Let $p, q \geq 2$ be two multiplicatively independent integers (i.e., the only integers satisfying $p^k = q^\ell$ are $k = \ell = 0$). If a subset $X \subseteq \mathbb{N}$ of integers is both $p$- and $q$-recognizable then it is a finite union of arithmetic progressions (i.e., $X$ is an ultimately periodic set). Recall that $X \subset \mathbb{N}$ is said to be $p$-recognizable if the language $\rho_p(X)$ of the $p$-ary representations (without leading zeroes) of the elements in $X$ is a regular language accepted by a finite automaton (see for instance [7, Chap. 5]).

This famous result has been widely studied from various points of view (we give here just a few references): extension to non-standard numeration systems [6, 10] or to the framework of $k$-regular sequences [2], study of the multidimensional case (known as Cobham-Semenov’s theorem) [4, 14], alternative proofs using the formalism of the first order logic [3, 12], . . . .

The original proof due to Cobham is widely considered as rather difficult [5]. In his book, S. Eilenberg proposed as a challenge to find an easier proof [7]. The major improvements in the simplification of the proof of Cobham’s theorem were made by G. Hansel in [8] where he makes use of the notion of syndeticity and
sketches the key-points leading to the result. Recall that an infinite set of integers \( X = \{x_0 < x_1 < \cdots \} \) is said to be \textit{syndetic} if there exists \( C > 0 \) such that for all \( n \geq 1, x_n - x_{n-1} \leq C \). (Notice that Hansel’s ideas about sydenticity also hold in a wider framework than \( p \)-ary numeration systems \([9]\).)

Afterwards, a great work of presentation relying on the main ideas found in \([8]\) was made by several authors \([1, 13]\). Unfortunately, in these last two documents a same mistake can be found (Statement \([1]\) below is not correct and Example \([2]\) is a counter-example). In this note, our modest contribution is to correct this error using as simple arguments as possible. In the spirit, we are naturally close to \([5]\) and \([8]\) but new ideas appear in our reasoning. Finally, we hope that this erratum added to \([13]\) or \([1]\) will now give a complete presentation of the proof of Cobham’s theorem.

Let us set \( \Sigma_p := \{0, \ldots, p-1\} \) as the alphabet of the \( p \)-ary digits. In \([1, 13]\), the following result is presented.

\textbf{Statement 1.} \textit{If an infinite \( p \)-recognizable set \( X \subseteq \mathbb{N} \) is such that \( 0^* \rho_p(X) \) is right dense, i.e., for all \( u \in \Sigma_p^* \) there exists \( v \in \Sigma_p^* \) such that \( uv \in 0^* \rho_p(X) \), then \( X \) is syndetic.}

\textbf{Example 2.} As stated above, Statement \([7]\) is not correct. An easy counter-example is given by the following set \( X \) of integers

\[ X = \bigcup_{i \geq 0} [2^{2i}, 2^{2i+1}]. \]

Indeed, this set is \( 2 \)-recognizable : \( \rho_2(X) = 1\{00, 01, 10, 11\}^* \), and trivially right dense but not syndetic.

In the literature, Statement \([1]\) is generally presented to obtain the following proposition.

\textbf{Proposition 3.} \([8]\) \textit{Prop. 5} \textit{Let} \( p, q \geq 2 \) \textit{be two multiplicatively independent integers. If an infinite set of integers if both \( p \)- and \( q \)-recognizable, then it is syndetic.}

In substance, this latter result can naturally be found in Cobham’s work (see \([5]\) Lemma 3). In this note, our aim is to give an alternative proof of Proposition \([3]\) not using Statement \([1]\). Our approach relies on five easy lemmas.

\section{Proof of the result}

We assume that the reader has some basic knowledge in automata theory (see for instance \([7]\)). If \( X \subseteq \mathbb{N} \) is a set of integers, we define a mapping (or a right-infinite
word) $1_X : \mathbb{N} \to \{0, 1\}$ such that $1_X(n) = 1$ if and only if $n \in X$. If $w$ is a finite word, $|w|$ denotes its length.

This first lemma will be useful in the proof of Lemma 6 and 7.

**Lemma 4.** Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA (Deterministic Finite Automaton) with $\delta : Q \times \Sigma^* \to Q$ as transition function. For any state $s \in Q$, the set

$$L_s := \{ |w| \in \mathbb{N} : w \in \Sigma^*, \delta(s, w) \in F \}$$

is such that $1_{L_s}$ is ultimately periodic, i.e., there exist $N \geq 0$ and $P > 0$ such that for all $n \geq N$, $1_{L_s}(n) = 1_{L_s}(n + P)$.

**Proof.** For any state $s \in Q$, we define a mapping

$$f_s : \mathbb{N} \to \mathcal{P}(Q) : n \mapsto \{\delta(s, w) : w \in \Sigma^n\}.$$

Since $\mathcal{P}(Q)$ is finite, there exist $a_s$ and $b_s$ such that $a_s < b_s$ and $f_s(a_s) = f_s(b_s)$. Obviously, for any $u, v \in \Sigma^*$, $\delta(s, uv) = \delta(s, u)v$. Consequently for all $n \geq 0$,

$$f_s(a_s + n) = \bigcup_{r \in f_s(a_s)} f_s(n) = \bigcup_{r \in f_s(b_s)} f_s(n) = f_s(b_s + n).$$

In other words, $f_s$ is ultimately periodic: $f_s(n) = f_s(n + b_s - a_s)$ if $n \geq a_s$. To conclude the proof, observe that $1_{L_s} = 1_{F_s}$ where $F_s = \{n \in \mathbb{N} : f_s(n) \cap F \neq \emptyset\}$. □

**Lemma 5.** Let $m, n, a, b, c, d \in \mathbb{N} \setminus \{0\}$ be arbitrary integers such that $n < m$ and $p, q$ be two multiplicatively independent integers. Then there exist integers $k, \ell \geq 1$ such that $nq^{c+\ell} \leq mp^{a+bk} < (m+1)p^{a+bk} \leq (n+1)q^{c+\ell}$.

**Proof.** It is enough to find integers $k, \ell$ satisfying

$$\frac{nq^c}{mp^a} \leq \left(\frac{p^b}{q^d}\right)^k \leq \frac{(n+1)q^c}{(m+1)p^a}.$$  

This is a direct consequence of Kronecker’s theorem (because $p^b$ and $q^d$ are still multiplicatively independent hence $\log p^b / \log q^d$ is irrational) [11]. □

**Lemma 6.** Let $p \geq 2$ and $X \subseteq \mathbb{N}$ be an infinite $p$-recognizable set. Then there exist integers $m, a, b \geq 1$ such that for all $k \in \mathbb{N}$, the set $X \cap \{mp^{a+bk}, (m+1)p^{a+bk}\}$ is nonempty. Moreover, the integer $m$ can be chosen arbitrarily large.

**Proof.** Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA recognizing $\rho_p(X)$. Since $X$ is infinite, there exists $m > 0$ arbitrarily large such that $\rho_p(m)$ is prefix of an infinite number of elements in $\rho_p(X)$. Let $s = \delta(q_0, \rho_p(m))$. By Lemma 4 there exist $\alpha \geq 0$ and $b > 0$ such that $1_{L_s}(n) = 1_{L_s}(n + b)$ for all $n \geq \alpha$.  

3
For any \( t \geq 0 \), the interval \( [mp', (m+1)p'] \) contains all the integers having a \( p \)-ary representation of the form \( \rho_p(m)w \) with \( |w| = t \). Since the set \( (\rho_p(m)\Sigma_p)^r \cap \rho_p(X) \) is infinite, there exists a word \( v \) such that \( \rho_p(m)v \) is the \( p \)-ary representation of an element in \( X \) with \( |v| > \alpha \). Take \( a = |v| \). Consequently, the interval \( [mp^a, (m+1)p^a] \) contains an element belonging to \( X \). The conclusion follows from the periodicity of \( 1_{L_s} \): \( 1_{L_s}(a) = 1_{L_s}(a + kb) = 1 \), for all \( k \geq 0 \).

Recall that a state \( s \) is said to be accessible (resp. coaccessible) if there exists a word \( w \) such that \( \delta(q_0, w) = s \) (resp. \( \delta(s, w) \in F \)). The trimmed minimal automaton of a language \( L \) is obtained by taking only states which are accessible and coaccessible.

**Lemma 7.** Let \( p \geq 2 \) and \( X \subseteq \mathbb{N} \) be an infinite \( p \)-recognizable set such that \( \mathcal{A} = (Q, q_0, F, \Sigma_p, \delta) \) is the trimmed minimal automaton of \( \rho_p(X) \). If there exists a state \( s \) such that \( \mathbb{N} \setminus L_s \) is infinite, then there exist integers \( m, a, b \geq 1 \) such that for all \( k \in \mathbb{N} \), the set \( X \cap [mp^{a+bk}, (m+1)p^{a+bk}] \) is empty.

**Proof.** Let \( s \) be a state such that \( \mathbb{N} \setminus L_s \) is infinite. Without loss of generality, we may assume that \( s \neq q_0 \) and there exists \( m > 0 \) such that \( \delta(q_0, \rho_p(m)) = s \). (Indeed, if \( \mathbb{N} \setminus L_{q_0} \) is infinite then the same property holds for some other state \( s \).)

We use the same reasoning as in the previous proof. Thanks to Lemma 4 there exist \( \alpha \geq 0 \) and \( b > 0 \) such that \( 1_{L_s}(n) = 1_{L_s}(n + b) \) for all \( n \geq \alpha \). Since \( \mathbb{N} \setminus L_s \) is infinite, there exists \( a > \alpha \) such that no word \( v \) of length \( a \) is such that \( \delta(s, v) \in F \).

In other words, if \( |v| = a \) then \( \rho_p(m)v \notin \rho_p(X) \) and the interval \( [mp^a, (m+1)p^a] \) does not contain any element of \( X \). Once again, the conclusion follows from the periodicity of \( 1_{L_s} \). \( \square \)

The last lemma is a simple consequence of the three previous ones.

**Lemma 8.** Let \( q > p \geq 2 \) be two multiplicatively independent integers and \( X \subseteq \mathbb{N} \) be an infinite \( p \)- and \( q \)-recognizable set of integers. If \( \mathcal{A} = (Q, q_0, F, \Sigma_p, \delta) \) is trimmed minimal automaton of \( \rho_q(X) \), then for any state \( r \in Q \), the set \( L_r \) is cofinite.

**Proof.** Assume to the contrary that \( \mathbb{N} \setminus L_r \) is infinite. By Lemma 7 there exist \( n, c, d \geq 1 \) such that for all \( \ell \in \mathbb{N} \), \( X \cap [nq^{c+d\ell}, (n+1)q^{c+d\ell}] \) is empty.

By Lemma 6 there also exist \( m, a, b \geq 1 \) such that for all \( k \in \mathbb{N} \), \( X \cap [mp^{a+bk}, (m+1)p^{a+bk}] \) is nonempty and \( m > n \).

To obtain a contradiction, simply observe that as a consequence of Lemma 5 there exist \( K, L \geq 1 \) such that \( nq^{c+dL} \leq mp^{a+bk} < (m+1)p^{a+bk} \leq (n+1)q^{c+dL} \). \( \square \)

We now have at our disposal all the necessary material to conclude this short note.
Proof of Proposition Assume that \( q > p \). Let \( \mathcal{A} = (Q, q_0, F, \Sigma, \delta) \) be the trimmed minimal automaton of \( \rho_q(X) \). For all \( n > 0 \), we write \( q_n := \delta(q_0, \rho_q(n)) \). Thanks to Lemma, \( L_{q_n} \) is cofinite. This means that for all \( n \geq 0 \), there exists \( C_n \) such that for all \( k \geq C_n \), \( k \) belongs to \( L_{q_n} \). Clearly, \( C_n \) depends only on the state \( q_n \) and there are a finite number of such states. Let \( C = \max\{C_n\} \). Consequently, for any \( n > 0 \), there exists a word \( w_n \) of length \( C \) such that \( \rho_q(n)w_n \in \rho_q(X) \). In other words, for any \( n > 0 \), there exist \( t_n \in [0, q^C] \) such that \( nq^C + t_n \in X \). We conclude that any interval of length \( 2q^C \) contains at least an element belonging to \( X \). 

References

[1] J.-P. Allouche, J. Shallit, *Automatic sequences, Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, (2004).

[2] J. P. Bell, A generalization of Cobham’s theorem for regular sequences, Preprint (2005).

[3] A. Bès, An extension of the Cobham-Semenov theorem, *J. Symbolic Logic* 65 (2000), 201–211.

[4] V. Bruyère, G. Hansel, C. Michaux, R. Villemaire, Logic and \( p \)-recognizable sets of integers, *Bull. Belg. Math. Soc.* 1 (1994), 191–238.

[5] A. Cobham, On the base dependence of sets of numbers recognizable by finite automata, *Math. Syst. Theory* 3 (1969), 186–192.

[6] F. Durand, A theorem of Cobham for non primitive substitutions, *Acta Arith.* 104 (2002), 225–241.

[7] S. Eilenberg, *Automata, Languages and Machines*, vol. A, Academic Press (1974).

[8] G. Hansel, À propos d’un théorème de Cobham, In D. Perrin Ed., *Actes de la fête des mots*, 55–59, Greco de programmation, CNRS, Rouen, (1982).

[9] G. Hansel, Systèmes de numération indépendants et syndéticité, *Theoret. Comput. Sci.* 204 (1998), 119–130.

[10] G. Hansel, T. Safer, Vers un théorème de Cobham pour les entiers de Gauss, *Bull. Belg. Math. Soc. Simon Stevin* 10 (2003), 723–735.

[11] G.H. Hardy, E.M. Wright, *Introduction to the Theory of Numbers*, Oxford Univ. Press, (1985).

[12] C. Michaux, R. Villemaire, Presburger arithmetic and recognizability of sets of natural numbers by automata: new proofs of Cobham’s and Semenov’s theorems, *Ann. Pure Appl. Logic* 77 (1996), 251–277.

[13] D. Perrin, Finite Automata, J. Van Leeuwen Ed. *Handbook of Theoret. Comput. Sci.*, vol. B, 1–57, Elsevier–MIT Press, (1990).

[14] F. Point and V. Bruyère, On the Cobham-Semenov theorem, *Theory Comput. Syst.* 30 (1997), 197–220.