Tomi Kärki

A Note on the Proof of Cobham's Theorem

Turku Centre for Computer Science

TUCS Technical Report
No 713, September 2005

A Note on the Proof of Cobham's Theorem

Tomi Kärki
Department of Mathematics
University of Turku
FIN-20014 Turku, Finland
topeka@utu.fi

TUCS Technical Report
No 713, September 2005

Abstract

The famous theorem of Cobham says that, for multiplicatively independent integers k and l, any subset of $\mathbb{N}$, which is both k- and l-recognizable, is recognizable. Many of its proofs are based on so called Hansel's lemma stating that such a k- and l-recognizable set is syndetic. We consider these proofs and point out that some of them are inadequate.

Keywords: recognizability, Cobham's theorem, Hansel's lemma, multiplicative independence, right dense set, syndetic set

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Definitions

Let Σ be an *alphabet* and a *word* w over Σ is a sequence of symbols from Σ. The *length* of w, denoted by $|w|$, is the total number of letters in w. Denote by Σ^* and Σ^+ the sets of all finite and finite nonempty words over Σ, respectively. Any subset of Σ^* is called a *language* on the alphabet Σ. Let $k \geq 2$ be an integer. Define $\Sigma_k = \{0, 1, \dots, k-1\}$ and, for any $n \in \mathbb{N}$, let $(n)_k \in \Sigma_k^+$ be the *normalized* base k representation of n. Normalized representation means a word with no extra leading zeros, i.e., a word belonging to the set $\Sigma^+ \setminus 0\Sigma^+$. A language $L \subseteq \Sigma$ is Σ-*automatic* if there exists a deterministic finite automaton with input alphabet Σ accepting the language L. A Σ_k-automatic set is shortly called k-*automatic*. A subset X of $\mathbb{N}$ is k-*recognizable* if the language $(X)_k = \{(x)_k \,|\, x \in X\}$ is k-automatic. A subset X is called *recognizable* if it is k-recognizable for all bases $k \geq 2$. There are many equivalent formulations of recognizability. For example, a set $L \subseteq \mathbb{N}$ is recognizable if and only if it is a finite union of arithmetic progressions; if and only if L is ultimately periodic; if and only if L is 1-*recognizable* ,i.e., $\{a^n \,|\, n \in L\}$ is $\{a\}$-recognizable. Integers k and l are *multiplicatively independent* if, for all nonzero integers p and q, we have $k^p \neq l^q$. A set $X = \{x_1 < x_2 < \cdots < x_n <\} \subseteq \mathbb{N}$ is *syndetic* if $\sup(x_{n+1} - x_n) < \infty$.

2 On the proofs of Cobhams's theorem

Cobham proved the following theorem in his original paper [3].

Theorem 1. *Let $k, l \geq 2$ be multiplicatively independent integers. If a set $X \subseteq \mathbb{N}$ is both k- and l-recognizable, then it is recognizable.*

In this paper Cobham deeply analyzed some properties of an automaton recognizing X. His proof is quite complicated and therefore Eilenberg announced in [4] that it is a challenge to find a simpler and shorter proof. This was obtained by Hansel in 1982 [5]. His proof consisted of two parts. First he showed that if X is both k- and l recognizable set, then it is syndetic. In the second part his idea was to show that the characteristic sequence of X is ultimately periodic by showing that there exists m such that the number of recurrent subwords of length m is bounded by m. A more precise version of this proof was made by Reutenauer [10] completing the outlines of Hansel's proof. In Reutenauer's paper the first part of Hansel's proof was called the little theorem of Cobham (le petit thorm de Cobham). Later on Michaux and Villemaire found another proof of Cobham's theorem based on logic and Presburger arithmetic [6]. Their idea was to make a contradiction with the syndetic result, which was called Hansel's lemma in [7]. It seems that the syndetic result is considered to be fairly easy since the attempts to make Cobham's proof more understandable are mainly focused on the second part. On the other hand, there are some totally different proofs. For example, Muchnik [8] gave a simple sophisticated proof which can also be applied to the multidimentional generalization of the theorem due to Semenov [12]. See also the survey by Bruyre, Hansel, Michaux and Villemaire [2].

3 On the proofs of Hansel's lemma

As mentioned in the previous section, Hansel's lemma is an essential part of some proofs of Cobham's theorem and it is stated as follows:

Theorem 2. *Let $k, l \geq 2$ be multiplicatively independent integers. If a set $X \subseteq \mathbb{N}$ is both k- and l-recognizable, then it is syndetic.*

Actually, neither Hansel nor Reutenauer proves this theorem. They just point out that is follows immediately from Lemma 3 in [3]. In 1990 Perrin gave a self-contained proof in [9] consisting of two propositions (Prop. III.8.10., Prop. III.8.11.). Many of the later proofs refer to this proof as a more accessible source than the works of Hansel and Reutenauer. Also recently published book on automatic sequences by Allouche and Shallit [1] uses a proof based on Perrin's ideas. Unfortunately, apart of its many merits, this proof contains some impreciseness. On the other hand, we note that there exists another kind of proof of Hansel's lemma in [7], the goal of which is a new proof of Semenov's Theorem.

Both Cobham's and Perrin's proofs are based on the fact that, for multiplicatively independent integers k and l, the set of quotients $\{k^p/l^q \mid p, q \geq 0\}$ is dense in the positive reals. In the first proposition Perrin uses the notion of right dense sets. We say that $X \subseteq \Sigma^*$ is *right dense* if for any word $u \in \Sigma^*$ there exists a word $v \in \Sigma^*$ such that uv belongs to the set X.

Proposition 1. *Let X be an infinite k-recognizable set of integers. For any integer l multiplicatively independent of k, the set $0^*(X)_l$ of (unnormalized) expansions of X at base l is right dense.*

The proof is similar to the proof of Cobham's Lemma 2, where Cobham proves also that for any word $u \in \Sigma$, the length of the word v such that uv belongs to the set X can be chosen to satisfy the congruence $|v| \equiv c \pmod{d}$ for any $d > c \geq 0$. This additional length condition is needed in the proof of Cobham's Lemma 3, which is a more precise formulation of Theorem 2. Perrin, instead, proves the following proposition, which together with Proposition 1 implies Theorem 2.

Proposition 2. *A k-recognizable set X is syndetic iff the set $0^*(X)_k$ of (unnormalized) expansions of X at base k is right dense.*

This statement is not exactly true, since we can easily construct a counter example. Consider, for example, the set of natural numbers with normalized k-base representation of even length, i.e., the set $X = \{n \in \mathbb{N} \,|\, (n)_k \equiv 0 \pmod{2}\}$. This set is clearly automatic and right dense. However, it is not syndetic, since there exists arbitrarily large intervals of the form $[k^{2l}, k^{2l+1} - 1]$ containing no elements of X. In the proof of Proposition III.8.11. it is shown that for any integer

n there exists two integers p and $t < k^p$ such that $nk^p + t \in X$. But the fact that the exponent p can be bounded does not imply syndetic result as stated. Instead, we would need a constant p for all n large enough like in the proof of Cobham's Lemma 3. This is the reason, why Cobham's proof is much more complicated and why he is obliged to apply his Lemma 2 several times.

Remark. We note, that also Rigo and Waxweiler have recently noticed the incompleteness of the proofs of Cobham's Theorem in [1] and [9]. In addition, they give an alternative proof for Hansel's lemma in their article [11].

References

[1] J.-P. Allouche, J. Shallit, Automatic sequences: Theory, Applications, Generalizations. Cambridge University Press, 2003.

[2] V. Bruyre, G. Hansel, C. Michaux and R. Villemaire, Logic and p-recognizable sets of integers, Bull. Belg. Math. Soc. **1**, 191–238, 1994. Corrigendum, Bull. Belg. Math. Soc. **1**, 577, 1994.

[3] A. Cobham, On the base-dependence of set of numbers recognizable by finite automata. Math. Systems Theory **3**, 186–192, 1969.

[4] S. Eilenberg, Automata, Languages, and Machines. Vol. A. Academic Press, 1974.

[5] G. Hansel, A propos d'un thorme de Cobham. In D. Perrin, editor, *Actes de la Fte des Mots, Greco de Programmation*, CNRS, Rouen, 55–59, 1982

[6] C. Michaux and R. Villemaire, Cobham's theorem seen through Bchi's theorem. In *Proc. 20th Int. Conf. on Automata, Languages, and Programming (ICALP), Vol. 700 of Lecture Notes in Computer Science,* 325–334, Springer-Verlag, 1993.

[7] C. Michaux and 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**, 251–277, 1996.

[8] A. Muchnik, Definable criterion for definability in Presburger Arithmetic and its application. preprint in Russian, Institute of New Technologies, 1991.

[9] D. Perrin, Finite automata, in J. Van Leeuwen, editor, *Handbook of Theoretical Computer Science, vol. B,*Elsevier, 2–57, 1990.

[10] C. Reutenauer, Dmonstration du thorme de Cobham sur les ensembles de nombres reconnaissables par automate fini, d'aprs Hansel, in *Sminaire d'Informatique Thorique, Anne 1983-84, Universit Paris 7, Paris*, 217–224, 1984.

[11] M. Rigo, L. Waxweiler, A note on syndeticity, recognizable sets and Cobham's theorem. To appear in Bull. EATCS.

[12] A.L. Semenov, Presburgerness of predicates regular in two number systems (in Russian), Sibirsk. Mat. Zh. **18**, 403–418, 1977, English translation, Siberian Math. J. **18**, 289-299, 1997.

TURKU CENTRE *for* COMPUTER SCIENCE

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration

- Institute of Information Systems Sciences

ISBN 952-12-1610-7
ISSN 1239-1891