



Numération : mathématiques et informatique

RENCONTRE ORGANISÉE PAR :
Boris Adamczewski, Anne Siegel and Wolfgang Steiner

23-27 mars 2009

Jason P. Bell

Cobham's theorem and its extensions

Vol. 1, n° 1 (2009), p. 11-16.

<http://acirm.cedram.org/item?id=ACIRM_2009__1_1_11_0>

Centre international de rencontres mathématiques
U.M.S. 822 C.N.R.S./S.M.F.
Luminy (Marseille) FRANCE

cedram

*Texte mis en ligne dans le cadre du
Centre de diffusion des revues académiques de mathématiques
<http://www.cedram.org/>*

Cobham's theorem and its extensions

Jason P. BELL

Abstract

Cobham's theorem says that if k and ℓ are two multiplicatively independent integers and $f(n)$ is a k - and ℓ -automatic sequence, then $f(n)$ is eventually periodic. We give a summary of recent work on automatic sequences and their relation to Cobham's theorem.

1. AUTOMATA AND COBHAM'S THEOREM

To talk about Cobham's theorem, it is necessary to first talk about *automatic sequences*. Automatic sequences have been employed in number theory, algebra, computer science, combinatorics, and other areas. As a result, there are many different ways of thinking about these sequences. We give a few equivalent formulations and illustrate it with the example of the *Thue-Morse sequence*.

Definition 1. Let $k \geq 2$ be a natural number. A sequence $\{f(n)\}_{n \geq 0}$ is k -automatic if there is a finite state automaton which accepts as input the base k expansion of n and outputs $f(n)$.

Cobham [12] gave an equivalent formulation in terms of *morphic sequences*. Automatic sequences are a subset of the collection of *morphic* sequences. We define morphic sequences here.

Let $\Sigma = \{a_1, \dots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \rightarrow \Sigma^*$ be a morphism. We say that a letter $a_i \in \Sigma$ is *mortal* if $\phi^j(a_i) = \epsilon$, the empty word, for some $j \geq 1$. Suppose that $\phi(a_1) = a_1x$ for some non-empty word x containing a non-mortal letter. Then we say that ϕ is prolongable on a_1 . Then we can form the right infinite word

$$\phi^\omega(a_1) := a_1x\phi(x)\phi^2(x)\dots$$

Note that the right-infinite word $\phi^\omega(a_1)$ is a fixed point of ϕ . Words defined in this manner are called *pure morphic words*. In general, a word u on a finite alphabet Δ is *morphic* if there is a set map $f : \Sigma \rightarrow \Delta$ such that $u = f(w)$ for some pure morphic word w on Σ .

In the case that the length of $\phi(a_1), \dots, \phi(a_d)$ are all the same length (uniform) and this common length is equal to a power of k , the sequence $f(n)$ formed by taking the n th letter of our morphic word is k -automatic. In fact, Cobham shows that every k -automatic sequence can be formed in this way. As an example, we consider the 2-automatic *Thue-Morse* sequence. This sequence is defined by the rule $f(n) = 1$ if the number of 1s in the binary expansion of n is odd; otherwise $f(n) = 0$. The finite state automaton which generates this sequence is easy to describe (see Allouche and Shallit [3, p. 153]). We note that this sequence can be described as a morphic sequence by using the alphabet $\Sigma = \{0, 1\}$ and the morphism $\phi : \Sigma^* \rightarrow \Sigma^*$ defined by

$$0 \mapsto 01 \quad 1 \mapsto 10.$$

Then

$$\phi^\omega(0) = 0110100110010110\dots$$

has the property that the n th term of the Thue-Morse sequence is the n th letter (starting with $n = 0$) of this morphic word.

Yet another formulation is in terms of *k-kernels*. Given a sequence $\{f(n)\}_{n \geq 0}$ we define the k -kernel of this sequence to be all distinct subsequences of the form $\{f(k^a n + b)\}_{n \geq 0}$, where $a \geq 0$ and $0 \leq b < k^a$. Note that a subsequence in the k -kernel runs over all indices n whose base

Text presented during the meeting "Numeration: mathematics and computer science" organized by Boris Adamczewski, Anne Siegel and Wolfgang Steiner. 23-27 mars 2009, C.I.R.M. (Luminy).

k -expansion have the property that the last a digits are precisely the base k digits of b (with $0 \leq b < k^a$). A sequence $f(n)$ is k -automatic if and only if its k -kernel is finite [3, Theorem 6.6.2].

We can see this in terms of the Thue-Morse sequence. If $f(n)$ is the Thue-Morse sequence, then $f(2n) = f(n)$ as the binary expansion of $2n$ has the same number of 1s as the binary expansion of n ; similarly, $f(2n + 1) = 1 - f(n)$. An easy induction argument shows there are exactly two sequences in the 2-kernel of $f(n)$.

Perhaps the least well-known of equivalent formulations is in terms of matrix semigroups. A complex sequence $f(n)$ is k -automatic if and only if there exist:

- (1) a natural number d ,
- (2) a set of $d \times d$ complex matrices A_0, \dots, A_{k-1} which generate a finite semigroup,
- (3) $d \times 1$ column vectors v and w

such that

$$f(n) = w^T A_{i_r} \cdots A_{i_0} v,$$

where $n = i_0 + i_1 k + \cdots + i_r k^r$. This linear algebraic formulation of automatic sequences has a lot of interesting consequences. We recall a famous theorem of Amitsur and Levitzki [7, p. 20].

Theorem 1.1. (*Amitsur-Levitzki*) *Let B_1, \dots, B_m be $d \times d$ matrices with entries in a commutative ring. Then*

$$(1.1) \quad \sum_{\sigma \in S_m} \text{sgn}(\sigma) B_{\sigma(1)} \cdots B_{\sigma(m)} = 0$$

whenever $m \geq 2d$.

We can think of a k -automatic sequence $f(n)$ as being a map $f : \{0, 1, \dots, k-1\}^* \rightarrow \Delta$, for some finite set Δ , with the property that $f(0W) = f(W)$ for $W \in \{0, 1, \dots, k-1\}^*$. Here the correspondence comes from noting that a number n has a base k -expansion which does not begin with the digit 0. An immediate corollary of the Amitsur-Levitzki theorem is that there exists some m such that

$$\sum_{\sigma \in S_m} \text{sgn}(\sigma) f(W_{\sigma(1)} \cdots W_{\sigma(m)}) = 0$$

for all words $W_1, \dots, W_m \in \{0, 1, \dots, k-1\}^*$. For more on this, see Berstel and Reutenauer [6]

Cobham's theorem [11] is one of the most important results in the theory of k -automatic sequences. It shows that being automatic with respect to two multiplicatively independent bases can only occur if the sequence is very nice. Two numbers k and ℓ are *multiplicatively independent* if there are no solutions in nonzero integers to the equation $k^a = \ell^b$.

Theorem 1.2. (*Cobham*) *Let k and ℓ be two multiplicatively independent integers. If a sequence $f(n)$ is both k - and ℓ -automatic then it is eventually periodic.*

In the next sections, we will talk about some recent extensions involving Cobham's theorem.

2. EXTENSIONS WITH MORPHIC SEQUENCES

We note that if $\Sigma = \{a_1, \dots, a_d\}$ be a finite alphabet and let $\phi : \Sigma^* \rightarrow \Sigma^*$ be a morphism. In §1, we showed how one can construct a morphic word with these data along with a simple substitution. Let w be a morphic word constructed in this manner. We can construct a $d \times d$ incidence matrix of ϕ in which the (i, j) -entry is the number of occurrences of a_j in $\phi(a_i)$. The Perron-Frobenius theorem states that this has a real positive eigenvalue which is also the maximum of the moduli of all eigenvalues of this matrix. We call this positive eigenvalue α the *Perron* number of the morphic word w and say that w is α -substitutive. We note that in the case that $\phi(a_1), \dots, \phi(a_d)$ all have length k then k is the Perron number of w . Moreover, the sequence is k -automatic. Hansel suggested the following generalization of Cobham's theorem.

Conjecture 2.1. (*Hansel*) *Suppose that w is a morphic word that is both α - and β -substitutive and $\alpha^a = \beta^b$ has no nonzero solutions in integers $(a, b) \in \mathbb{Z}^2$. Then w is eventually periodic.*

Some earlier work had been done on this conjecture by Fabre [16] and Fagnot [17]. The biggest results in this area are due to Durand [13, 14]. Durand [13] first showed that Hansel's conjecture is true for *primitive* substitutions. If w is constructed using an alphabet $\Sigma = \{a_1, \dots, a_d\}$ and

morphism $\phi : \Sigma^* \rightarrow \Sigma^*$ along with a substitution, then we say that w is *primitive* if the incidence matrix of ϕ has some power such that it has no nonzero entries. Durand [14] then showed that Hansel's conjecture holds for a large class of substitutions that included all known results in the area.

3. EXTENSIONS WITH k -KERNELS

Allouche and Shallit [2] extended the notion of k -automatic sequences, introducing k -regular sequences. To define these, we have to use a little commutative algebra.

Let R be a commutative ring. Given a sequence $\{f(n)\}_{n=0}^{\infty}$ taking values in some R -module, we create an R -module $M(\{f(n)\}; k)$ which is defined to be the R -module generated by all sequences $\{f(k^i n + j)\}_{n=0}^{\infty}$, where $i \geq 0$ and $0 \leq j < k^i$.

Definition 2. *Let R be a commutative ring and let k be a positive integer. A sequence is (R, k) -regular if $M(\{f(n)\}; k)$ is finitely generated as an R -module.*

Since the k -kernel of a sequence $\{f(n)\}$ spans $M(\{f(n)\}; k)$ as an R -module, we see that a k -automatic sequence with values in R is necessarily (R, k) -regular for any ring R .

Unlike automatic sequences, which only assume finitely many values, regular sequences can assume infinitely many values. For this reason it is unrealistic to assume that the correct analogue of Cobham's theorem for regular sequences is that an (R, k) - and (R, ℓ) -regular sequence is eventually periodic if k and ℓ are multiplicatively independent. There is, however, a larger class of sequences which gives the correct analogue.

Definition 3. *Given a commutative ring R and R -module M , we say that a map*

$$f : \mathbb{N} \rightarrow M$$

satisfies a linear recurrence over R , if there exist a positive integer m and constants $c_1, \dots, c_m \in R$ such that

$$f(n) = \sum_{i=1}^m c_i f(n-i) \quad \text{for } n \geq m.$$

If $\{f(n)\}$ satisfies a linear recurrence over a ring R and assumes only finitely many values, then $\{f(n)\}$ is eventually periodic (cf. Everest et al. [15, §3.1]). Furthermore, given an eventually periodic sequence $\{f(n)\}$ there exist numbers m and N such that $\{f(n)\}$ satisfies the linear recurrence $f(n+m) = f(n)$ for $n \geq N$. The following result can be proved.

Theorem 3.1. *Let R be a commutative ring, let k and ℓ be multiplicatively independent positive integers, and let $\{f(n)\}$ be a sequence which is both (R, k) - and (R, ℓ) -regular. Then $\{f(n)\}$ satisfies a linear recurrence over R .*

In light of the above remarks, this is indeed a generalization of Cobham's theorem. In the case that $R = \mathbb{Z}$, an (R, k) -regular sequence is called k -regular. In this case we get a simple characterization of sequences which are both k - and ℓ -regular if k and ℓ are multiplicatively independent.

Theorem 3.2. *Let $\{f(n)\}$ be an integer-valued sequence and let k and ℓ be two multiplicatively independent positive integers. Then $\{f(n)\}$ is both k - and ℓ -regular if and only if*

$$\sum_{n=0}^{\infty} f(n)x^n \in \mathbb{Z}[[x]]$$

is the power series expansion of a rational function whose poles are all roots of unity.

Theorems 3.1 and 3.2 are due to the author [5].

We note that Theorem 3.2 appears as a conjecture in Allouche and Shallit [3, §16.8, item 16.3]. In this form, it is a special case of a conjecture due to van der Poorten. Given a power series

$$\mathbf{F}(x) = \sum_{n=0}^{\infty} f(n)x^n \in \mathbb{Z}[[x]],$$

we say that $\mathbf{F}(x)$ is *k-Mahler* if $\mathbf{F}(x)$ satisfies a functional equation of the form

$$(3.2) \quad \mathbf{F}(x^{k^m}) = \sum_{i=0}^{m-1} p_i(x) \mathbf{F}(x^{k^i}).$$

k-regular sequences are a special subset of the set of *k*-Mahler sequences.

Conjecture 3.3. (*van der Poorten*) *Let $\mathbf{F}(x)$ be a power series which is both *k*- and ℓ -Mahler. If *k* and ℓ are multiplicatively independent then $\mathbf{F}(x)$ is the power series expansion of a rational function.*

Some work has been done on this by various authors [23, 4].

Finally, it is interesting to ask whether we can obtain a generalization of Cobham's theorem by replacing finiteness of the *k*-kernel by some weaker condition. Note that, generically, we expect a sequence $f(n)$ to have k^a distinct subsequences of the form $\{f(k^a n + b)\}_{n \geq 0}$ with $0 \leq b < k^a$. With this in mind, we introduce a function to deal with the *growth* of the *k*-kernel.

Definition 4. *Given a sequence $\{f(n)\}_{n \geq 0}$, we define $G(f(n), k; m)$ to be the number of distinct subsequences of the form $f(k^a n + b)$ with $0 \leq a \leq m$ and $0 \leq b < k^a$. We call the function $G(f(n), k; m)$ the *k*-growth of $f(n)$ and think of this as a function of m .*

As we said earlier, the *k*-growth, $G(f(n), k; m)$, of $f(n)$ generally grows exponentially as a function of m . Note that $G(f(n), k; m)$ is eventually constant if and only if $f(n)$ is *k*-automatic. Note that the growth could grow subexponentially or polynomially without being eventually constant, however. With this in mind, we define a new dimension for the sequence $f(n)$.

Definition 5. *Given a sequence $f(n)$ we define the *k*-dimension of $f(n)$ to be*

$$\text{Dim}_k(f(n)) := \limsup_m \frac{\log G(f(n), k; m)}{\log m}.$$

As an example, let $k \geq 2$ and consider the sequence $f(n)$ which is defined to be 1 if $n = k^{a^2}$ for some natural number a ; otherwise $f(n) = 0$. Then one can easily check that $G(f(n), k, m) = m + 1$ and so $\text{Dim}_k(f(n)) = 1$. The definition of *k*-dimension is inspired by the notion of growth in groups and algebras. Finiteness of the corresponding dimensions is very important in these contexts. In particular, see the theorem of Gromov [21, Theorem 11.1]. In light of this, we make the following conjecture.

Conjecture 3.4. *Let $f(n)$ be a sequence taking values in a finite set and let *k* and ℓ be two multiplicatively independent numbers. If $\text{Dim}_k(f(n))$ and $\text{Dim}_\ell(f(n))$ are both finite, then $f(n)$ is eventually periodic.*

4. EXTENSIONS WITH HAHN POWER SERIES

One of the beautiful applications of Cobham's theorem comes from a result of Christol, Kamae, Mendès France and Rauzy [10]. The motivation for their work comes from a number theoretic problem concerning the expansion of algebraic numbers in integer bases. It appears at the end of a paper of Mendès France [22]. It can be stated as follows. Let $\mathbf{a} = (a_n)_{n \geq 0}$ be a binary sequence and consider the two real numbers

$$\alpha = \sum_{n \geq 0} \frac{a_n}{2^n} \quad \text{and} \quad \beta = \sum_{n \geq 0} \frac{a_n}{3^n}.$$

Then, the problem is to show that these numbers are both algebraic if and only if they are both rational. At first glance, this problem seems contrived, but behind it hides the more fundamental question of the structure of representations of real numbers in two multiplicatively independent integer bases. Unfortunately, problems of this type are extremely difficult.

However, when considering addition and multiplication without carry things become easier. In particular, Christol, Kamae, Mendès France and Rauzy [10] gave an analogue of this problem in terms of power series.

Theorem 4.1 (Christol *et al.*). Let p_1 and p_2 be distinct prime numbers and let q_1 and q_2 be powers of prime p_1 and p_2 respectively. Let $(a_n)_{n \geq 0}$ be a sequence with values in a finite set A with cardinality at most $\min\{q_1, q_2\}$. Let i_1 and i_2 be two injections from A into \mathbb{F}_{q_1} and \mathbb{F}_{q_2} respectively. Then, the formal power series

$$f(t) = \sum_{n \geq 0} i_1(a_n)t^n \in \mathbb{F}_{q_1}((t)) \quad \text{and} \quad g(t) = \sum_{n \geq 0} i_2(a_n)t^n \in \mathbb{F}_{q_2}((t))$$

are both algebraic (respectively over $\mathbb{F}_{q_1}(t)$ and $\mathbb{F}_{q_2}(t)$) if and only if they are rational functions.

As was remarked by Christol *et al.* [10], Theorem 4.1 is a straightforward consequence of two important results. On one side, Christol's theorem [9] describes precisely in terms of automata the algebraic closure of $\mathbb{F}_q(t)$ in $\mathbb{F}_q((t))$ (q being a power of a prime p). On the other side, one finds Cobham's theorem [11] proving that for multiplicatively independent positive integers k and ℓ , a function $h : \mathbb{N} \rightarrow \mathbb{F}_q$ that is both k - and ℓ -automatic is eventually periodic.

Christol's theorem gives a very concrete description of the elements of $\mathbb{F}_q((t))$ (q a power of a prime p) that are algebraic over $\mathbb{F}_q(t)$; it shows in fact that being an algebraic power series is equivalent to the sequence of coefficients being p -automatic. As Kedlaya [19] points out, this result does not give the complete picture, as the field $\mathbb{F}_q((t))$ is far from being algebraically closed. Indeed, for an algebraically closed field \mathbb{K} of characteristic 0, the field

$$\bigcup_{i=1}^{\infty} \mathbb{K}((t^{1/i}))$$

is itself algebraically closed and contains in particular the algebraic closure of $\mathbb{K}(t)$; but in positive characteristic, things are rather different. The algebraic closure of $\mathbb{F}_q((t))$ is more complicated, due to the existence of wildly ramified field extensions. For instance, Chevalley remarked [8] that the Artin-Schreier polynomial $x^p - x - 1/t$ does not split in the field $\bigcup_{n=1}^{+\infty} \mathbb{F}_q((t^{1/n}))$.

It turns out that the appropriate framework to describe the algebraic closure of $\mathbb{F}_p(t)$ is provided by the fields of generalized power series $\mathbb{F}_q((t^{\mathbb{Q}}))$ introduced by Hahn [18]. This consists of all power series in which we allow rational exponents (as opposed to just integer exponents). In order to make this a ring, we insist that our power series have well-ordered support. That is, the set of all rational numbers α such that t^α has a nonzero coefficient in our power series must be a well-ordered set.

The work of Kedlaya [20] is precisely devoted to a description of the algebraic closure of $\mathbb{F}_p(t)$ in such fields of generalized power series. For this purpose, Kedlaya introduces the notion of a p -quasi-automatic function over the rationals. His extension of Christol's theorem is that if q is a power of a prime p , then a generalized power series

$$\sum_{\alpha \in \mathbb{Q}} h(\alpha)t^\alpha \in \mathbb{F}_q((t^{\mathbb{Q}}))$$

is algebraic if and only if the function $h : \mathbb{Q} \rightarrow \mathbb{F}_q$ is p -quasi-automatic.

Motivated by the work of Kedlaya, Adamczewski and Bell proved an analogue of Cobham's theorem for maps that are quasi-automatic with respect to two multiplicatively independent bases [1].

REFERENCES

- [1] B. Adamczewski and J. Bell. Function fields in positive characteristic: expansions and Cobham's theorem. *J. Algebra* **319** (2008), no. 6, 2337–2350.
- [2] J.-P. Allouche and J. Shallit. The ring of k -regular sequences. *Theoret. Comput. Sci.* **98** (1992), 163–197.
- [3] J.-P. Allouche and J. Shallit. *Automatic sequences. Theory, applications, generalizations*. Cambridge University Press, Cambridge, 2003.
- [4] P.-G. Becker. K -regular power series and Mahler-type functional equations. *J. Number Theory* **49** (1994), 269–286.
- [5] J. Bell. A generalization of Cobham's theorem for regular sequences. *Sém. Lothar. Combin.* 54A (2005/07), Art. B54Ap, 15 pp.
- [6] J. Berstel and C. Reutenauer. *Rational Series and Their Languages* EATCS Monographs on Theoretical Computer Science (12), W. Brauer, G. Rozenberg, A. Saloma (Eds.) Springer-Verlag Berlin, Heidelberg 1988.

- [7] B. Bollobás. *Graph theory. An introductory course*. Springer-Verlag, New York-Berlin, 1979.
- [8] C. Chevalley. *Introduction to the Theory of Algebraic Functions of One Variable*. Mathematical Surveys No. VI, Amer. Math. Soc., 1951.
- [9] G. Christol. Ensembles presque périodiques k -reconnaissables. *Theor. Comput. Sci.* **9** (1979), 141–145.
- [10] G. Christol, T. Kamae, M. Mendès France & G. Rauzy. Suites algébriques, automates et substitutions. *Bull. Soc. Math. France* **108** (1980), 401–419.
- [11] A. Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Math. Systems Theory* **3** (1969), 186–192.
- [12] A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.
- [13] F. Durand. A generalization of Cobham’s theorem. *Theory Comput. Systems* **31** (1998), 169–185.
- [14] F. Durand. A theorem of Cobham for non-primitive substitutions. *Acta Arith.* **104** (2002), no. 3, 225–241.
- [15] G. Everest, A. van der Poorten, I. Shparlinski, and T. Ward. *Recurrence sequences*. Mathematical Surveys and Monographs, 104. American Mathematical Society, Providence, RI, 2003.
- [16] S. Fabre. Une généralisation du théorème de Cobham. *Acta Arith.* **67** (1994), 197–208.
- [17] I. Fagnot. On the subword equivalence problem for morphic words. *Discrete Appl. Math.* **75** (1997), no. 3, 231–253.
- [18] H. Hahn. Über die nichtarchimedische Größensysteme (1907), reprinted in *Gesammelte Abhandlungen I*, Springer-Verlag, 1995.
- [19] K. Kedlaya. The algebraic closure of the power series field in positive characteristic. *Proc. Amer. Math. Soc.* **129** (2001), 3461–3470.
- [20] K. Kedlaya. Finite automata and algebraic extensions of function fields. *J. Théor. Nombres Bordeaux* **18** (2006), 379–420.
- [21] G. Krause and T. Lenagan. *Growth of Algebras and Gelfand–Kirillov Dimension*, revised edition. Grad. Stud. Math., vol. 22, Amer. Math. Soc., Providence, RI, 2000.
- [22] M. Mendès France. Sur les décimales des nombres algébriques réels, in *Sémin. Théor. Nombres, Bordeaux, 1979–1980*, Exp. No. 28, 7 pp., Univ. Bordeaux I, Talence, 1980.
- [23] B. Randé. Équations fonctionnelles de Mahler et applications aux suites p -régulières. PhD thesis, Université Bordeaux I, 1992.

Jason Bell, Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada •
 jpb@math.sfu.ca