Automata in Number Theory *

Boris Adamczewski[1] *and Jason Bell*[2]

[1] CNRS, Université de Lyon
Institut Camille Jordan
43 boulevard du 11 novembre 1918
F-69622 Villeurbanne Cedex, France
email: Boris.Adamczewski@math.cnrs.fr

[2] Department of Pure Mathematics
University of Waterloo
Waterloo, ON, Canada
N2L 3G1
email: jpbell@uwaterloo.ca

October 7, 2018 19 h 47

2010 Mathematics Subject Classification: 11B85,11J81,11J87

Key words: Finite automata, automatic sequences, automatic sets, prime numbers, algebraic numbers, Diophantine approximation, linear recurrences, Laurent series, generalized power series.

Contents

*Work supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 648132.

1 Introduction

Among infinite sequences or infinite sets of integers, some are well-behaved, such as periodic sequences and arithmetic progressions, whereas others, such as random sequences and random sets, are completely chaotic and cannot be described in a simple way. Finite automata are one of the most basic models of computation and thus lie at the bottom of the hierarchy associated with Turing machines. Such machines can naturally be used to generate sequences with values over a finite set, and also as devices to recognize certain subsets of the integers.

One of the main interests of these automatic sequences/sets arises from the fact that they are in many ways very well-behaved without necessarily being trivial. One can thus consider that they lie somewhere between order and chaos, even if, in many respects, they are well-behaved.

In this chapter, we survey some of the connections between automatic sequences/sets and number theory. Several substantial advances have recently been made in this area and we give an overview of some of these new results. This includes discussions about prime numbers, the decimal expansion of algebraic numbers, the search for an analogue of the Skolem-Mahler-Lech theorem in positive characteristic and the description of an algebraic closure of the field $\mathbb{F}_p(t)$.

2 Automatic sequences and automatic sets of integers

In this section, we recall some basic facts about automatic sequences and automatic sets of integers. The main reference on this topic is the book of Allouche and Shallit [7]. An older reference is Eilenberg [20]. In [20] k-automatic sets are called k-recognizable.

2.1 Automatic sequences

Let $k \geqslant 2$ be an integer. An infinite sequence $(a_n)_{n\geqslant 0}$ is said to be k-automatic if a_n is a finite-state function of the base-k representation of n. This means that there exists a deterministic finite automaton with output (DFAO) taking the base-k expansion of n as input and producing the term a_n as output. We say that a sequence is generated by a finite automaton, or for short is *automatic*, if it is k-automatic for some $k \geqslant 2$.

A more concrete definition of k-automatic sequences can be given as follows. Let A_k denote the alphabet $\{0, 1, \ldots, k-1\}$. By definition, a k-automaton is a 6-tuple

$$\mathcal{A} = (Q, A_k, \delta, q_0, \Delta, \tau)\,,$$

where Q is a finite set of states, $\delta : Q \times A_k \to Q$ is the transition function, q_0 is the initial state, Δ is the output alphabet and $\tau : Q \to \Delta$ is the output function. For a state q in Q and for a finite word $w = w_1 w_2 \cdots w_n$ on the alphabet A_k, we define $\delta(q, w)$ recursively by $\delta(q, w) = \delta(\delta(q, w_1 w_2 \cdots w_{n-1}), w_n)$. Let $n \geqslant 0$ be an integer and let $w_r w_{r-1} \cdots w_1 w_0$ in $(A_k)^{r+1}$ be the base-k expansion of n starting with the most significant digit. Thus $n = \sum_{i=0}^{r} w_i k^i := [w_r w_{r-1} \cdots w_0]_k$. We let $w(n)$ denote the word $w_r w_{r-1} \cdots w_0$. Then a sequence $(a_n)_{n\geqslant 0}$ is said to be *k-automatic* if there exists a k-automaton $\mathcal{A}$ such that $a_n = \tau(\delta(q_0, w(n)))$ for all $n \geqslant 0$.

Example 2.1. The Thue–Morse sequence $t := (t_n)_{n\geqslant 0}$ is probably the most famous example of an automatic sequence. It is defined as follows: $t_n = 0$ if the sum of the binary digits of n is even, and $t_n = 1$ otherwise. We thus have

$$t = 01101001100101\cdots\,.$$

It is easy to check that the Thue–Morse sequence can be generated by the following finite 2-automaton: $\mathcal{A} = (\{A, B\}, \{0, 1\}, \delta, A, \{0, 1\}, \tau)$, where $\delta(A, 0) = \delta(B, 1) = A$, $\delta(A, 1) = \delta(B, 0) = B$, $\tau(A) = 0$ and $\tau(B) = 1$.

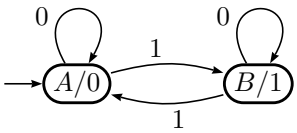

Figure 1. A DFAO generating Thue–Morse word.

Example 2.2. Let $w = (w_n)_{n\geqslant 0}$ be the 3-automatic sequence generated by the DFAO represented in Figure 2. Note that though this 3-automaton has only two states, it seems to be non-trivial to give a simple expression of w_n as a function of the ternary expansion of n.

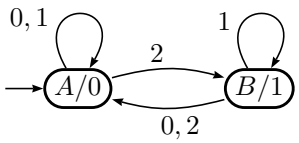

Figure 2. A DFAO generating the sequence w.

2.1.1 Morphisms of free monoids For a finite set A, we let A^* denote the free monoid generated by A. The empty word ε is the identity element of A^*. Let A and B be two finite sets. A map from A to B^* extends uniquely to a homomorphism between the free monoids A^* and B^*. We call such a homomorphism from A^* to B^* a *morphism*. If there is a positive integer k such that each element of A is mapped to a word of length k, then the morphism is called k-*uniform* or simply *uniform*. A *coding* is a 1-uniform morphism.

A morphism σ from A^* to itself is said to be *prolongable* if there exists a letter a such that $\sigma(a) = aw$, where the word w is such that $\sigma^n(w)$ is a nonempty word for every $n \geqslant 0$. In that case, the sequence of finite words $(\sigma^n(a))_{n\geqslant 0}$ converges in $A^\omega = A^* \cup A^{\mathbb{N}}$, endowed with its usual topology, to an infinite word denoted $\sigma^\omega(a)$. This infinite word is clearly a fixed point for σ (extended by continuity to infinite words) and we say that $\sigma^\omega(a)$ is generated by the morphism σ.

For instance, the morphism τ defined over the alphabet $\{0, 1\}$ by $\tau(0) = 01$ and $\tau(1) = 10$ is a 2-uniform morphism that generates the Thue–Morse sequence

$$t = \tau^\omega(0) = 01101001100101\cdots .$$

Uniform morphisms and automatic sequences are strongly connected, as the following classical result of Cobham shows [15]. A notable consequence of Theorem 2.1 is that finite automata are Turing machines that produce sequences in linear time.

Theorem 2.1 (Cobham). *An infinite word is k-automatic if and only if it is the image by a coding of a word that is generated by a k-uniform morphism.*

Example 2.3. Let us consider the 3-uniform morphism ω defined over the monoid $\{0, 1, 2\}^*$ by $\omega(0) = 012$, $\omega(1) = 020$, and $\omega(2) = 021$. This morphism has a unique fixed point

$$x = \omega^\omega(0) = 012020021012021012012\cdots .$$

Letting ϕ denote the coding that maps 0 and 1 to 0, and 2 to 1, we thus obtain that

$$y := y_1y_2\cdots := \phi(x) = 001010010001010001001\cdots$$

is a 3-automatic word.

Example 2.4. The word w defined in Example 2.2 is the unique fixed point generated by the binary morphism ψ satisfying $\psi(0) = 001$ and $\psi(1) = 010$.

2.1.2 Kernels An important notion in the study of k-automatic sequences is the notion of k-kernel. The k-kernel of a sequence $a = (a_n)_{n\geqslant 0}$ is defined as the set of subsequences

$$N_k(a) = \left\{(a_{k^in+j})_{n\geqslant 0} \colon i \geqslant 0,\ 0 \leqslant j < k^i\right\} .$$

This notion gives rise to another useful characterization of k-automatic sequences which was first proved by Eilenberg in [20].

Theorem 2.2 (Eilenberg). *A sequence is k-automatic if and only if its k-kernel is finite.*

Example 2.5. The 2-kernel of the Thue–Morse sequence t has only two elements: t and the sequence $\overline{t}$ obtained by exchanging the symbols 0 and 1 in t.

2.2 Automatic sets of integers

Another important aspect of finite automata is that they can naturally be used as a device to recognize sets of integers.

2.2.1 Automatic subsets of $\mathbb{N}$ A set $\mathcal{N} \subset \mathbb{N}$ is said to be recognizable by a finite k-automaton, or for short *k-automatic*, if the characteristic sequence of $\mathcal{N}$, defined by $a_n = 1$ if $n \in \mathcal{N}$ and $a_n = 0$ otherwise, is a k-automatic sequence. This means that there exists a finite k-automaton that reads as input the base-k expansion of n and accepts this integer (producing as output the symbol 1) if n belongs to $\mathcal{N}$; otherwise this automaton rejects the integer n, producing as output the symbol 0.

Example 2.6. The simplest automatic sets are arithmetic progressions. Moreover, arithmetic progressions have the very special property of being k-automatic sets for every integer $k \geqslant 2$ (see Cobham's theorem in Chapter 26).

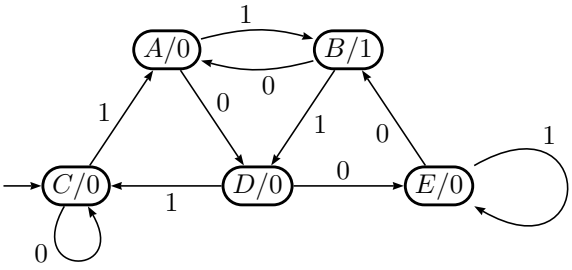

Figure 3. A 2-DFAO recognizing the arithmetic progression $5\mathbb{N} + 3$.

Example 2.7. The set $\{1, 2, 4, 8, 16, \ldots\}$ formed by the powers of 2 is also a typical example of a 2-automatic set.

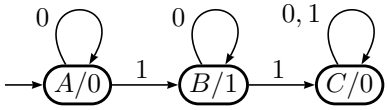

Figure 4. A 2-DFAO recognizing the powers of 2.

Example 2.8. In the same spirit, the set formed by taking all integers that can be expressed as the sum of at most two powers of 3 is 3-automatic (see Figure 5).

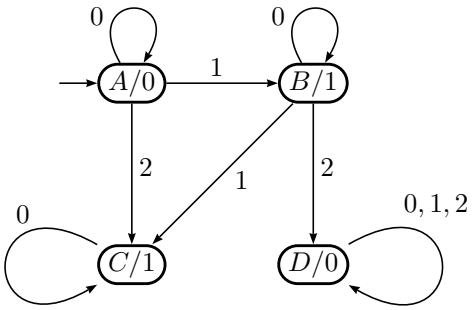

Figure 5. A 3-DFAO recognizing those integers that are the sum of tat most wo powers of 3.

There are also much stranger automatic sets. For instance, the set of integers whose binary expansion has an odd number of digits, does not contain three consecutive 1's, and contains an even number of two consecutive 0's is a 2-automatic set. Furthermore, the class of k-automatic sets is closed under various natural operations such as intersection, union and complement. On the other hand, some classical sets of integers, such as the set of prime numbers and the set of perfect squares, cannot be recognized by a finite automaton (see Theorem 3.1 and [56, 49]).

2.2.2 Automatic subsets of $\mathbb{N}^d$ and multidimensional automatic sequences Salon [58] extended the notion of automatic sets to include subsets of $\mathbb{N}^d$, where $d \geqslant 1$. To describe Salon's construction, we let A_k denote the alphabet $\{0, 1, \ldots, k-1\}$. We then consider an automaton

$$\mathcal{A} = \left(Q, A_k^d, \delta, q_0, \Delta, \tau\right) ,$$

where Q is a finite set of states, $\delta : Q \times A_k^d \to Q$ is the transition function, q_0 is the initial state, Δ is the output alphabet and $\tau : Q \to \Delta$ is the output function. Just as in the one-dimensional case, for a state q in Q and for a finite word $w = w_1 w_2 \cdots w_n$ on the alphabet A_k^d, we recursively define $\delta(q, w)$ by $\delta(q, w) = \delta(\delta(q, w_1 w_2 \cdots w_{n-1}), w_n)$. We call such an automaton a *d-dimensional k-automaton*.

We identify $\left(A_k^d\right)^*$ with the subset of $\left(A_k^*\right)^d$ consisting of all d-tuples $(u_1, \ldots, u_d)$ such that $u_1, \ldots, u_d$ all have the same length. Each nonnegative integer n can be written uniquely as

$$n = \sum_{j=0}^{\infty} e_j(n) k^j ,$$

in which $e_j(n) \in \{0, \ldots, k-1\}$ and $e_j(n) = 0$ for all sufficiently large j. Let $(n_1, \ldots, n_d)$ be a d-tuple of nonnegative integers and let

$$h := \max\left(\lfloor \log n_1 / \log\ k \rfloor, \cdots, \lfloor \log n_d / \log\ k \rfloor\right) ,$$

that is, if a_i represents the number of digits in the base-k expansion of n_i, then $h+1$ is the maximum of $a_1, \ldots, a_r$. We can then produce an element

$$w_k(n_1, \ldots, n_d) := (w_1, \ldots, w_d) \in \left(A_k^d\right)^*$$

corresponding to $(n_1, \ldots, n_d)$ by defining

$$w_i := e_h(n_i)e_{h-1}(n_i)\cdots e_0(n_i)\,.$$

In other words, we are taking the base-k expansions of $n_1, \ldots, n_r$ and then "padding" the expansions of each n_i at the beginning with 0's if necessary to ensure that each expansion has the same length. We say that a map $f : \mathbb{N}^d \to \Delta$ is k-automatic if there is a d-dimensional k-automaton $\mathcal{A} = \left(Q, A_k^d, \delta, q_0, \Delta, \tau\right)$ such that

$$f(n_1, \ldots, n_d) = \tau(\delta(q_0, w_d(n_1, \ldots, n_d)))\,.$$

Similarly, we define a k-automatic subset of $\mathbb{N}^d$ to be a subset S such that the characteristic function of S, $f : \mathbb{N}^d \to \{0,1\}$, defined by $f(n_1, \ldots, n_d) = 1$ if $(n_1, \ldots, n_d) \in S$; and $f(n_1, \ldots, n_d) = 0$, otherwise, is k-automatic.

Example 2.9. Let $f : \mathbb{N}^2 \to \{0,1\}$ be defined by $f(n,m) = 1$ if the sum of the binary digits of n added to the sum of the binary digits of m is even, and $f(n,m) = 0$ otherwise. Then $f(m,n)$ is a 2-automatic map. One can check that f can be generated by the following 2-dimensional 2-automaton: $\mathcal{A} = \left(\{A,B\}, \{0,1\}^2, \delta, A, \{0,1\}, \tau\right)$, where $\delta(A,(0,0)) = \delta(A,(1,1)) = \delta(B,(1,0)) = \delta(B,(0,1)) = A$, $\delta(A,(1,0)) = \delta(A,(0,1)) = \delta(B,(0,0)) = \delta(B,(1,1)) = B$, $\tau(A) = 1$ and $\tau(B) = 0$.

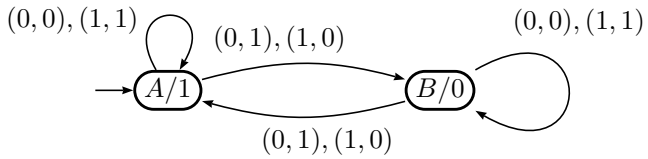

Figure 6. A DFAO generating the map f defined in Example 2.9.

Just as k-automatic sequences can be characterized by the finiteness of the k-kernel, multidimensional k-automatic sequences have a similar characterization.

Definition 2.1. Let d be a positive integer and let Δ be a finite set. We define the *k-kernel* of a map $f : \mathbb{N}^d \to \Delta$ to be the collection of all maps of the form

$$g(n_1, \ldots, n_d) := f(k^a n_1 + b_1, \ldots, k^a n_d + b_d)$$

where $a \geqslant 0$ and $0 \leqslant b_1, \ldots, b_d < k^a$.

For example, if $f : \mathbb{N}^2 \to \{0,1\}$ is the map defined in Example 2.9, then the 2-kernel of f consists of the 2 maps $f_1(m,n) := f(m,n)$, $f_2(m,n) = f(2m+1, 2n)$. Just as Eilenberg [20] showed that being k-automatic is equivalent to having a finite k-kernel for k-automatic sequences, Salon [58, Theorem 1] showed that a similar characterization of multidimensional k-automatic maps holds.

Theorem 2.3 (Salon). *Let d be a positive integer and let Δ be a finite set. A map $f : \mathbb{N}^d \to \Delta$ is k-automatic if and only if its k-kernel is finite.*

3 Prime numbers and finite automata

In this section, we briefly discuss some results concerning primes and finite automata.

3.1 Primes and randomness

An efficient way to produce conjectures about prime numbers comes from the so-called Cramér probabilistic model (see [16, 64, 65]). It is based on the principle that the set $\mathcal{P}$ of prime numbers behaves roughly like a random sequence, in which an integer of size about n has—as inspired by the prime number theorem—a 1 in $\log n$ chance of being prime. Of course, this probabilistic model has some limitations: for instance prime numbers are all odd with only one exception (see [53] for more about such limitations). Thus the set of prime numbers should be thought of as being a hybrid set rather than as a pseudorandom set (see the discussion in [66]). However, the Cramér model allows one to predict precise answers concerning occurrences of large gaps between consecutive prime numbers and concerning small gaps between primes (twin prime conjecture) and of some special patterns in $\mathcal{P}$ such as arithmetic progressions (Hardy–Littlewood conjectures). Some spectacular breakthrough were made recently in the two latter topics. See in particular [26] and [25].

A consequence of this probabilistic way of thinking is that the set $\mathcal{P}$ should be sufficiently random that it cannot be recognized by a finite automaton. This result was in fact proved to be true by Minsky and Papert [49] in 1966.

Theorem 3.1 (Minsky and Papert). *The set of prime numbers cannot be recognized by a finite automaton.*

Schützenberger [60] (also see [29]) even proved the stronger result that an automatic set always contains infinitely many composite numbers.

Theorem 3.2 (Schützenberger). *No infinite subset of the set of prime numbers can be recognized by a finite automaton.*

The intriguing question we are now left with is: *how can we prove that a set of integers is not automatic?* There are actually several different approaches: one can use the k-kernel and show that it is infinite or one can use some density properties (the logarithmic frequency of an automatic set exists; also if an automatic set has a positive density then it is rational). Another very efficient tool is the so-called pumping lemma, which is recalled below. For more details about the different ways of proving that a sequence is not automatic, we refer the reader to [7].

Lemma 3.3 (Pumping lemma). *Let $\mathcal{N} \subset \mathbb{N}$ be a k-automatic set. Then for every sufficiently large integer n in $\mathcal{N}$, there exist finite words w_1, w_2 and w_3, with $|w_2| \geqslant 1$, such that $n = [w_1w_2w_3]_k$ and $[w_1w_2^iw_3]_k$ belong to $\mathcal{N}$ for all $i \geqslant 1$.*

Sketch of proof. Let $n = [a_ra_{r-1}\cdots a_0]_k$ be an element of $\mathcal{N}$ and assume that r is larger than the number of states in the underlying automaton. By the pigeonhole principle, there is a state that is encountered twice when reading the input $a_0a_1\cdots a_r$, say just after reading a_i and a_j, $i < j$. Then setting $w_1 = a_r\cdots a_{j+1}$, $w_2 = a_j\cdots a_{i+1}$ and $w_3 = a_i\cdots a_0$, gives the result. □

Proof of Theorem 3.2. Let us assume that $\mathcal{N}$ is an infinite k-automatic set consisting only of prime numbers. Let p be an element of $\mathcal{N}$ that is sufficiently large to apply the pumping lemma. By the pumping lemma, there exist finite words w_1, w_2 and w_3, with $|w_2| \geqslant 1$, such that $p = [w_1w_2w_3]_k$ and such that all integers of the form $[w_1w_2^iw_3]_k$, with $i \geqslant 1$, belong to $\mathcal{N}$. However, it is not difficult to see, by using Fermat's little theorem, that $[w_1w_2^pw_3]_k \equiv [w_1w_2w_3]_k(\mod p)$ and thus $[w_1w_2^pw_3]_k \equiv 0(\mod p)$. It follows that the integer $[w_1w_2^pw_3]_k$ belongs to $\mathcal{N}$ but is not a prime number. Hence we obtain a contradiction. □

3.2 Primes in automatic sets

We have just seen that the set of all prime numbers is not automatic. However, it is believed that many automatic sets should contain infinitely many prime numbers. The most basic example of such a result is the famous Dirichlet theorem.

Theorem 3.4 (Dirichlet). *Let a and b be two relatively prime positive integers. Then the arithmetic progression $a\mathbb{N} + b$ contains infinitely many primes.*

Note that the special case of the arithmetic progression $2\mathbb{N} + 1$ was known by Euclid and his famous proof that there are infinitely many prime numbers. A more complete discussion about Dirichlet's theorem can be found in [54]. Beyond Dirichlet's theorem, the more general result concerning automatic sets and prime numbers is Theorem 3.5 from [23]. Recall that an automaton is *irreducible* if for all pairs of states (A, B) there is a path from A to B. Recall also that a positive integer is an *r-almost prime* if it is the product of at most r prime numbers. It is well-known that results about almost-primes are much easier to prove than those concerning primes (compare for instance Chen's theorem [12] with known results about the twin prime conjecture and the Goldbach conjecture).

Theorem 3.5 (Fouvry and Mauduit). *Given an automatic set $\mathcal{N} \subset \mathbb{N}$ associated with an irreducible automaton, there exists a positive integer r such that $\mathcal{N}$ contains infinitely many r-almost primes.*

Theorem 3.5 is not too difficult to prove using results similar to Chen's theorem. In contrast, to prove that there are infinitely many primes in sparse automatic sets such as $\{2^n - 1,\ n \geqslant 1\}$ and $\{2^n + 1,\ n \geqslant 1\}$ appears to be extremely difficult. This would solve two long-standing conjectures about the existence of infinitely many Fermat primes and Mersenne primes.

3.3 A problem of Gelfond: the sum of digits of prime numbers

Given a natural number n and a base b, we let $s_b(n)$ denote the sum of the digits of n in base b. Given two natural numbers a and m with $0 \leqslant a < m$ and $\gcd(m, b-1) = 1$, one can then look at the set of positive integers n such that $s_b(n) \equiv a \pmod m$. This set is known to be recognizable by a finite b-automaton. In 1968, Gelfond [24] asked about the collection of prime numbers that belong to this set. Theorem 3.5 implies that such a set contains infinitely r-almost primes for some r, but until recently it was still not known whether it contains infinitely many primes. Remarkably, Mauduit and Rivat [46] proved a much stronger result that gives the exact proportion of primes that belong to this automatic set. As usual with analytic number theory, the proof of their result—which relies on strong estimates of exponential sums—is long and difficult. As an example, an immediate corollary of the work of Mauduit and Rivat is that half of prime numbers belong to the Thue–Morse set $\{1, 2, 4, 7, 8, 11, 13, \ldots\}$.

Theorem 3.6 (Mauduit and Rivat). *One has*

$$\lim_{N\to\infty} \frac{\{0 \leqslant n \leqslant N, n \in \mathcal{P} \text{ and } s_2(n) \equiv 1 \pmod 2\}}{\{0 \leqslant n \leqslant N, n \in \mathcal{P}\}} = \frac{1}{2}.$$

4 Expansions of algebraic numbers in integer bases

The decimal expansions of classical constants like $\sqrt{2}$, π and e appear to be very mysterious and have baffled mathematicians for a long time. Numerical observations suggest that a complex underlying structure exists and several famous mathematicians have suggested possible rigorous definitions to try to formalize what "complex structure" actually means (see, for instance, [10, 50, 30]). These mathematicians were mainly influenced by notions from probability theory, dynamical systems, or theoretical computer science. These pioneering works lead us to a cluster of interesting conjectures concerning expansions of irrational periods in integer bases. However, even some of the simplest questions one can ask about the decimal expansions of classical irrational constants are still far out of reach.

The seminal work of Turing [67] gives rise to a rough classification of real numbers. On one side we find computable real numbers; that is, real numbers whose binary (or more generally base-b) expansion can be produced by a Turing machine, while on the other side lie uncomputable real numbers which, in some sense, "evade computers." Though most real numbers belong to the second class (the first one being countable), classical mathematical constants are usually computable. Following the pioneering ideas of Turing, Hartmanis and Stearns [30] proposed the emphasis of the quantitative aspect of the notion of computability, and to take into account the number $T(n)$ of operations needed by a (multitape) Turing machine to produce the first n digits of the expansion. In this regard, a real number is considered to be simple if its base-b expansion can be produced quickly by a Turing machine. A general problem is then to determine where our mathematical constants take place in such a classification. It is a source of challenging open questions such as the Hartmanis–Stearns problem which asks whether there exists an irrational algebraic number computable in linear time; that is, with $T(n) = O(n)$.

In 1968, Cobham [14] suggested to restrict this problem to a particular class of Turing machines, namely to the case of finite automata. Several attempts at a resolution to this problem are due to Cobham in 1968 [14] and to Loxton and van der Poorten [39, 40] during the 1980s. Both of these works are based on the so-called Mahler transcendence method [41]. The aim of this section is to give a proof, due to Adamczewski and Bugeaud [2], of Cobham's conjecture following a completely different approach based on a deep Diophantine result known as the Schmidt subspace theorem.

Theorem 4.1 (Adamczewski and Bugeaud). *The base-b expansion of an algebraic irrational number cannot be generated by a finite automaton.*

4.1 Rational approximations and transcendence of some automatic numbers

Given an integer $k \geqslant 2$, a real number is said to be *k-automatic* if there exists an integer $b \geqslant 2$ such that its base-b expansion is a k-automatic sequence.

4.1.1 Liouville's inequality In 1844, Liouville [38] proved that transcendental numbers exist. Moreover, he constructed explicit examples of such numbers. His approach relies on the famous Liouville inequality recalled below.

Proposition 4.2 (Liouville's inequality). *Let ξ be an algebraic number of degree $d \geqslant 2$. Then there exists a positive real number c_ξ such that*

$$\left|\xi - \frac{p}{q}\right| \geqslant \frac{c_\xi}{q^d}$$

for every rational number p/q with $q \geqslant 1$.

Proof. Let P denote the minimal polynomial of ξ, let P' denote its derivative, and set

$$c_\xi := 1/(1 + \max_{|\xi - x| < 1} |P'(x)|)\,.$$

If $|\xi - p/q| \geqslant 1$, then our choice of c_ξ ensures that $|\xi - p/q| \geqslant c_\xi/q^d$.

Let us now assume that $|\xi - p/q| < 1$. Since P is the minimal polynomial of ξ, it does not vanish at p/q and $q^d P(p/q)$ is a nonzero integer. Consequently,

$$|P(p/q)| \geqslant \frac{1}{q^d}\cdot \tag{4.1}$$

Since $|\xi - p/q| < 1$, the mean value theorem implies the existence of a real number t in $(p/q - 1, p/q + 1)$ such that

$$|P(p/q)| = |P(\xi) - P(p/q)| = \left|\xi - \frac{p}{q}\right| \cdot |P'(t)|\,,$$

which ends the proof in view of Inequality (4.1) and the definition of c_ξ. □

Liouville's inequality can be used to easily construct transcendental numbers. Indeed, if ξ is an irrational real number such that for every integer $d \geqslant 2$ there exists a rational number p/q satisfying $|\xi - p/q| < q^{-d}$, then ξ is transcendental. Real numbers enjoying this property are termed *Liouville numbers*. The number $\mathcal{L}$ below is a typical example of Liouville number, often considered as the first example of a transcendental number.

Theorem 4.3 (Liouville). *The real number*

$$\mathcal{L} := \sum_{n=1}^{+\infty} \frac{1}{10^{n!}}$$

is transcendental.

Proof of Theorem 4.3. Let $j \geqslant d \geqslant 2$ be two integers. Then, there exists an integer p_j such that

$$\frac{p_j}{10^{j!}} = \sum_{n=1}^{j} \frac{1}{10^{n!}} \cdot$$

Observe that

$$\left| \mathcal{L} - \frac{p_j}{10^{j!}} \right| = \sum_{n>j} \frac{1}{10^{n!}} < \frac{2}{10^{(j+1)!}} < \frac{1}{(10^{j!})^d} \cdot$$

It then follows from Proposition 4.2 that $\mathcal{L}$ cannot be algebraic of degree less than d. Since d is arbitrary, $\mathcal{L}$ is transcendental. □

Adamczewski and Cassaigne [3] confirmed a conjecture of Shallit by proving that no Liouville number can be generated by a finite automaton. In other words, there is no automatic real number that can be proved to be transcendental by the elementary approach described above. However, we will see in the sequel how some deep improvements of Liouville's inequality can be used in a similar way to prove the transcendence of automatic numbers.

4.1.2 Roth's theorem The following famous improvement of Liouville's inequality was established by Roth [57] in 1955. This result is the best possible in the sense that the exponent $2 + \varepsilon$ in (4.2) cannot be lowered.

Theorem 4.4 (Roth). *Let ξ be a real algebraic number and let ε be a positive real number. Then there are only a finite number of rational numbers p/q such that $q \geqslant 1$ and*

$$\left| \xi - \frac{p}{q} \right| < \frac{1}{q^{2+\varepsilon}} \cdot \tag{4.2}$$

We give an immediate application of Roth's theorem to the transcendence of automatic real numbers.

Corollary 4.5. *For every integer $k \geqslant 3$, the k-automatic real number*

$$\sum_{n=1}^{+\infty} \frac{1}{10^{k^n}}$$

is transcendental.

Proof. Use the same argument as in the proof of Theorem 4.3. □

However, Roth's theorem gives no information on the arithmetical nature of the 2-automatic real number

$$\sum_{n=1}^{+\infty} \frac{1}{10^{2^n}},$$

and indeed this number has bounded partial quotients.

Let us now consider the word w defined in Example 2.2. We associate with w the real number

$$\xi_w := \sum_{n \geqslant 0} \frac{w_n}{10^{n+1}} = 0.001\,001\,010\,001\,001\,010\,001 \cdots .$$

A characteristic of the number $\mathcal{L}$ and the numbers defined in Corollary 4.5 is that large blocks of zeros appear in their decimal expansion much more frequently than one would expect if the numbers we were dealing with were randomly selected. In contrast, the decimal expansion of ξ_w contains no occurrence of more than three consecutive zeros. However, the combinatorial structure of w can be used to reveal more hidden good rational approximations to ξ_w that imply the following result.

Theorem 4.6. *The 2-automatic real number ξ_w is transcendental.*

Proof. Let ψ be the binary morphism defined in Example 2.4. For every positive integer j, set $u_j := \psi^j(0)$, $s_j := |u_j|$ and let us consider the rational number ρ_j defined by

$$\rho_j := 0.u_j^\omega .$$

An easy computation shows that there exists an integer p_j such that

$$\rho_j = \frac{p_j}{10^{s_j} - 1} \cdot \tag{4.3}$$

The rational number ρ_j turns out to be a very good approximation to ξ_w. Indeed, by definition of w, the decimal expansion of ξ_w begins with $\psi^j(0)\psi^j(0)\psi^{j-1}(0)$, which is also a prefix of u_j^ω. Consequently, the first $(2 + 1/3)s_j = 7 \cdot 3^{j-1}$ digits in the decimal expansion of ξ_w and of ρ_j are the same. We thus obtain that

$$|\xi_w - \rho_j| < 10^{-(2+1/3)s_j} . \tag{4.4}$$

Consequently, we infer from (4.4) and (4.3) that

$$\left| \xi_w - \frac{p_j}{10^{s_j} - 1} \right| < \frac{1}{(10^{s_j} - 1)^{2.3}} \cdot$$

Furthermore, the rational numbers ρ_j are all different since $\psi^n(0)$ is not a prefix of the infinite word $(\psi^m(0))^\omega$ when $n > m$. It thus follows from Roth's theorem that ξ_w is transcendental. □

4.1.3 A p-adic version of Roth's theorem The following non-Archimedean extension of Roth's theorem was proved in 1957 by Ridout [55]. For every prime number ℓ, we let $|\cdot|_\ell$ denote the ℓ-adic absolute value, which is normalized such that $|\ell|_\ell = \ell^{-1}$. Thus given an integer n, we have $|n|_\ell = \ell^{-j}$ where j denotes the largest integer for which ℓ^j divides n.

Theorem 4.7. *Let ξ be an algebraic number and ε be a positive real number. Let S be a finite set of distinct prime numbers. Then there are only a finite number of rational numbers p/q such that $q \geqslant 1$ and*

$$\left(\prod_{\ell\in S} |p|_\ell \cdot |q|_\ell\right) \cdot \left|\xi - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}}\cdot$$

We point out a first classical consequence of Ridout's theorem.

Corollary 4.8. *The real number*

$$\mathcal{K} := \sum_{n=1}^{+\infty} \frac{1}{10^{2^n}}$$

is transcendental.

Proof. Let j be a positive integer and let us consider the rational number $\rho_j := \displaystyle\sum_{n=1}^{j} 10^{-2^n}$. There exists an integer p_j such that $\rho_j = p_j/q_j$ with $q_j := 10^{2^j}$. Observe that

$$|\mathcal{K} - \rho_j| = \sum_{n>j} \frac{1}{10^{2^n}} < \frac{2}{10^{2^{j+1}}} = \frac{2}{(q_j)^2}\,,$$

and set $S := \{2, 5\}$. An easy computation gives that

$$\left(\prod_{\ell\in S} |q_j|_\ell \cdot |p_j|_\ell\right) \leqslant \prod_{\ell\in S} |q_j|_\ell = \frac{1}{q_j}$$

and thus

$$\left(\prod_{\ell\in S} |q_j|_\ell \cdot |p_j|_\ell\right) \cdot |\mathcal{K} - p_j/q_j| < \frac{2}{(q_j)^3}\cdot$$

Theorem 4.7 then implies that $\mathcal{K}$ is transcendental. □

Let us now consider the 3-automatic word y defined in Example 2.3. Let us associate with y the real number

$$\xi_y := \sum_{n\geqslant 1} \frac{y_n}{10^n} = 0.001\,010\,010\,001\,010\,001\,001\cdots\,.$$

Unfortunately, the word y does not have sufficiently large initial repetitive patterns to prove the transcendence of ξ_y by means of Roth's theorem as we did in Theorem 4.6. To

overcome this difficulty we use a trick based on Ridout's theorem that was first introduced by Ferenczi and Mauduit [22].

Theorem 4.9. *The 3-automatic real number ξ_y is transcendental.*

Proof. For every integer $j \geqslant 0$, set $u_j := \phi(\omega^j(012020))$, $v_j := \phi(\omega^j(021012))$, $r_j := |u_j|$ and $s_j := |v_j|$. Let us also consider the rational number ρ_j defined by

$$\rho_j := 0.u_j v_j^\omega \, .$$

An easy computation shows that there exists an integer p_j such that

$$\rho_j = \frac{p_j}{10^{r_j}(10^{s_j}-1)} \cdot \tag{4.5}$$

On the other hand, one can check that y begins with the word

$$\phi(\omega^j(01202002101201021012 0)) = u_j v_j v_j \phi(\omega^j(0)) \, .$$

Since $\phi(\omega^j(0))$ is a prefix of v_j, we obtain that the first $r_j + 2s_j + |\phi(\omega^j(0))| = 19 \cdot 3^j$ digits in the decimal expansion of ξ_y and of ρ_j are the same. We thus have

$$|\xi_y - \rho_j| < \frac{1}{10^{19 \cdot 3^j}} \cdot \tag{4.6}$$

Note that we obtain very special rational approximations to ξ_y: their denominators can be divided by a very large power of 10. More precisely, letting $S := \{2, 5\}$, we have

$$\prod_{\ell \in S} |10^{r_j}(10^{s_j}-1)|_\ell = \frac{1}{10^{r_j}} = \frac{1}{10^{6 \cdot 3^j}} \cdot \tag{4.7}$$

Set $q_j := 10^{r_j}(10^{s_j}-1)$. We infer from (4.5), (4.6) and (4.7) that

$$\left(\prod_{\ell \in S} |p_j|_\ell \cdot |q_j|_\ell \right) \cdot \left| \xi_y - \frac{p_j}{q_j} \right| < \frac{1}{10^{25 \cdot 3^j}} \, , \tag{4.8}$$

for every positive integer j. Since $q_j < 10^{r_j + s_j} = 10^{12 \cdot 3^j}$, we deduce from (4.8) that

$$\left(\prod_{\ell \in S} |p_j|_\ell \cdot |q_j|_\ell \right) \cdot \left| \xi_y - \frac{p_j}{q_j} \right| < \frac{1}{q_j^{2+1/12}} \, ,$$

for every integer j large enough. On the other hand, it can be shown that the word y is not eventually periodic, which implies that the set of rational numbers $\{p_j/q_j : j \geqslant 1\}$ is infinite. It thus follows from Theorem 4.7 that ξ_y is transcendental, concluding the proof. □

4.2 The Schmidt subspace theorem and a proof of Cobham's conjecture

A wonderful multidimensional generalization of Roth's theorem was obtained by Schmidt in the early 1970s (see [59]). It is now referred to as the Schmidt subspace theorem or,

for short, as the subspace theorem. We state below a heavily simplified p-adic version of this theorem. However, Theorem 4.10 turns out to be strong enough for our purpose.

Theorem 4.10. *Let $m \geqslant 2$ be an integer and ε be a positive real number. Let S be a finite set of distinct prime numbers. Let $L_1, \ldots, L_m$ be m linearly independent (over the field of algebraic numbers) linear forms with real algebraic coefficients. Then the set of solutions $\mathbf{x} = (x_1, \ldots, x_m)$ in $\mathbb{Z}^m$ to the inequality*

$$\left(\prod_{i=1}^{m}\prod_{\ell\in S}|x_i|_\ell\right)\cdot\prod_{i=1}^{m}|L_i(\mathbf{x})| \leqslant (\max\{|x_1|,\ldots,|x_m|\})^{-\varepsilon}$$

lies in finitely many proper vector subspaces of $\mathbb{Q}^m$.

Let us note that Roth's theorem easily follows from Theorem 4.10. Let $0 < \xi < 1$ be a real algebraic number and let ε be a positive real number. Consider the two independent linear forms $\xi X - Y$ and X. Choosing $S = \{\emptyset\}$, Theorem 4.10 implies that all the integer solutions (p, q) to

$$|q| \cdot |q\xi - p| < |q|^{-\varepsilon} \tag{4.9}$$

are contained in a finite union of proper vector subspaces of $\mathbb{Q}^2$. There thus is a finite set of equations $x_1X + y_1Y = 0, \ldots, x_tX + y_tY = 0$ such that, for every solution (p, q) to (4.9), there exists an integer k with $x_kp + y_kq = 0$. This means that there are only finitely many rational solutions to $|\xi - p/q| < |q|^{-2-\varepsilon}$, which immediately gives Roth's theorem.

Proof of Theorem 4.1. Let $0 < \xi < 1$ be an automatic irrational real number. Then there is an integer base $b \geqslant 2$ such that the base-b expansion of ξ is a k-automatic word for some integer $k \geqslant 2$. Let a denote the base-b expansion of ξ.

By Theorem 2.1, there exist a coding φ from an alphabet $A = \{1, 2, \ldots, r\}$ to the alphabet $\{0, 1, \ldots, b-1\}$ and a k-uniform morphism σ from A into itself such that

$$a = \varphi(u)\,,$$

where u is a fixed point of σ. By the pigeonhole principle, the prefix of length $r + 1$ of u can be written in the form $w_1cw_2cw_3$, where c is a letter and w_1, w_2, w_3 are (possibly empty) finite words.

For every integer $j \geqslant 1$, set $u_j = \varphi(\sigma^j(w_1))$, $v_j = \varphi(\sigma^j(cw_2))$ and $v'_j = \varphi(\sigma^j(c))$. Since σ is a k-uniform morphism and φ is a coding, we get that

$$|u_j| = s \cdot k^j, \quad |v_j| = t \cdot k^j \quad \text{and} \quad |v'_j| = k^j\,,$$

where $s := |u_1|$ and $t := |v_1|$. Thus the base-b expansion of ξ begins with the word $u_jv_jv'_j$, that is,

$$\xi = 0.u_jv_jv'_j\cdots\,.$$

Let ρ_j be the rational number whose base-b expansion is the infinite word $u_jv_j^\omega$, that is,

$$\rho_j = 0.u_jv_j^\omega\,.$$

A simple computation shows that there exists an integer p_j such that

$$\rho_j = \frac{p_j}{b^{s\cdot k^j}(b^{t\cdot k^j} - 1)}\,.$$

Since ρ_j and ξ have the same first $(s+t+1)\cdot k^j$ digits, we have

$$|\xi - \rho_j| < \frac{1}{b^{(s+t+1)\cdot k^j}}\,.$$

Henceforth, we assume that ξ is an algebraic number, and we will reach a contradiction. Consider the three linearly independent linear forms with real algebraic coefficients:

$$\begin{aligned} L_1(X_1,X_2,X_3) &= \xi X_1 - \xi X_2 - X_3\,,\\ L_2(X_1,X_2,X_3) &= X_1\,,\\ L_3(X_1,X_2,X_3) &= X_2\,. \end{aligned}$$

For $j \geqslant 1$, evaluating them on the integer triple

$$\mathbf{x}_j := (x_1^{(j)}, x_2^{(j)}, x_3^{(j)}) := (b^{(s+t)\cdot k^j}, b^{s\cdot k^j}, p_j)\,,$$

we obtain that

$$\prod_{i=1}^{3} |L_i(\mathbf{x}_j)| \leqslant b^{(2s+t-1)\cdot k^j}\,. \tag{4.10}$$

On the other hand, letting S be the set of prime divisors of b, we get that

$$\prod_{i=1}^{3}\prod_{\ell\in S} |x_i^{(j)}|_\ell \leqslant \prod_{\ell\in S} |b^{(s+t)\cdot k^j}|_\ell \cdot \prod_{\ell\in S} |b^{s\cdot k^j}|_\ell = b^{-(2s+t)\cdot k^j}\,. \tag{4.11}$$

Combining (4.10) and (4.11), we get that

$$\left(\prod_{i=1}^{3}\prod_{\ell\in S} |x_i^{(j)}|_\ell\right) \cdot \prod_{i=1}^{3} |L_i(\mathbf{x}_j)| \leqslant b^{-k^j}\,.$$

Set $\varepsilon = 1/(s+t)$. We thus obtain

$$\left(\prod_{i=1}^{3}\prod_{\ell\in S} |x_i^{(j)}|_\ell\right) \cdot \prod_{i=1}^{3} |L_i(\mathbf{x}_j)| \leqslant \left(\max\{b^{(s+t)\cdot k^j}, b^{s\cdot k^j}, p_j\}\right)^{-\varepsilon}\,,$$

for every positive integer j.

We then infer from Theorem 4.10 that all integer points $\mathbf{x}_j$ lie in a finite number of proper vector subspaces of $\mathbb{Q}^3$. Thus there exist a nonzero integer triple (z_1,z_2,z_3) and an infinite set of distinct positive integers $\mathcal{J}$ such that

$$z_1 b^{(s+t)\cdot k^j} + z_2 b^{s\cdot k^j} + z_3 p_j = 0\,, \tag{4.12}$$

for every j in $\mathcal{J}$. Recall that $p_j/b^{(s+t)\cdot k^j}$ tends to ξ when j tends to infinity. Dividing (4.12) by $b^{(s+t)\cdot k^j}$ and letting j tend to infinity along $\mathcal{J}$, we get that ξ is a rational number since (z_1,z_2,z_3) is a nonzero triple. This provides a contradiction. □

5 The Skolem-Mahler-Lech theorem in positive characteristic

5.1 Zeros of linear recurrences over fields of characteristic zero

The Skolem-Mahler-Lech theorem is a celebrated result which describes the set of solutions in n to the equation $a(n) = 0$, where $a(n)$ is a sequence satisfying a linear recurrence over a field of characteristic 0. We recall that if $\mathbb{K}$ is a field and $a(n)$ is a $\mathbb{K}$-valued sequence, then $a(n)$ *satisfies a linear recurrence* over $\mathbb{K}$ if there exists a natural number d and values $c_1, \ldots, c_d \in \mathbb{K}$ such that

$$a(n) = \sum_{i=1}^{d} c_i a(n-i)$$

for all sufficiently large values of n. The zero set of the linear recurrence a is defined by

$$\mathcal{Z}(a) := \{n \in \mathbb{N} \colon a(n) = 0\} \ .$$

Theorem 5.1 (Skolem-Mahler-Lech). *Let a be a linear recurrence over a field of characteristic 0. Then $\mathcal{Z}(a)$ is a union of a finite set and a finite number of arithmetic progressions.*

This theorem was first proved for linear recurrences over the rational numbers by Skolem [63]. It was next proved for linear recurrences over the algebraic numbers by Mahler [42]. The version above was proven first by Lech [35] and later by Mahler [43], [44]. This history of this theorem can be found in the book by Everest van der Poorten, Shparlinski, and Ward [21]. The techniques used by Lech to prove the Skolem-Mahler-Lech theorem are a modification of a method first used by Skolem [63]. The idea of the proof is to first note that it is no loss of generality to assume that $\mathbb{K}$ is a finitely generated extension of $\mathbb{Q}$. We can then embed $\mathbb{K}$ in a p-adic field $\mathbb{Q}_p$ for some prime p. One can then show that there exists a natural number a such that for each $i = 0, \ldots, a-1$, there is a p-adic analytic map θ_i on $\mathbb{Z}_p$ such that $\theta_i(n) = f(an+i)$ for all sufficiently large positive integers $n \in \mathbb{N}$. If $f(an+i)$ is zero for infinitely many natural numbers n, then the map θ_i has infinitely many zeros in $\mathbb{Z}_p$. Since an analytic function cannot have infinitely many zeros in a compact subset of its domain of convergence unless that function is identically zero, this implies that either $f(an+i) = 0$ for all n sufficiently large, or there are only finitely many n for which $f(an+i) = 0$, which gives the result.

There are many different proofs and extensions of the Skolem-Mahler-Lech theorem in the literature [9, 28, 68, 21]. These proofs all use p-adic methods in some way, although the result is valid in any field of characteristic 0. A well-known aspect of Theorem SML is that it is an ineffective result. Indeed, it is still an open problem whether the set $\mathcal{Z}(a)$ can always be determined for a given linear recurrence $a(n)$ defined over a field of characteristic 0 (see the discussions in [21] and [66]). In particular, it is still unknown whether the fact that $\mathcal{Z}(a)$ is empty or not is a decidable question.

5.2 Zeros of linear recurrences over fields of positive characteristic

5.2.1 Pathological examples over fields of positive characteristic It is interesting to note that the Skolem-Mahler-Lech theorem does not hold for fields $\mathbb{K}$ of positive characteristic. The simplest counter-example was given by Lech [35]. Let p be a prime and let $\mathbb{K} = \mathbb{F}_p(t)$ be the field of rational functions in one variable over $\mathbb{F}_p$. Let

$$a(n) := (1+t)^n - t^n - 1\,.$$

It is easy to check that $a(n)$ satisfies the recurrence

$$a(n) - (2+2t)a(n-1) + (1+3t+t^2)a(n-2) - (t+t^2)a(n-3) \;=\; 0$$

for $n > 3$. On the other hand, we have

$$a(p^j) = (1+t)^{p^j} - t^{p^j} - 1 = 0$$

and $a(n) \neq 0$ if n is not a power of p, and so we have

$$\mathcal{Z}(a) = \{1, p, p^2, p^3, \ldots\}\,.$$

In fact, there are even more pathological examples, which show that the correct analogue of the Skolem-Mahler-Lech theorem in positive characteristic is much more subtle. For example, consider the sequence $a(n)$ in $\mathbb{F}_2(x,y,z)$ defined by

$$a(n) := (x+y+z)^n - (x+y)^n - (x+z)^n - (y+z)^n + x^n + y^n + z^n\,.$$

We note that if V denotes the $\mathbb{K}$-vector space consisting of all $\mathbb{K}$-valued sequences and $S : V \to V$ is the "shift" linear operator that sends a sequence $a(1), a(2), \ldots$ to the sequence $0, a(1), a(2), \ldots$, then $a(n)$ satisfies a linear recurrence if and only if there is a nonzero polynomial $P(t)$ with coefficients in $\mathbb{K}$ such that when $P(S)$ is applied to the sequence $a(n)$ we obtain a sequence whose terms are eventually zero. Then one can see that the operator

$$(1-(x+y+z)S)(1-(x+y)S)(1-(y+z)S)(1-xS)(1-yS)(1-zS)$$

sends the sequence $a(n)$ to a sequence whose terms are eventually zero.

We claim that the zero set of $a(n)$ is precisely all natural numbers n of the form 2^i+2^j or of the form 2^i. To see this, observe that $a(2^i) = 0$ follows simply from the fact that $(b+c)^{2^i} = b^{2^i} + c^{2^i}$ for elements b and c in a field of characteristic 2. To check that $a(2^i+2^j) = 0$ we note that

$$\begin{aligned} G(x_1,y_1,z_1;x_2,y_2,z_2) \;&:=\; (x_1+y_1+z_1)(x_2+y_2+z_2) \\ &-\; (x_1+y_1)(x_2+y_2) - (x_1+z_1)(x_2+z_2) \\ &-\; (y_1+z_1)(y_2+z_2) + x_1x_2 + y_1y_2 + z_1z_2 \end{aligned}$$

is identically zero in every field. Notice that if $c_1, c_2, c_3 \in \mathbb{F}_2$ then

$$(c_1x + c_2y + c_3z)^{2^i+2^j} = (c_1x^{2^i} + c_2y^{2^i} + c_3z^{2^i})(c_1x^{2^j} + c_2y^{2^j} + c_3z^{2^j})\,.$$

Hence

$$a(2^i+2^j) = G(x^{2^i}, y^{2^i}, z^{2^i}; x^{2^j}, y^{2^j}, z^{2^j}) = 0\,.$$

On the other hand, if n is not a power of 2 or of the form $2^i + 2^j$, then we can write $n = 2^i + 2^j + 2^k m$ where $i > j$, $2^j > 2^k m$ and m is an odd positive integer. Note that

$$\begin{aligned}(x+y+z)^n &= (x+y+z)^{2^i}(x+y+z)^{2^j}\left((x+y+z)^{2^k}\right)^m \\ &= (x^{2^i}+y^{2^i}+z^{2^i})(x^{2^j}+y^{2^j}+z^{2^j})(x^{2^k}+y^{2^k}+z^{2^k})^m\,.\end{aligned}$$

Consider the coefficient of $x^{2^i}y^{2^j}z^{2^k m}$ in $(x+y+z)^n$. The only way to get this term is to take x^{2^i} from the first term in the product, y^{2^j} from the second term, and $z^{2^k m}$ from the third term. Hence the coefficient is 1. Since $(x+y+z)^n$ is the only term in $a(n)$ that has monomials of the form $x^b y^c z^d$ with $b, c, d > 0$ appearing, we see that $a(n)$ is nonzero if the binary expansion of n has more than two 1's.

5.2.2 Derksen's theorem We now give a remarkable result due to Derksen [17]. We have seen that the zero set of a linear recurrence in a field of characteristic $p > 0$ is often more pathological than in characteristic zero. At the same time, in our pathological examples, the base-p expansion of a number n gives insight into whether the nth term of our linearly recurrent sequence vanishes. In fact, Derksen [17] shows that the zero set of a linearly recurrent sequence can always be described in terms of automata.

Theorem 5.2 (Derksen). *Let a be a linear recurrence over a field of characteristic p. Then the set $\mathcal{Z}(a)$ is a p-automatic set.*

Derksen gives a further refinement of this result; however the main ingredient of his proof is the fact that the zero set is p-automatic. Furthermore, each step in Derksen's proof can be made effective!

We prove an extension of Derksen's result for algebraic power series in several variables in the next section. To explain the connection between Derksen's result and power series, we recall the following classical result.

Proposition 5.3. *Let $\mathbb{K}$ be a field and let $a(n)$ be a $\mathbb{K}$-valued sequence. The following conditions are equivalent.*

(i) *The sequence $a(n)$ satisfies a linear recurrence over $\mathbb{K}$.*
(ii) *There is a natural number d, a matrix $A \in M_d(\mathbb{K})$, and vectors v and w in $\mathbb{K}^d$ such that $a(n) = w^T A^n v$.*
(iii) *$\sum_{n\geqslant 0} a(n)t^n$ is the power series expansion of a rational function in $\mathbb{K}(t)$.*

Proof. $(i) \implies (ii)$. Suppose that $a(n)$ satisfies a linear recurrence

$$a(n) := \sum_{j=1}^{d} c_j a(n-j)$$

for all $n \geqslant d$. We let

$$v(i) := [a(i)\ a(i+1)\ \cdots\ a(i+d-1)]^T$$

and

$$w := [1\ 0\ 0\ \cdots\ 0]^T\,.$$

Finally, we let

$$A := \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ldots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ c_d & c_{d-1} & c_{d-2} & c_{d-3} & \cdots & c_1 \end{pmatrix}.$$

Then one easily sees that $v(i+1) = Av(i)$ and so $w^T A^n v = a(n)$, where $v = v(0)$. Thus (i) implies (ii).

$(ii) \implies (iii)$. Set

$$f(t) := \sum_{n=0}^{\infty} (w^T A^n v) t^n .$$

By the Cayley-Hamilton theorem, A satisfies a polynomial $A^d + \sum_{j=0}^{d-1} c_j A^j = 0$ and hence

$$w^T A^{n+d} v + \sum_{j=0}^{d-1} c_j w^T A^{j+n} v = 0$$

for all n. It follows that $f(t)(1 + \sum_{j=0}^{d-1} c_j t^{d-j})$ is a polynomial in t and so $f(t)$ is the power series expansion of a rational function.

$(iii) \implies (i)$. Suppose that $f(t) = \sum_{n=0}^{\infty} a(n) t^n$ is the power series expansion of a rational function $P(t)/Q(t)$ with $P(t)$ and $Q(t)$ polynomials and $Q(t)$ nonzero. We may assume that $Q(0) = 1$. We write $Q(t) = 1 + \sum_{j=1}^{d} c_j t^j$. Then $P(t) = f(t)Q(t)$ and so $a(n) + \sum_{j=1}^{d} c_j a(n-j) = 0$ for all n larger than the degree of $P(t)$. It follows that $a(n)$ satisfies a linear recurrence. □

5.3 Vanishing coefficients of algebraic power series

In light of Proposition 5.3, we may interpret Derksen's result as a statement about the zero coefficients of the power series expansion of a rational power series over a field of characteristic $p > 0$. In this section, we show that this interpretation gives rise to a far-reaching generalization of Derksen's result.

We first note that rational power series are a subset of *algebraic power series* (choosing $m = 1$ in the definition below).

Definition 5.1. Let $\mathbb{K}$ be a field. We say that a power series

$$f(t) = \sum_{n=0}^{\infty} a(n) t^n \in \mathbb{K}[[t]]$$

is *algebraic* if it is algebraic over the field of rational functions $\mathbb{K}(t)$, that is, if there exists a natural number m and polynomials $A_0(t), \ldots, A_m(t) \in \mathbb{K}[t]$, with $A_m(t)$ nonzero, such that

$$\sum_{j=0}^{m} A_j(t) f(t)^j = 0 .$$

More generally, we say that $f(t_1,\dots,t_d) \in \mathbb{K}[[t_1,\dots,t_d]]$ is *algebraic* if there exist polynomials $A_0,\dots,A_m \in \mathbb{K}[t_1,\dots,t_d]$, not all zero, such that

$$\sum_{j=0}^{m} A_j(t_1,\dots,t_d) f(t_1,\dots,t_d)^j \;=\; 0\,.$$

Given a multivariate power series $f(t_1,\dots,t_d) = \sum_{n_1,\dots,n_d} a_{n_1,\dots,n_d} t_1^{n_1}\cdots t_d^{n_d} \in \mathbb{K}[[t_1,\dots,t_d]]$, we let $\mathcal{Z}(f)$ denote the set of vanishing coefficients, that is,

$$\mathcal{Z}(f) = \{(n_1,\dots,n_d) \in \mathbb{N}^d \colon a_{n_1,\dots,n_d} = 0\}\,.$$

It is interesting to note that the Skolem-Mahler-Lech theorem in characteristic 0 has no analogue for multivariate rational functions. For instance,

$$f(t_1,t_2) = \sum_{m,n} (2^m - n^2) t_1^m t_2^n$$

is a bivariate rational power series in $\mathbb{Q}[[t_1,t_2]]$ with

$$\mathcal{Z}(f) = \{(m,n) \;:\; m \equiv 0 \pmod 2, n = 2^{m/2}\}\,.$$

Thus we cannot expect the zero set to be given in terms of arithmetic progressions or even in terms of finite automata.

To see some of the complexities that can occur in the multivariate case, consider the power series

$$f(t_1,t_2) = \sum_{m,n\geqslant 0} (3^m - 2^n - 1) t_1^m t_2^n\,.$$

We see that

$$f(t_1,t_2) = (1-3t_1)^{-1}(1-t_2)^{-1} - (1-t_1)^{-1}(1-2t_2)^{-1} - (1-t_1)^{-1}(1-t_2)^{-1},$$

and so it is a rational power series. On the other hand, the coefficient of $t_1^m t_2^n$ is zero if and only if $3^m = 2^n + 1$. It is now known that this occurs only when (m,n) is $(2,3)$ or $(1,1)$, due to Mihăilescu's solution to Catalan's conjecture [48]. In general, finding the zero set often involves difficult diophantine problems.

Remarkably, in positive characteristic an analogue of Derksen's result holds for multivariate rational power series, as shown in [1]—in fact it even holds for multivariate algebraic power series! In the sequel of this chapter, we will use $\mathbf{n}$ and $\mathbf{j}$ to represent, respectively, the d-tuple of natural numbers $(n_1,\dots,n_d)$ and $(j_1,\dots,j_d)$. We will also let $\mathbf{t}^{\mathbf{n}}$ denote the monomial $t_1^{n_1}\cdots t_d^{n_d}$.

Theorem 5.4 (Adamczewski and Bell). *Let $\mathbb{K}$ be a field of characteristic $p > 0$ and let $f(\mathbf{t}) \in \mathbb{K}[[\mathbf{t}]]$ be the power series expansion of an algebraic function over $\mathbb{K}(\mathbf{t})$. Then $\mathcal{Z}(f)$ is a p-automatic subset of $\mathbb{N}^d$.*

We note that this immediately implies Theorem 5.2 by taking $d = 1$ and taking $f(t)$ to be a rational function. On the other hand, by taking $\mathbb{K}$ to be a finite field, Theorem 5.4 reduces to the difficult part of the multivariate version of Christol's theorem (see Theorem 6.2). As with Derksen's proof, it seems that Theorem 5.4 can be made effective.

Furthermore, given any p-automatic set $\mathcal{N}$ in $\mathbb{N}^d$, $\mathcal{N}$ is the zero set of the power series $\sum_{\mathbf{n}\notin\mathcal{N}} \mathbf{t}^{\mathbf{n}} \in \mathbb{F}_p((\mathbf{t}))$ which is known to be algebraic over $\mathbb{F}_p(\mathbf{t})$ by Theorem 6.2. At this level of generality, we thus have a nice correspondence between p-automatic sets and the zero set of algebraic multivariate functions over fields of characteristic p.

5.3.1 Proof of Theorem 5.4 In order to prove this result we need to introduce some notation.

Let p be a prime number and let d be a natural number. For each $\mathbf{j} = (j_1, \ldots, j_d) \in \{0, 1, \ldots, p-1\}^d$, we define $e_{\mathbf{j}} : \mathbb{N}^d \to \mathbb{N}^d$ by

$$e_{\mathbf{j}}(n_1, \ldots, n_d) := (pn_1 + j_1, \ldots, pn_d + j_d)\,. \tag{5.1}$$

We let Σ denote the semigroup generated by the collection of all $e_{\mathbf{j}}$ under composition.

Remark 5.5. Note that if Δ is a finite set, then $f : \mathbb{N}^d \to \Delta$ is p-automatic if and only the set of functions $\{f \circ e : \ e \in \Sigma\}$ is a finite set.

We also recall that a field $\mathbb{K}$ of characteristic $p > 0$ is *perfect* if the map $x \mapsto x^p$ is surjective on $\mathbb{K}$. Let p be a prime number and let $\mathbb{K}$ be a perfect field of characteristic p. For a power series $f(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} a(\mathbf{n})\mathbf{t}^{\mathbf{n}} \in \mathbb{K}[[\mathbf{t}]]$, we define

$$E_{\mathbf{j}}(f(\mathbf{t})) \ := \sum_{\mathbf{n}\in\mathbb{N}^d} (a \circ e_{\mathbf{j}}(\mathbf{n}))^{1/p}\mathbf{t}^{\mathbf{n}} \tag{5.2}$$

for $\mathbf{j} \in \{0, 1, \ldots, p-1\}^d$. We let Ω denote the semigroup generated by the collection of $E_{\mathbf{j}}$ under composition. We let $\Omega(f)$ denote the $\mathbb{K}$-vector space spanned by all power series of the form $E \circ f$ with $E \in \Omega$. We note that if $g \in \Omega(F)$ then $E \circ g \in \Omega(f)$ for all $E \in \Omega$.

A theorem of Sharif and Woodcock [62] gives a concrete characterization of the algebraic power series over a perfect field of positive characteristic.

Theorem 5.6 (Sharif and Woodcock). *Let p be a prime number and let $\mathbb{K}$ be a perfect field of characteristic p. A power series $f(\mathbf{t}) \in \mathbb{K}[[\mathbf{t}]]$ is algebraic if and only if $\Omega(f)$ is a finite-dimensional $\mathbb{K}$-vector space.*

One can rephrase the theorem of Sharif and Woodcock in terms of the coefficients of an algebraic power series.

Lemma 5.7. *Let p be a prime number, let $\mathbb{K}$ be a perfect field of characteristic p, and let $a : \mathbb{N}^d \to \mathbb{K}$ be a sequence with the property that*

$$f(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} a(\mathbf{n})\mathbf{t}^{\mathbf{n}} \in \mathbb{K}[[\mathbf{t}]]$$

is a nonzero algebraic function over $\mathbb{K}(\mathbf{t})$. Then there exists a natural number m and there exist maps $a_1, \ldots, a_m : \mathbb{N}^d \to \mathbb{K}$ with the following properties.

(i) *The formal power series $f_i(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} a_i(\mathbf{n})\mathbf{t}^{\mathbf{n}}$, $1 \leqslant i \leqslant m$, form a basis of $\Omega(f)$ as a $\mathbb{K}$-vector space.*

(ii) $f_1 = f$.
(iii) *If* $b : \mathbb{N}^d \to \mathbb{K}$ *has the property that* $g(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} b(\mathbf{n})\mathbf{t}^{\mathbf{n}}$ *belongs to* $\Omega(f)$*, then* $b \circ e_{\mathbf{j}} \in \mathbb{K}\, a_1^p + \cdots + \mathbb{K}\, a_m^p$ *for every* $\mathbf{j} \in \{0, \dots, p-1\}^d$.

Proof. Since $f(\mathbf{t})$ is algebraic, $\dim_{\mathbb{K}}(\Omega(f))$ is finite by Theorem 5.6. We can thus pick maps $a_1, \dots, a_m : \mathbb{N}^d \to \mathbb{K}$ such that the m power series $f_i(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} a_i(\mathbf{n})\mathbf{t}^{\mathbf{n}}$ form a basis of $\Omega(f)$, and with $f_1 = f$. Let $b : \mathbb{N}^d \to \mathbb{K}$ be such that $g(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} b(\mathbf{n})\mathbf{t}^{\mathbf{n}}$ belongs to $\Omega(f)$. Observe that the power series g can be decomposed as

$$g(\mathbf{t}) = \sum_{\mathbf{j}\in\{0,\dots,p-1\}^d} \mathbf{t}^{\mathbf{j}} E_{\mathbf{j}}(g(\mathbf{t}))^p \,. \tag{5.3}$$

By assumption, $E_{\mathbf{j}}(g(\mathbf{t})) \in \mathbb{K}\, f_1(\mathbf{t}) + \cdots + \mathbb{K}\, f_m(\mathbf{t})$ and hence $E_{\mathbf{j}}(g(\mathbf{t}))^p \in \mathbb{K}\, f_1(\mathbf{t})^p + \cdots + \mathbb{K}\, f_m(\mathbf{t})^p$. Let $\mathbf{j} \in \{0, 1, \dots, p-1\}^d$. Considering the coefficient of $\mathbf{t}^{p\mathbf{n}+\mathbf{j}}$ in Equation (5.3), we see that $b \circ e_{\mathbf{j}}(\mathbf{n})$ is equal to the coefficient of $\mathbf{t}^{p\mathbf{n}}$ in $E_{\mathbf{j}}(g(\mathbf{t}))^p$, which belongs to $\mathbb{K}\, a_1(\mathbf{n})^p + \cdots + \mathbb{K}\, a_m(\mathbf{n})^p$. □

Before proving Theorem 5.4, we first fix a few notions. Given a finitely generated field extension $\mathbb{K}_0$ of $\mathbb{F}_p$, we let $\mathbb{K}_0^{\langle p\rangle}$ denote the subfield consisting of all elements of the form x^p with $x \in \mathbb{K}_0$. Given $\mathbb{F}_p$-vector subspaces V and W of $\mathbb{K}_0$ we let VW denote the $\mathbb{F}_p$-subspace of $\mathbb{K}_0$ spanned by all products of the form vw with $v \in V, w \in W$. We let $V^{\langle p\rangle}$ denote the $\mathbb{F}_p$-vector subspace consisting of all elements of the form v^p with $v \in V$. We note that since $\mathbb{K}_0$ is a finitely generated field extension of $\mathbb{F}_p$, $\mathbb{K}_0$ is a finite-dimensional $\mathbb{K}_0^{\langle p\rangle}$-vector space. If we fix a basis

$$\mathbb{K}_0 = \bigoplus_{i=1}^{r} \mathbb{K}_0^{\langle p\rangle} h_i$$

then we have *projections* $\pi_1, \dots, \pi_r : \mathbb{K}_0 \to \mathbb{K}_0$ defined by

$$x = \sum_{i=1}^{r} \pi_i(x)^p h_i \,. \tag{5.4}$$

Remark 5.8. For $1 \leqslant i \leqslant r$ and $a, b, c \in \mathbb{K}_0$ we have

$$\pi_i(c^p a + b) = c\pi_i(a) + \pi_i(b) \,.$$

The last ingredient of the proof is a technical (but very useful) result due to Derksen, which we state here without proof.

Proposition 5.9 (Derksen). *Let* $\mathbb{K}_0$ *be a finitely generated field extension of* $\mathbb{F}_p$ *and let* $\pi_1, \dots, \pi_r : \mathbb{K}_0 \to \mathbb{K}_0$ *be as in Equation (5.4). Let* V *be a finite-dimensional* $\mathbb{F}_p$*-vector subspace of* $\mathbb{K}_0$*. Then there exists a finite-dimensional* $\mathbb{F}_p$*-vector subspace* W *of* $\mathbb{K}_0$ *containing* V *such that* $\pi_i(WV) \subseteq W$ *for* $1 \leqslant i \leqslant r$.

Proof of Theorem 5.4. By enlarging $\mathbb{K}$ if necessary, we may assume that $\mathbb{K}$ is perfect. By Lemma 5.7 we can find maps $a_1, \dots, a_m : \mathbb{N}^d \to \mathbb{K}$ with the following properties.

(1) the power series $f_i(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} a_i(\mathbf{n})\mathbf{t}^{\mathbf{n}}$, $1 \leqslant i \leqslant m$, form a basis for $\Omega(f)$.

(2) $f_1 = f$.
(3) If $b : \mathbb{N}^d \to \mathbb{K}$ has the property that $g(\mathbf{t}) := \sum_{\mathbf{n}\in\mathbb{N}^d} b(\mathbf{n})\mathbf{t}^{\mathbf{n}}$ belongs to $\Omega(f)$, then $b \circ e_{\mathbf{j}} \in \mathbb{K}\, a_1^p + \cdots + \mathbb{K}\, a_m^p$ for every $\mathbf{j} \in \{0, \ldots, p-1\}^d$.

In particular, given $1 \leqslant i \leqslant m$ and $\mathbf{j} \in \{0, 1, \ldots, p-1\}^d$, there are elements $\lambda(i, \mathbf{j}, k)$, $1 \leqslant k \leqslant m$, such that

$$a_i \circ e_{\mathbf{j}} = \sum_{k=1}^{m} \lambda(i, \mathbf{j}, k) a_k^p \,. \tag{5.5}$$

Since $f_1, \ldots, f_m$ are algebraic power series, there exists a finitely generated field extension of $\mathbb{F}_p$ such that all coefficients of $f_1, \ldots, f_m$ are contained in this field extension. It follows that the subfield $\mathbb{K}_0$ of $\mathbb{K}$ generated by the coefficients of $f_1(\mathbf{t}), \ldots, f_m(\mathbf{t})$ and all the elements $\lambda(i, \mathbf{j}, k)$ is a finitely generated field extension of $\mathbb{F}_p$.

Since $\mathbb{K}_0$ is a finite-dimensional $\mathbb{K}_0^{\langle p\rangle}$-vector space, we can fix a basis $\{h_1, \ldots, h_r\}$ of $\mathbb{K}_0$, that is,

$$\mathbb{K}_0 = \bigoplus_{i=1}^{r} \mathbb{K}_0^{\langle p\rangle} h_i \,.$$

Then we have *projections* $\pi_1, \ldots, \pi_r : \mathbb{K}_0 \to \mathbb{K}_0$ defined by

$$c = \sum_{i=1}^{r} \pi_i(c)^p h_i \,. \tag{5.6}$$

We let V denote the finite-dimensional $\mathbb{F}_p$-vector subspace of $\mathbb{K}_0$ spanned by the elements $\lambda(i, \mathbf{j}, k)$, $1 \leqslant i, k \leqslant m$ and $\mathbf{j} \in \{0, 1, \ldots, p-1\}^d$, and by 1. By Equation (5.5), we have

$$a_i \circ e_{\mathbf{j}} \in \sum_{k=1}^{m} V a_k^p \,, \tag{5.7}$$

for $1 \leqslant i \leqslant m$ and $\mathbf{j} \in \{0, 1, \ldots, p-1\}^d$. By Proposition 5.9 there exists a finite-dimensional $\mathbb{F}_p$-vector subspace W of $\mathbb{K}_0$ containing V such that $\pi_i(WV) \subseteq W$ for $1 \leqslant i \leqslant r$. Set

$$U := Wa_1 + \cdots + Wa_m \subseteq \{b : \ b : \mathbb{N}^d \to \mathbb{K}_0\} \,.$$

We note that $\operatorname{Card} U \leqslant (\operatorname{Card} W)^m < \infty$. Note also that if $\ell \in \{1, \ldots, r\}$, $i \in \{1, \ldots, m\}$, and $j \in \{0, 1, \ldots, p-1\}^d$ then by Equation (5.7) and Remark 5.8 we have

$$\pi_\ell(Wa_i \circ e_{\mathbf{j}}) \subseteq \pi_\ell(WVa_1^p + \cdots + WVa_m^p) \subseteq \sum_{k=1}^{m} \pi_\ell(WV) a_k$$

$$\subseteq \sum_{k=1}^{m} Wa_k = U \,.$$

Thus by Remark 5.8, if $b \in U$ and $\mathbf{j} \in \{0, 1, \ldots, p-1\}^d$, then $b_\ell := \pi_\ell(b \circ e_{\mathbf{j}}) \in U$ for $1 \leqslant \ell \leqslant r$. In particular, $b(p\mathbf{n} + \mathbf{j}) = 0$ if and only if $b_1(\mathbf{n}) = b_2(\mathbf{n}) = \cdots = b_r(\mathbf{n}) = 0$. Given $b : \mathbb{N}^d \to \mathbb{K}_0$, we let $\chi_b : \mathbb{N}^d \to \{0, 1\}$ be defined by

$$\chi_b(\mathbf{n}) \ = \ \begin{cases} 0, \text{ if } b(\mathbf{n}) \neq 0 \\ 1, \text{ if } \ b(\mathbf{n}) = 0 \,. \end{cases}$$

Set

$$\mathcal{S} := \{\chi_{b_1} \cdots \chi_{b_t} \ : \ t \geqslant 0, b_1, \ldots, b_t \in U\}.$$

We note that since $\chi_b^2 = \chi_b$ for all $b \in U$ and U is finite, $\mathcal{S}$ is finite. Note that if $b \in U$ and $\mathbf{j} \in \{0, 1, \ldots, p-1\}^d$, then $b_\ell := \pi_\ell(b \circ e_{\mathbf{j}}) \in U$ for $1 \leqslant \ell \leqslant r$. By the above remarks,

$$(\chi_b \circ e_{\mathbf{j}})(\mathbf{n}) = \prod_{\ell=1}^{r} \chi_{b_\ell}(\mathbf{n}),$$

and so we see that if $\chi \in \mathcal{S}$ then $\chi \circ e \in \mathcal{S}$ for all $e \in \Sigma$. Since $\mathcal{S}$ is finite, this proves that $\chi : \mathbb{N}^d \to \{0, 1\}$ is p-automatic. In particular, since $a(\mathbf{n}) = a_1(\mathbf{n}) \in U$, we obtain that χ_a is p-automatic. In other words, the set of $\mathbf{n} \in \mathbb{N}^d$ such that $a(\mathbf{n}) = 0$ is a p-automatic set. This ends the proof. □

6 The algebraic closure of $\mathbb{F}_p(t)$

6.1 Christol's theorem

One of the most beautiful results in the theory of automatic sequences is Christol's theorem, which characterizes those Laurent series with coefficients in a finite field that are algebraic over the field of rational functions.

Theorem 6.1 (Christol). *Let $\mathbb{K}$ be a finite field of characteristic $p > 0$. Then $f(t) = \sum_{n \geqslant 0} a(n) t^n \in \mathbb{K}((t))$ is algebraic over $\mathbb{K}(t)$ if and only if the sequence $a(n)$ is p-automatic.*

Christol's theorem consists of two parts: the "easy" direction in which one shows that if the sequence of coefficients of a Laurent series is p-automatic, then the Laurent series is algebraic, and the "hard" direction in which one must show that the coefficients of an algebraic Laurent series form a p-automatic sequence. The hard direction is generally proved using Ore's lemma, which is the observation that if $f(t)$ is algebraic over a field $\mathbb{K}(t)$, then the set $\{f, f^p, f^{p^2}, \ldots\}$ is linearly dependent over $\mathbb{K}(t)$. Christol's theorem was generalized to multivariate Laurent series by Salon [58].

Theorem 6.2 (Salon). *Let $\mathbb{K}$ be a finite field of characteristic $p > 0$. Then $f(\mathbf{t}) = \sum_{\mathbf{n} \in \mathbb{N}^d} a(\mathbf{n}) \mathbf{t}^{\mathbf{n}} \in \mathbb{K}((\mathbf{t}))$ is algebraic if and only if the sequence $a(\mathbf{n})$ is p-automatic.*

Salon's theorem turns out to be a special case of Theorems 5.6 and 5.4.

Proof of Theorem 6.2. We suppose first that $a : \mathbb{N}^d \to \mathbb{K}$ is p-automatic and we consider the power series

$$f(\mathbf{t}) := \sum_{\mathbf{n} \in \mathbb{N}^d} a(\mathbf{n}) \mathbf{t}^{\mathbf{n}}.$$

Using the notation of Equations 5.1 and 5.2, we infer from Remark 5.5 that there are only finitely many distinct functions of the form $a \circ e$ where e runs over Σ. Consequently, there

are only finitely many functions of the form $E \circ f$ where E runs over Ω. Thus $\Omega(f)$ is finite-dimensional and Theorem 5.6 implies that $f(\mathbf{t})$ is algebraic.

We next suppose that $f(\mathbf{t})$ is algebraic and let $c \in \mathbb{K}$. Since $f(\mathbf{t})$ is algebraic, then so is $f(\mathbf{t}) - c$ and by Theorem 5.4 the set S_c of d-tuples of natural numbers $\mathbf{n}$ such that $a(\mathbf{n}) - c = 0$ is p-automatic. It follows that the sequence $a_c : \mathbb{N}^d \to \mathbb{K}$ given by $a_c(\mathbf{n}) = 1$ if $\mathbf{n} \in S_c$ and $a_c(\mathbf{n}) = 0$ otherwise is p-automatic. Thus $a(\mathbf{n}) = \sum_{c \in \mathbb{K}} c a_c(\mathbf{n})$ is also p-automatic, as p-automatic sequences taking values in a field are closed under the taking of finite sums and scalar multiplication. □

While Christol's theorem gives a concrete description of the elements of $\mathbb{F}_q((t))$ that are algebraic over $\mathbb{F}_q(t)$, it does not give the whole picture. As Kedlaya [32] points out, the field $\mathbb{F}_q((t))$ is far from being algebraically closed. Indeed, for an algebraically closed field $\mathbb{K}$ of characteristic 0, a classical result of Puiseux is that the field

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

is itself algebraically closed and contains, in particular, the algebraic closure of $\mathbb{K}(t)$. However, over field of positive characteristic, the situation is more subtle. In particular, the algebraic closure of $\mathbb{F}_q((t))$ is much more complicated to describe, due to the existence of wildly ramified field extensions. For instance, Chevalley remarked [13] that the Artin-Schreier polynomial $x^p - x - 1/t$ does not split in the Puiseux field $\bigcup_{n=1}^{+\infty} \mathbb{F}_q((t^{1/n}))$.

6.2 Generalized power series

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 [27]. We briefly describe this construction.

We recall that a subset S of a totally ordered group is said to be *well-ordered* if every nonempty subset of S has a minimal element or, equivalently, if there is no infinite decreasing sequence within S. Given a commutative ring R and a totally ordered Abelian group G we construct a commutative ring, which we denote $R((t^G))$, which is defined to be the collection of all elements of the form

$$f(t) := \sum_{\alpha \in G} r_\alpha t^\alpha$$

which satisfy the following conditions.

(i) $r_\alpha \in R$ for all $\alpha \in G$.

(ii) The support of $f(t)$ is well ordered, that is, the subset $\{\alpha \colon r_\alpha \neq 0_R\}$ is a well-ordered set.

Addition and multiplication are defined via the rules

$$\sum_{\alpha \in G} r_\alpha t^\alpha + \sum_{\alpha \in G} s_\alpha t^\alpha = \sum_{\alpha \in G} (r_\alpha + s_\alpha) t^\alpha$$

and

$$\left(\sum_{\alpha\in G} r_\alpha t^\alpha\right)\left(\sum_{\alpha\in G} s_\alpha t^\alpha\right) \;=\; \sum_{\alpha\in G}\sum_{\beta\in G}(r_\beta s_{\alpha-\beta})t^\alpha\,.$$

We note that the fact that the supports of valid series expansions are well-ordered means that no problems with possible infinite sums appearing in the expression for the coefficients in a product of two generalized power series will occur. We call the ring $R((t^G))$ the *ring of generalized power series over R with exponents in G*.

We recall that a group is *divisible* if for every $g \in G$ and $n \geqslant 1$, there exists some $h \in G$ such that $h^n = g$. For an algebraically closed field $\mathbb{K}$ and a totally ordered divisible Abelian group G, the field $\mathbb{K}((t^G))$ is known to be algebraically closed [31] (see also [32, 61]). In what follows, we will only consider the particular case of the divisible group $\mathbb{Q}$ and of a finite field $\mathbb{F}_q$ (q being a power of a prime p).

We then have the series of containments

$$\mathbb{F}_q(t) \;\subset\; \mathbb{F}_q((t)) \subset\; \mathbb{F}_q((t^{\mathbb{Q}}))\,.$$

Though $\mathbb{F}_q((t^{\mathbb{Q}}))$ is not algebraically closed, it is sufficient for our purpose to consider such fields. Indeed, taking $\bigcup_{n\geqslant 1}\mathbb{F}_{p^n}$ as an algebraic closure of $\mathbb{F}_p$, it follows from the remark above that the field $\left(\bigcup_{n\geqslant 1}\mathbb{F}_{p^n}\right)((t^{\mathbb{Q}}))$ is algebraically closed. For example, the Artin-Schreier polynomial $x^p - x - 1/t$ does split in $\mathbb{F}_p((t^{\mathbb{Q}}))$. Indeed, we can check that the generalized power series

$$c + \sum_{i=1}^{\infty} t^{-1/p^i},\; c \in \mathbb{F}_p\,,$$

are the roots of this polynomial.

6.3 Kedlaya's theorem

Kedlaya [32] considered whether one can, as in Christol's theorem, give an automaton-theoretic characterization of the elements of $\mathbb{F}_q((t^{\mathbb{Q}}))$ that are algebraic over $\mathbb{F}_q(t)$. The work of Kedlaya [33] is thus precisely devoted to a description of the algebraic closure of $\mathbb{F}_p(t)$ as a subfield of generalized power series. For this purpose, Kedlaya introduces the notion of a p-quasi-automatic function over the rational numbers.

Kedlaya uses automata to produce power series whose exponents take values in the rational numbers. Hence it is necessary to create automata which accept rational numbers as opposed to just accepting integers. We now explain how Kedlaya does this.

Let $k > 1$ be a positive integer. We set

$$\Sigma_k' \;=\; \{0, 1, \ldots, k-1, \bullet\}$$

and we let $\mathcal{L}(k)$ denote the language on the alphabet Σ_k' consisting of all words on Σ_k' with exactly one occurrence of the letter '$\bullet$' (the radix point) and whose first and last letters are not equal to 0. This is a regular language [33, Lemma 2.3.3]. We let S_k denote

the set of nonnegative k-adic rationals, that is,

$$S_k \;=\; \{a/k^b \,:\, a, b \in \mathbb{Z},\; a \geqslant 0\}\,.$$

We note that there is a bijection $[\,\cdot\,]_k : \mathcal{L}(k) \to S_k$ given by

$$s_1 \cdots s_{i-1} \bullet s_{i+1} \cdots s_n \in \mathcal{L}(k) \;\mapsto\; \sum_{j=1}^{i-1} s_j k^{i-1-j} + \sum_{j=i+1}^{n} s_j k^{i-j}\,,$$

where $s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n \in \{0, 1, \ldots, k-1\}$. So, for example, we have $[110\bullet 32]_4 = [20\bullet 875]_{10} = 167/8$. We also note that the fact that we exclude strings whose initial and terminal letters are 0 means that we have the awkward looking expression $[\,\bullet\,]_k = 0$.

Definition 6.1. We say that a map $h : S_k \to \Delta$ is *k-automatic* if there is a finite state machine which takes words on Σ_k' as input such that for each $W \in \mathcal{L}_k$, $h([W]_k)$ is generated by the machine using the word W as input.

Since the support of a generalized power series is well-ordered, we need a more general notion of automatic functions defined over the set of rationals. For this purpose, we always implicitly consider sets Δ containing a special element called zero, which we let 0 denote (of course, when Δ is a subset of $\mathbb{R}$ or $\mathbb{N}$, or if it denotes a finite field, zero will preserve its usual meaning). Then we will talk about functions $h : \mathbb{Q} \to \Delta$ as being k-automatic if their support is contained in S_k and the restriction of h to S_k is k-automatic (the support of such a function being defined as the set $S = \{\alpha \in \mathbb{Q} \colon h(\alpha) \neq 0\}$).

Example 6.1. For $w \in \mathcal{L}(2)$, define

$$h([w]_2) \;=\; \begin{cases} 0, \text{if there are an even number of } 1\text{'s in } w; \\ 1, \text{otherwise.} \end{cases}$$

Then $h : S_2 \to \{0, 1\}$ is K_2-automatic.

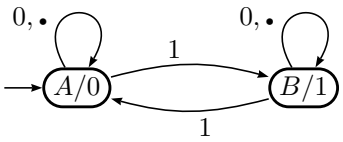

Figure 7. The DFAO associated with the function h of Example 6.1.

Definition 6.2. Let k be a positive integer, let Δ be a finite set containing a special element 0, and let $h : \mathbb{Q} \to \Delta$. We say that h is k-quasi-automatic if it satisfies the following conditions.

(i) The support S of h is well-ordered.

(ii) There exist a positive integer a and an integer b such that the set $aS + b$ consists of nonnegative k-adic rationals and the map $h((x - b)/a)$ is a k-automatic function from S_k to Δ.

We are now ready to state Kedlaya's theorem.

Theorem 6.3 (Kedlaya). *Let p be a prime, let q be a power of p, and let $a : \mathbb{Q} \to \mathbb{F}_q$. Then $\sum_{\alpha\in\mathbb{Q}} a(\alpha)t^\alpha$ is algebraic over $\mathbb{F}_q(t)$ if and only if the function $a : \mathbb{Q} \to \mathbb{F}_q$ is p-quasi-automatic.*

In light of Salon's result [58], Kedlaya asked whether his theorem has an extension to multivariate generalized power series $\mathbb{F}_q((t_1^{\mathbb{Q}}, \dots, t_m^{\mathbb{Q}}))$. As far as we know, this problem has not yet been solved.

7 Update

Since the writing of this chapter in 2010 there has been additional work related to the topics we just discussed. Here we point out a few such references. Concerning Section 3, we mention the papers of Mauduit and Rivat [47], Martin, Mauduit, and Rivat [45], and Müllner [51]. Concerning Section 4, we mention an extension of Theorem 4.1 to deterministic pushdown automata due to Adamczewski, Cassaigne and Le Gonidec [4]. Also, a new proof of Theorem 4.1 and some generalizations have been obtained recently by using the so-called Mahler method (see Philippon [52], Adamczewski and Faverjon [5, 6]). Concerning Section 5, we mention the work of Derksen and Masser [18, 19], Leitnik [36, 37], and Bell and Moosa [8]. Concerning Section 6, we mention the papers of Kedlaya [34] and Bridy [11].

References

[1] B. Adamczewski and J. Bell. On vanishing coefficients of algebraic power series over fields of positive characteristic. *Invent. Math.*, 187:343–393, 2012. 90

[2] B. Adamczewski and Y. Bugeaud. On the complexity of algebraic numbers. I. Expansions in integer bases. *Ann. of Math. (2)*, 165:547–565, 2007. 79

[3] B. Adamczewski and J. Cassaigne. Diophantine properties of real numbers generated by finite automata. *Compositio Math.*, 142(6):1351–1372, 2006. 80

[4] B. Adamczewski, J. Cassaigne, and M. Le Gonidec. On the computational complexity of algebraic numbers: the Hartmanis–Stearns problem revisited. Technical report, arXiv:1601.02771, 2017. 98

[5] B. Adamczewski and C. Faverjon. Méthode de Mahler: relations linéaires, transcendance et applications aux nombres automatiques. *Proc. Lond. Math. Soc. (3)*, 115(1):55–90, 2017. 98

[6] B. Adamczewski and C. Faverjon. Méthode de Mahler, transcendance et relations linéaires : aspects effectifs. Technical report, arXiv:1610.09136, 2018. arXiv version 2016, to appear in J. Théor. Nombres Bordeaux. 98

[7] J.-P. Allouche and J. O. Shallit. *Automatic sequences, theory, applications, generalizations*. Cambridge University Press, 2003. 71, 76

[8] J. Bell and R. Moosa. F-sets and finite automata. Technical report, arXiv:1712.03800, 2017. 98

[9] J.-P. Bézivin. Une généralisation du théorème de Skolem-Mahler-Lech. *Quart. J. Math. Oxford Ser. (2)*, 40(158):133–138, 1989. 86

[10] E. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo (2)*, 27:247–271, 1909. 78

[11] A. Bridy. Automatic sequences and curves over finite fields. *Algebra Number Theory*, 11(3):685–712, 2017. 98

[12] J. R. Chen. On the representation of a large even integer as the sum of a prime and the product of at most two primes. *Kexue Tongbao (Foreign Lang. Ed.)*, 17:385–386, 1966. 77

[13] C. Chevalley. *Introduction to the theory of algebraic functions of one variable*. Mathematical Surveys, No. VI. Amer. Math. Soc., New York, N. Y., 1951. 95

[14] A. Cobham. On the Hartmanis-Stearns problem for a class of tag machines. In *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory*, pages 51–60, 1968. Also appeared as IBM Research Technical Report RC-2178, August 23 1968. 79

[15] A. Cobham. Uniform tag sequences. *Math. Systems Theory*, 6:164–192, 1972. 72

[16] H. Cramér. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.*, 2:23–46, 1936. 76

[17] H. Derksen. A Skolem-Mahler-Lech theorem in positive characteristic and finite automata. *Invent. Math.*, 168(1):175–224, 2007. 88

[18] H. Derksen and D. Masser. Linear equations over multiplicative groups, recurrences, and mixing I. *Proc. Lond. Math. Soc. (3)*, 104(5):1045–1083, 2012. 98

[19] H. Derksen and D. Masser. Linear equations over multiplicative groups, recurrences, and mixing II. *Indag. Math. (N.S.)*, 26(1):113–136, 2015. 98

[20] S. Eilenberg. *Automata, languages, and machines. Vol. A*. Academic Press [A Subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58. 71, 73, 75

[21] G. Everest, A. Van Der Poorten, I. Shparlinski, and T. Ward. *Recurrence sequences*, volume 104 of *Math. Surveys Monogr.* Amer. Math. Soc., Providence, RI, 2003. 86

[22] S. Ferenczi and C. Mauduit. Transcendence of numbers with a low complexity expansion. *J. Number Theory*, 67:146–161, 1997. 83

[23] E. Fouvry and C. Mauduit. Sommes des chiffres et nombres presque premiers. *Math. Ann.*, 305:571–599, 1996. 77

[24] A. O. Gelfond. Sur les nombres qui ont des propriétés additives et multiplicatives données. *Acta Arith.*, 13:259–265, 1967/1968. 78

[25] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples. I. *Ann. of Math. (2)*, 170:819–862, 2009. 76

[26] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167:481–547, 2008. 76

[27] H. Hahn. *Gesammelte Abhandlungen/Collected works. Band 1/Vol. 1*. Springer-Verlag, Vienna, 1995. With biographical sketches by Karl Popper and by L. Schmetterer and K. Sigmund, and commentaries on Hahn's work by H. Heuser, H. Sagan and L. Fuchs, Edited by Schmetterer and Sigmund and with a foreword by Popper. 95

[28] G. Hansel. Une démonstration simple du théorème de Skolem-Mahler-Lech. *Theoret. Comput. Sci.*, 43(1):91–98, 1986. 86

[29] J. Hartmanis and H. Shank. On the recognition of primes by automata. *J. Assoc. Comput. Mach.*, 15:382–389, 1968. 76

[30] J. Hartmanis and R. E. Stearns. On the computational complexity of algorithms. *Trans. Amer. Math. Soc.*, 117:285–306, 1965. 78

[31] I. Kaplansky. Maximal fields with valuations. *Duke Math. J.*, 9:303–321, 1942. 96

[32] K. S. Kedlaya. The algebraic closure of the power series field in positive characteristic. *Proc. Amer. Math. Soc.*, 129(12):3461–3470 (electronic), 2001. 95, 96

[33] K. S. Kedlaya. Finite automata and algebraic extensions of function fields. *J. Théorie Nombres Bordeaux*, 18(2):379–420, 2006. 96

[34] K. S. Kedlaya. On the algebraicity of generalized power series. *Beitr. Algebra Geom.*, 58(3):499–527, 2017. 98

[35] C. Lech. A note on recurring series. *Ark. Mat.*, 2:417–421, 1953. 86, 87

[36] D. J. Leitner. Linear equations over multiplicative groups in positive characteristic. *Acta Arith.*, 153(4):325–347, 2012. 98

[37] D. J. Leitner. Linear equations over multiplicative groups in positive characteristic II. *J. Number Theory*, 180:169–194, 2017. 98

[38] J. Liouville. Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même reductible à des irrationelles algébriques. *C. R. Acad. Sci. Paris*, 18:883–885, 910–911, 1844. 79

[39] J. H. Loxton and A. J. Van Der Poorten. Arithmetic properties of the solutions of a class of functional equations. *J. Reine Angew. Math.*, 330:159–172, 1982. 79

[40] J. H. Loxton and A. J. Van Der Poorten. Arithmetic properties of automata: regular sequences. *J. Reine Angew. Math.*, 392:57–69, 1988. 79

[41] K. Mahler. Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen. *Math. Ann.*, 101:342–366, 1929. Corrigendum, **103** (1930), 532. 79

[42] K. Mahler. Eine arithmetische eigenshaft der taylor-koeffizienten rationaler funktionen. In *Proc. Kon. Nederlandsche Akad. v. Wetenschappen*, volume 38, pages 50–60. 1935. 86

[43] K. Mahler. On the Taylor coefficients of rational functions. *Math. Proc. Cambridge Phil. Soc.*, 52:39–48, 1956. 86

[44] K. Mahler. Addendum to the paper "On the Taylor coefficients of rational functions". *Math. Proc. Cambridge Phil. Soc.*, 53:544, 1957. 86

[45] B. Martin, C. Mauduit, and J. Rivat. Fonctions digitales le long des nombres premiers. *Acta Arith.*, 170(2):175–197, 2015. 98

[46] C. Mauduit and J. Rivat. Sur un problème de Gelfond: la somme des chiffres des nombres premiers. *Ann. of Math. (2)*, 171:1591–1646, 2010. 78

[47] C. Mauduit and J. Rivat. Prime numbers along Rudin-Shapiro sequences. *J. Eur. Math. Soc. (JEMS)*, 17(10):2595–2642, 2015. 98

[48] P. Mihăilescu. Primary cyclotomic units and a proof of Catalan's conjecture. *J. Reine Angew. Math.*, 572:167–195, 2004. 90

[49] M. Minsky and S. Papert. Unrecognizable sets of numbers. *J. Assoc. Comput. Mach.*, 13:281–286, 1966. 74, 76

[50] M. Morse and G. A. Hedlund. Symbolic Dynamics. *Amer. J. Math.*, 60:815–866, 1938. 78

[51] C. Müllner. Automatic sequences fulfill the Sarnak conjecture. *Duke Math. J.*, 166(17):3219–3290, 2017. 98

[52] P. Philippon. Groupes de Galois et nombres automatiques. *J. Lond. Math. Soc. (2)*, 92(3):596–614, 2015. 98

[53] J. Pintz. Cramér vs. Cramér. On Cramér's probabilistic model for primes. *Funct. Approx. Comment. Math.*, 37(part 2):361–376, 2007. 76

[54] P. Ribenboim. *The new book of prime number records*. Springer-Verlag, New York, 1996. 77

[55] D. Ridout. Rational approximations to algebraic numbers. *Mathematika*, 4:125–131, 1957. 82

[56] R. W. Ritchie. Finite automata and the set of squares. *J. Assoc. Comput. Mach.*, 10:528–531, 1963. 74

[57] K. F. Roth. Rational approximations to algebraic numbers. *Mathematika*, 2:1–20, 1955. Corrigendum, p. 168. 80

[58] O. Salon. Suites automatiques à multi-indices et algébricité. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(12):501–504, 1987. 74, 75, 94, 98

[59] W. M. Schmidt. *Diophantine Approximation*, volume 785 of *Lecture Notes in Mathematics*. Springer-Verlag, 1980. 83

[60] M.-P. Schützenberger. A remark on acceptable sets of numbers. *J. Assoc. Comput. Mach.*, 15:300–303, 1968. 76

[61] J.-P. Serre. *Local fields*, volume 67 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1979. Translated from the French by Marvin Jay Greenberg. 96

[62] H. Sharif and C. F. Woodcock. Algebraic functions over a field of positive characteristic and Hadamard products. *J. Lond. Math. Soc. (2)*, 37(3):395–403, 1988. 91

[63] T. Skolem. Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen und diophantischer Gleichungen. In *8. Skand. Mat. Kongr., Stockholm*, pages 163–188. 1934. 86

[64] K. Soundararajan. The distribution of prime numbers. In *Equidistribution in number theory, an introduction*, volume 237 of *NATO Sci. Ser. II Math. Phys. Chem.*, pages 59–83. Springer, Dordrecht, 2007. 76

[65] K. Soundararajan. Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım. *Bull. Amer. Math. Soc. (N.S.)*, 44:1–18, 2007. 76

[66] T. Tao. *Structure and randomness*. Amer. Math. Soc., Providence, RI, 2008. Pages from year one of a mathematical blog. 76, 86

[67] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.*, 42:230–265, 1936. 78

[68] A. J. Van Der Poorten. Some facts that should be better known, especially about rational functions. In *Number theory and applications (Banff, AB, 1988)*, volume 265 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 497–528. Kluwer Acad. Publ., Dordrecht, 1989. 86