Contents

1

Chapter 1
Number theoretic aspects of regular sequences

Michael Coons and Lukas Spiegelhofer

Abstract We present a survey of results concerning regular sequences and related objects. Regular sequences were defined in the early 90s by Allouche and Shallit as a combinatorially, algebraically, and analytically interesting generalisation of automatic sequences. In this chapter, after an historical introduction, we follow the development from automatic sequences to regular sequences, and their associated generating functions, to Mahler functions. We then examine size and growth properties of regular sequences. The last half of the chapter focuses on the algebraic, analytic and Diophantine properties of Mahler functions. In particular, we survey the rational-transcendental dichotomies of Mahler functions, due to Bézivin, and of regular numbers, due to Bell, Bugeaud and Coons.

1.1 Introduction

The concept of 'number' is central to mathematics and paramount to number theory. From the mathematical standpoint, one of the most important ways to view and

---

Michael Coons
School of Mathematical and Physical Sciences, University of Newcastle, Callaghan, Australia
e-mail: michael.coons@newcastle.edu.au

Lukas Spiegelhofer
Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wien, Austria
e-mail: lukas.spiegelhofer@tuwien.ac.at

* The research of M. Coons was supported in part by Australian Research Council grant DE140100223. Lukas Spiegelhofer was supported by the Austrian Science Fund (FWF), projects F5502-N26 and F5505-N26, which are part of the Special Research Program "Quasi Monte Carlo Methods: Theory and Applications", and also by the ANR–FWF joint project MuDeRa (Multiplicativity, Determinism and Randomness). The authors thank the Erwin Schrödinger Institute for Mathematics and Physics where part of this chapter was written during the workshop on "Normal Numbers: Arithmetic, Computational and Probabilistic Aspects."

treat numbers is algebraically, that is, to consider the integers as the ring $\mathbb{Z}$ under the operations addition and multiplication and the rationals $\mathbb{Q}$ as the field of fractions of $\mathbb{Z}$. Of course, from there interest is extended to the algebraic numbers, the field $\overline{\mathbb{Q}}$ of numbers, which are zeroes of polynomials with integer coefficients. The study of algebraic numbers and their properties is a continual fount of results and questions that for centuries has provided the foundational structures of mathematics and will—beyond doubt—form a significant part of these foundations for centuries to come.

The numbers of the preceding paragraph are abstract and in that sense do not really need to be represented. Yet, when one wishes to give an example of an integer, say 2 or 10 or 1729, one must write something down; if you wish to use only tick marks, treating the example 1729 will require large amounts of both time and space. Thus we have adopted the base system, with base 10—the number of fingers the average human has—as the most popular base for humans. The concept of 'base-expansion' is inseparable from modern computation and is fundamental to computer science. The use and importance of base expansions (predominantly binary) has become even more important with the advent of digital computers.

For those of us with interests at the interface of mathematics and theoretical computer science, the characterisation of relationships between the algebraic viewpoint and the base-expansion viewpoint is an extremely important and interesting area of research. Two specific questions stand out here and form the backdrop of our chapter.

1.1.1 Two important questions

The first is an old question of Borel [15] concerning the probabilistic properties (probabilités dénombrables) of base expansions of real algebraic numbers.

Question 1.1 (Borel, 1909). Is the base expansion of an irrational algebraic real number normal?

Recall that a real number x is called *simply normal to the base k* (or *k-simply normal*) if each of $0,1,\ldots,k-1$ occurs in the base-k expansion of x with equal frequency $1/k$. This number x is then called *normal to the base k* (or *k-normal*) provided it is k^m-simply normal for all positive integers m, and the number x is just called *normal* if this is true for all integers $k \geqslant 2$.

While Borel's question is asked from the standpoint of probability, Hartmanis and Stearns [31] were interested in the—at least morally related—question of computability. To state their question, we remind the reader that a real number x is *computable in real time* provided there is a multitape Turing machine that can compute the first n bits of x in time $O(n)$.

Question 1.2 (Hartmanis and Stearns, 1965). Do there exist irrational algebraic real numbers which are computable in real time?

Presumably, the answers to these questions are 'yes' and 'no,' respectively, though we stress here that our presumption is extremely presumptive. These presumptive answers reflect the well-observed notion that algebraic manipulations tend to do strange things to base expansions. In fact, compared to what is expected, very little is known about the digital properties of real algebraic numbers. For those interested, Bugeaud's recent work [17] provides a comprehensive exposition.

While Questions 1.1 and 1.2 are posed to study the digital properties of real algebraic numbers, in this chapter, we concern ourselves with a flipped version of these questions: *what are the number theoretic properties of real numbers whose expansions are highly structured?*

Real numbers with eventually periodic base expansions are the simplest numbers and sequences one can consider in our context. These numbers are not normal, are computable, and of course are algebraic—they are the zeroes of linear polynomials. This perceived exception to Questions 1.1 and 1.2 is why the word 'irrational' appears in these questions. Indeed, the rational numbers are in many ways fundamentally different from the irrational algebraic numbers. For examples, see Dirichlet's approximation theorem and Roth's theorem [51] on the irrationality exponent of algebraic numbers. The digital properties of rational numbers have been almost completely classified (up to some deep questions about the orbits of primitive roots).

From a computational point of view, the next step is to consider real numbers whose base-k expansion is k-automatic[1] for some integer $k \geqslant 2$. This is where things become extremely interesting. In fact, here the base starts to matter. Recall that if a number is rational, then its base expansion is *eventually periodic in every base*. This is not true for numbers that are k-automatic for some integer $k \geqslant 2$. Cobham [20] showed that if a real number is both k-automatic and l-automatic for two integers k and l that are multiplicatively independent[2], then that real number is rational.

This difference from rationals continues with the complexity of base expansions. For a rational written in base k, the number of strings of digits of length n that occur in the expansion is bounded by a constant, while for a k-automatic real number, the number of strings can increase with n. But not too fast; this number is $O(n)$, and so an automatic number is not normal since a normal number must have all k^n possible strings occur.

For Borel's question it may seem hopeful to then wonder if the set of automatic numbers contains an irrational algebraic number, but the negative answer to this question, which became known somewhat as the Cobham–Loxton–van der Poorten Conjecture, was settled[3] by Adamczewski and Bugeaud in 2007 [2].

Theorem 1.3 (Adamczewski and Bugeaud). *The base expansion of an irrational real algebraic number cannot be output by a finite automaton.*

---

[1] For a detailed account of automatic sequences, see the monograph of Allouche and Shallit [5].

[2] Two integers k and l are multiplicatively independent provided $\log k/\log l$ is irrational.

[3] This result is inherent in the work of Cobham. In the 1980s Loxton and van der Poorten [38] claimed to have proved that an automatic number is either rational or transcendental, but a few unresolvable flaws were found in their argument. This is why their name is associated with the conjecture.

1.1.2 Three (or four) hierarchies in one

According to Loxton [37], "the result about the decimal expansion of algebraic irrationals and finite automata suggests an alternative theoretical approach to randomness. We can try to assign a measure of computational complexity to a sequence by means of the following hierarchy:

(L0) [eventually] periodic sequences,
(L1) [...] sequences generated by finite automata,
(L2) sequences generated by automata with one push-down store,
(L3) sequences generated by non-deterministic automata with one push-down store, and
(L4) sequences generated by Turing machines.

Essentially, the n-th term of an [automatic] sequence is computed from the input n without any memory of earlier terms. A push-down store allows an arbitrary number of terms of the sequence to be stored and recalled later, the first one in being the last one out. Two push-down stores are equivalent to the doubly infinite tape of a Turing machine, which explains why the classification stops as it does. A random sequence is now one which cannot be generated by any machine less powerful than a Turing machine."

The well-informed reader will recognise Loxton's hierarchy as a subset of the Chomsky–Schützenberger hierarchy of formal languages. This type of language-theoretical hierarchy, while classical and certainly of interest, lacks the mathematical structure to delve into such arithmetic questions that we will address here—especially at the higher levels of the hierarchy.

We present here a more natural hierarchy for such questions based on the work of Mahler and the generalisation of automatic sequences presented by Allouche and Shallit. This hierarchy will be one of sequences, numbers, and functions simultaneously. From the standpoint of integer sequences, the *Mahler hierarchy* is as follows:

(M0) eventually periodic sequences,
(M1) automatic sequences,
(M2) regular sequences,
(M3) coefficient sequences of Mahler functions, and
(M4) integer sequences[4].

Levels (M0) and (M1) are taken from Loxton's hierarchy. Regular sequences were introduced in 1992 by Allouche and Shallit [4]. Following their treatment[5], let $\mathbb{C}$

---

[4] We make no comment on the randomness properties of integer sequences, but will be content with their generality as is.

[5] Allouche and Shallit gave a more general treatment for sequences taking values in Noetherian rings. In our applications, the most important settings are those of the integers and complex numbers, depending on the type of result presented. For our purposes, for results on sequences and numbers, the integers will be the standard setting, and for results on power series those with complex coefficients will be the most important.

denote the field of complex numbers and define the *k-kernel* of $f:\mathbb{Z}_{\geqslant 0}\to\mathbb{C}$ as the set

$$\mathrm{Ker}_k(f) := \left\{\{f(k^\ell n+r)\}_{n\geqslant 0} : \ell\geqslant 0, 0\leqslant r<k^\ell\right\}. \tag{1.1}$$

Definition 1.4 (Allouche and Shallit). Let $k\geqslant 1$ be an integer. A sequence f taking values in $\mathbb{C}$ is called *k-regular* provided the $\mathbb{C}$-vector space $\langle \mathrm{Ker}_k(f)\rangle_{\mathbb{C}}$ spanned by $\mathrm{Ker}_k(f)$ is finite dimensional over $\mathbb{C}$.

Allouche and Shallit introduced regular sequences a direct generalisation of automatic sequences based on the k-kernel. Their generalisation rests on a result of Cobham [21], who showed the following.

Theorem 1.5 (Cobham). *A sequence f is k-automatic if and only if $\mathrm{Ker}_k(f)$ is finite.*

While the notion of k-regularity is certainly worth studying in its own right, it becomes much more important when viewed as a bridge between the areas of theoretical computer science and number theory. As Allouche and Shallit showed, this notion is a direct extension of that of automatic sequences. Moreover, it is an extension that is algebraically, analytically, and arithmetically interesting and important.

The algebraic properties start with a correspondence between regular sequences and finite sets of matrices. Indeed, Allouche and Shallit [4, Lemma 4.1] showed that for a Noetherian ring R, *an R-valued sequence f is k-regular if and only if there exist a positive integer d, a finite set of matrices $\mathscr{A}_f=\{\mathbf{A}_0,\dots,\mathbf{A}_{k-1}\}\subseteq R^{d\times d}$, and vectors $\mathbf{v},\mathbf{w}\in R^d$ such that*

$$f(n)=\mathbf{w}^T\mathbf{A}_{i_0}\cdots\mathbf{A}_{i_s}\mathbf{v}, \tag{1.2}$$

where $(n)_k=i_s\cdots i_0$ is the base-k expansion of n.

The analytic importance comes via a result of Becker [7] relating regular sequences to Mahler functions. Recall the following definition; see the works of Mahler [39, 40, 41, 42].

Definition 1.6. A power series $F(z)\in\mathbb{C}[[z]]$ is *k-Mahler* for an integer $k\geqslant 2$ provided there is an integer $d\geqslant 1$ and polynomials $a_0(z),\dots,a_d(z)\in\mathbb{C}[z]$ with $a_0(z)a_d(z)\neq 0$ such that

$$a_0(z)F(z)+a_1(z)F(z^k)+\cdots+a_d(z)F(z^{k^d})=0. \tag{1.3}$$

The minimal such d is called the *degree* of the Mahler function.

The above-mentioned result of Becker states that *if $\{f(n)\}_{n\geqslant 0}$ is a k-regular sequence, then the generating function $F(z)=\sum_{n\geqslant 0}f(n)z^n$ is a k-Mahler function.* This established that those sequences in level (M3) contain those in (M2).

The arithmetic interest and importance of k-regular sequences are precisely the content of this chapter. We will present properties and results to this effect in the context of the Mahler hierarchy. It is important to note that while the Mahler hierarchy is stated in terms of sequences, it can be stated in term of numbers and functions as well.

Definition 1.7. If a sequence $\{f(n)\}_{n\geqslant 0}$ is k-automatic (resp. k-regular), then we call the generating function $F(z)=\sum_{n\geqslant 0} f(n)z^n$ *k-automatic* (resp. *k-regular*) as well, and refer to F(z) as a *k-automatic function* (resp. a *k-regular function*).

In this way the levels (M1)–(M4) of the Mahler hierarchy can be translated to a hierarchy of functions as

(M1) automatic functions,
(M2) regular functions,
(M3) Mahler functions, and
(M4) general power series.

The 'number' version of the hierarchy is stated *mutatis mutandis* using the following definition.

Definition 1.8. Let $k\geqslant 2$ and $b\geqslant 2$ be integers. If $F(z)$ is a k-automatic function (resp. k-regular or k-Mahler), then we call the special value $F(1/b)$ a k-automatic number (resp. k-regular or k-Mahler).

Note that our notion of k-automatic number is more general than the traditional definitions; we call something an automatic number if it is the special value of and automatic function. In most of the literature a real number is called k-automatic if its base-k expansion can be produced by an automaton. This is not the case for all of the numbers in our class. For example, the number $\sum_{n\geqslant 0} 3^{-2^n}$ is 2-automatic under our definition, though its base-2 expansion is not 2-automatic. Being able to treat such numbers is just one example of the strength and generality of using the framework of the Mahler hierarchy.

1.2 From automatic to regular to Mahler

In this section, we describe automatic and regular sequences based on their k-kernel and develop their properties as coefficient sequences of Mahler functions. We first recall the definitions from the context of the k-kernel with a little more generality than the previous section, then we give many simple properties and provide some examples.

We take Cobham's result (Theorem 1.5) as our definition of automaticity.

Notation 1.9 *Unless otherwise specified, a sequence f will be one that takes values in a commutative ring R, which when necessary to avoid complication will be taken as a subring of the complex numbers.*

Definition 1.10. A sequence f is k-automatic if and only if $\mathrm{Ker}_k(f)$ is finite.

Example 1.11. The canonical example of an automatic sequence is the *Thue–Morse sequence*. The Thue–Morse sequence $\{t(n)\}_{n\geqslant 0}$ over the alphabet $\{-1,1\}$ is given by $t(n):=(-1)^{s(n)}$ where $s(n)$ is the number of 1s in the binary expansion of the number n. Using this definition, it is immediate that the sequence $\{t(n)\}_{n\geqslant 0}$ is 2–automatic. That is, there is a deterministic finite automaton that takes the binary expansion of n as input and outputs the value $t(n)$; see Figure 1.1.

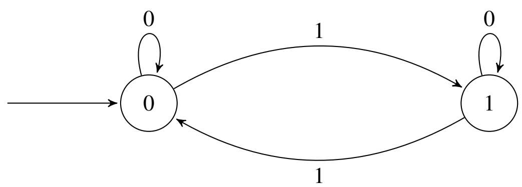

Fig. 1.1 The 2-automaton that produces the Thue–Morse sequence.

To show that t is 2-automatic using the Cobham-inspired definition based on the k-kernel, it is enough to note that $t(2n) = t(n)$ and $t(2n+1) = -t(n)$, so that $\mathrm{Ker}_2(t)$ has only two elements; namely, the sequences $t(n)$ and $-t(n)$.

As stated by Allouche in Shallit in their foundational paper [4], “unfortunately, the range of automatic sequences is necessarily finite, and this restricts their descriptive power.”

Definition 1.12. The sequence f taking values in a ring R is k-regular provided the k-kernel of f is contained in a finitely generated R-module.

Example 1.13. Let $\{s(n)\}_{n\geqslant 0}$ be *Stern's diatomic sequence*, which is determined by the relations $s(0) = 0$, $s(1) = 1$, and for $n \geqslant 0$, by

$$s(2n) = s(n), \quad \text{and} \quad s(2n+1) = s(n) + s(n+1).$$

These recursions immediately imply that the 2-kernel of s is contained in the $\mathbb{Z}$-module generated by $\{s(n)\}_{n\geqslant 0}$ and $\{s(n+1)\}_{n\geqslant 0}$, so that s is 2-regular. Note that s takes infinitely many values as well—$s(2^n+1) = n+1$—so that s is not 2-automatic.

The definition of k-regularity implies that there are a finite number of sequences $f_1, \ldots, f_d$ such that each element of the k-kernel of f is an R-linear combination of $f_1, \ldots, f_d$. This finite number of sequences can be taken in many ways, though two of these ways stand out. The first is to use an R-module basis for the R-module generated by the k-kernel of f. This is useful for proving results where minimality or irreducibility is important. The second is to take a spanning set directly from the k-kernel itself. This set is useful for more combinatorial results since it provides useful and usable recurrences, especially for manipulating sums. We record this result in the following lemma, the proof of which can be found in [4], though it is a worthy (and easy) exercise for the reader wishing to sharpen their teeth a bit on these ideas.

Lemma 1.14 (Allouche and Shallit). *The following are equivalent:*

(a) f is k-regular,

(b) the R-module generated by $\mathrm{Ker}_k(f)$ *is generated by a finite number of elements of* $\mathrm{Ker}_k(f)$*,*

(c) there exists an integer E *such that for all* $e_j > E$*, each subsequence* $f(k^{e_j}n + a_j)$ *with* $0 \leqslant a_j < k^{e_j}$ *can be expressed as an R-linear combination*

$$f(k^{e_j}n + a_j) = \sum_i c_{ij} f(k^{h_{ij}}n + b_{ij}),$$

where $h_{ij} \leqslant E$ *and* $0 \leqslant b_{ij} < k^{h_{ij}}$*,*

(d) there exist an integer d *and* d *sequences* $f = f_1, \ldots, f_d$ *such that for* $1 \leqslant i \leqslant d$ *the* k *sequences* $f_i(kn + a)$*,* $0 \leqslant a < k$*, are R-linear combinations of the* f_i*,*

(e) there exist an integer d*,* d *sequences* $f = f_1, \ldots, f_d$ *and* k *matrices* $\mathbf{A}_0, \ldots, \mathbf{A}_{k-1} \in R^{d\times d}$ *such that if* $\mathbf{v}(n) = [f_1, \ldots, f_d]^T$*, then* $\mathbf{v}(kn + a) = \mathbf{A}_a \mathbf{v}(n)$ *for* $0 \leqslant a < k$*.*

One of the most fundamental and important characterisations of k-regular sequence is their matrix formulation [4, Lemma 4.1].

Lemma 1.15 (Allouche and Shallit). *A sequence* f *is* k*-regular if and only if there exist a positive integer* d*, a finite set of matrices* $\mathscr{A}_f = \{\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}\} \subseteq R^{d\times d}$*, and vectors* $\mathbf{v}, \mathbf{w} \in R^d$ *such that*

$$f(n) = \mathbf{w}^T \mathbf{A}_{i_0} \cdots \mathbf{A}_{i_s} \mathbf{v}, \tag{1.4}$$

where $(n)_k = i_s \cdots i_0$ *is the base-*k *expansion of* n*.*

Proof. We prove only the right-hand implication; the other is left as an exercise for the reader.

Suppose that f is k-regular and $(n)_k = i_s \cdots i_0$ is the base-k expansion of n. By Lemma 1.14(e), there exist an integer d, d sequences $f = f_1, \ldots, f_d$ and k matrices $\mathbf{A}_0, \ldots, \mathbf{A}_{k-1} \in R^{d\times d}$ such that if $\mathbf{v}(n) = [f_1, \ldots, f_d]^T$, then $\mathbf{v}(kn + a) = \mathbf{A}_a \mathbf{v}(n)$ for $0 \leqslant a < k$. Since $f = f_1$, setting $\mathbf{v} := \mathbf{v}(0)$ and $\mathbf{e}_1 := [1\ 0\ \cdots\ 0]^T$, we have that for each $n \geqslant 0$ that

$$f(n) = \mathbf{e}_1^T \mathbf{A}_{i_0} \cdots \mathbf{A}_{i_s} \mathbf{v}.$$

Setting $\mathbf{w} := \mathbf{e}_1$ gives the desired result. $\square$

Definition 1.16. Let f be a k-regular sequence taking values in the ring R. If $\mathscr{A}_f = \{\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}\} \subseteq R^{d\times d}$ is a finite set of matrices and $\mathbf{v}, \mathbf{w} \in R^d$ vectors such that

$$f(n) = \mathbf{w}^T \mathbf{A}_{i_0} \cdots \mathbf{A}_{i_s} \mathbf{v},$$

where $(n)_k = i_s \cdots i_0$ is the base-k expansion of n, then we call the tuple $(\mathbf{w}, \mathscr{A}_f, \mathbf{v})$ the *linear representation of* f.

Example 1.17. As we saw in a previous example, the Stern sequence is 2-regular. Using Lemma 1.14(e) and following the notation of Lemma 1.15, one can show that the Stern sequence has linear representation

$$\left([1\ 0], \{\mathbf{A}_0, \mathbf{A}_1\} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\}, [1\ 0]\right).$$

We define the convolution of two sequences f and g by

$$f \star g(n) := \sum_{i+j=n} f(i)g(j).$$

The following result, which provides for the algebraic structure of the set of k-regular sequences, is due to Allouche and Shallit [4, Theorem 3.1 and Corollary 3.2], though we offer here a slightly different proof.

Theorem 1.18 (Allouche and Shallit). *The set of k-regular sequences forms a ring under standard addition and convolution.*

Proof. It is clear that the set of k-regular sequences forms a group under addition.

To see that the set is closed under convolution, let f and g be two k-regular sequences, whose k-kernels are contained in the R-modules generated by $f_1, f_2, \ldots, f_d$ and $g_1, g_2, \ldots, g_e$, respectively. To prove the theorem it is enough to show that the k-kernel of $f \star g$ is contained in the R-module

$$C := \left\langle \left\{ \{(f_i \star g_j)(n)\}_{n \geqslant 0} : 1 \leqslant i \leqslant d, 1 \leqslant j \leqslant e \right\} \right\rangle_R.$$

To see this, suppose that $c \in \mathrm{Ker}_k(f \star g)$ and that $\ell \geqslant 0$ and r ($0 \leqslant r < k^\ell$) are such that $c(n) = (f \star g)(k^\ell n + r)$ for all $n \geqslant 0$. Then there are $\alpha_1, \ldots, \alpha_d, \beta_1, \ldots, \beta_e \in R$ such that

$$f(k^\ell n + r) = \sum_{i=0}^{d} \alpha_i f_i(n) \quad \text{and} \quad g(k^\ell n + r) = \sum_{j=0}^{e} \beta_j g_j(n).$$

Now

$$\begin{aligned} c(n) = (f \star g)(k^\ell n + r) &= \sum_{a=0}^{n} f(k^\ell a + r) g(k^\ell (n-a) + r) \\ &= \sum_{a=0}^{n} \sum_{i=0}^{d} \alpha_i f_i(a) \sum_{j=0}^{e} \beta_j g_j(n-a) \\ &= \sum_{i=0}^{d} \sum_{j=0}^{e} \alpha_i \beta_j \sum_{a=0}^{n} f_i(a) g_j(n-a) \\ &= \sum_{i=0}^{d} \sum_{j=0}^{e} \alpha_i \beta_j (f_i \star g_j)(n) \qquad (1.5) \end{aligned}$$

is an element of C, which proves the theorem. □

Equality (1.5) essentially gives a description of the matrix representation of the k-regular convolution $f \star g$, but working this out can in practice be extremely complicated—the bookkeeping involved is nothing short of a nightmare. From the number-theoretic perspective, the most useful special case of convolution is $1 \star f$, which is the sequence of partial sums of f. Fortunately, in this case the details are not so unfriendly. The following result is due to Dumas [27, Lemma 1], which we reproduce here with a few fixed typos.

Proposition 1.19 (Dumas). *Let f be a k-regular sequence, with matrix presentation as in* (1.4)*. Then the sequence $g(m) = (1 \star f)(m) = \sum_{1 \leqslant n \leqslant m} f(n)$ is k-regular and*

$$g(m) = \mathbf{x}^T \mathbf{G}_{i_s} \cdots \mathbf{G}_{i_0} \mathbf{y},$$

where $(m)_k = i_s \ldots i_0$, $\mathbf{x}^T := [0_{1\times d}\ \mathbf{w}^T]$, $\mathbf{y}^T := [\mathbf{v}^T\ 0_{1\times d}]$ and for $b \in \{0, \ldots, k-1\}$,

$$\mathbf{G}_b := \begin{bmatrix} \mathbf{B}_0 & 0 \\ \mathbf{B}_0 - \mathbf{B}_{b+1} - \mathbf{A}_0 & \mathbf{I}_{d\times d} \end{bmatrix},$$

where $\mathbf{B}_b := \sum_{\ell=b}^{k-1} \mathbf{A}_\ell$ for $b = 0, \ldots, k-1$ and $\mathbf{B}_k := 0$.

Proof. Let $m \geqslant 1$ be an integer with $(m)_k = b_r b_{r-1} \cdots b_0$, and write $g(m) := \sum_{1 \leqslant n \leqslant m} f(n)$. It is quite clear that

$$g(m) = \mathbf{x}^T \left(\sum_{1 \leqslant n \leqslant m} \mathbf{A}_{(n)_k} \right) \mathbf{v} = \mathbf{x}^T \left(\sum_{0 \leqslant i \leqslant r} \left(\sum_{1 \leqslant j \leqslant b_i} \mathbf{A}_j \right) \sum_{\substack{|w| \leqslant i \\ w \in \{0,\ldots,k-1\}^*}} \mathbf{A}_w \right) \mathbf{v}, \quad (1.6)$$

where we use that convention that if $b_i = 0$, then $\sum_{1 \leqslant j \leqslant b_i} \mathbf{A}_j = 0$, and when $i = 0$ that $\sum_{|w| \leqslant i, w \in \{0,\ldots,k-1\}^*} \mathbf{A}_w = \mathbf{I}_{d\times d}$.

Now, in the notation presented in the statement of the proposition, it is quite clear that

$$\mathbf{B}_0^i = \sum_{\substack{|w| \leqslant i \\ w \in \{0,\ldots,k-1\}^*}} \mathbf{A}_w,$$

where our above convention is preserved since we understand $\mathbf{B}_0^0 = \mathbf{I}_{d\times d}$. Also, we note that

$$\mathbf{B}_0 - \mathbf{B}_{b_i+1} - \mathbf{A}_0 = \sum_{1 \leqslant j \leqslant b_i} \mathbf{A}_j,$$

where again our above convention is preserved since for $b_i = 0$, we have $\mathbf{B}_0 - \mathbf{B}_1 - \mathbf{A}_0 = 0$.

With this information of the preceding paragraph, we interpret the equality (1.6) as

$$g(m) = \mathbf{x}^T \left(\sum_{0 \leqslant i \leqslant r} \big(\mathbf{B}_0 - \mathbf{B}_{b_i+1} - \mathbf{A}_0 \big) \mathbf{B}_0^i \right) \mathbf{v}. \quad (1.7)$$

But this is exactly the output of the matrix representation for $g(m)$ as described in the statement of the proposition. □

The importance of the ring structure under addition and convolution begins with the following immediate corollary of Theorem 1.18.

Corollary 1.20 (Allouche and Shallit). *The set of k-regular functions forms a ring under standard addition and multiplication.*

This importance continues with the relationship to Mahler functions as provided by Becker [7]. Following Becker, we require the following definition and lemma regarding the Cartier operators.

Definition 1.21. Given a positive integer $k \geqslant 2$, we define the *Cartier operators* $\Lambda_0, \ldots, \Lambda_{k-1} : \mathbb{C}[[z]] \to \mathbb{C}[[z]]$ by

$$\Lambda_i\left(\sum_{n\geqslant 0} c(n)z^n\right) = \sum_{n\geqslant 0} c(kn+i)z^n,$$

for $i = 0, \ldots, k-1$.

Lemma 1.22. *Let $F(z), G(z) \in \mathbb{C}[[z]]$. For $i = 0, \ldots, k-1$ we have*

(a) $\Lambda_i(F(z^k)G(z)) = F(z)\Lambda_i(G(z))$, and
(b) $F(z) = \sum_{i=0}^{k-1} z^i \Lambda_i(F)(z^k)$,

where $\Lambda_i(F)(z^k)$ is understood as $\Lambda_i(F(z))$ evaluated at z^k, so that if $F(z) = \sum_{n\geqslant 0} f(n)z^n$, then $\Lambda_i(F)(z^k) = \sum_{n\geqslant 0} f(kn+i)z^{kn}$.

Proof. This is left as an exercise. □

Theorem 1.23 (Becker). *A k-regular function is a k-Mahler function.*

Proof. For convenience, we will assume that the k-regular function takes values in the complex numbers. This proof can be easily modified to give a result for any Noetherian ring R provided you work with the field of fractions of R.

Let $f = f_1, \ldots, f_d$ be a basis for the $\mathbb{C}$-vector space spanned by the k-kernel of f and set $F_i(z) := \sum_{n\geqslant 0} f_i(n)z^n$. Further, define the $\mathbb{C}(z)$-vector space V by

$$V := \langle \{F_i(z) : i = 1, \ldots, d\} \rangle_{\mathbb{C}(z)},$$

so that the set $\{F_i(z) : i = 1, \ldots, d\}$ is a basis for V, and define the operator

$$\Phi : V \to \mathbb{C}((z))$$

by $\Phi(G(z)) = G(z^k)$. We claim that $V = \Phi(V)$.

To show that $V \subset \Phi(V)$ we note that for each $i = 1, \ldots, d$

$$F_i(x) = \sum_{j=0}^{k-1} \sum_{n\geqslant 0} f_i(kn+j)(x^k)^n x^j,$$

and since each $\{f_i(kn+j)\}_{n\geqslant 0}$ is in the k-kernel of f it is a $\mathbb{C}$-linear combination of the basis sequences $f_1, \ldots, f_d$. Thus we may write

$$F_i(x) = \sum_{j=1}^{d} p_{i,j}(x) F_j(x^k) \tag{1.8}$$

where for each i, j we have $p_{i,j}(x) \in \mathbb{C}[x]$ and $\deg p_{i,j}(x) \leqslant k-1$. But since $\{F_i(z) : i = 1, \dots, d\}$ is a basis for V, we thus have that $\{F_i(z^k) : i = 1, \dots, d\}$ spans $\Phi(V)$, and so the relationship in (1.8) shows that $V \subset \Phi(V)$.

For the other inclusion, we set $\mathbf{F}(x) := [F_1(x), \dots, F_d(x)]^T$ and note that (1.8) gives

$$\mathbf{F}(x) = \mathbf{A}(x)\mathbf{F}(x^k), \tag{1.9}$$

where $\mathbf{A}(x) = (p_{i,j}(x))_{1 \leqslant i,j \leqslant d} \in \mathbb{C}[x]^{d \times d}$. Also, since $\{F_i(z) : i = 1, \dots, d\}$ is a basis for V, the matrix $\mathbf{A}(z)$ is nonsingular; if this were not the case, there would be a vector $\mathbf{v}(z) \in \mathbb{C}(z)^d$ such that $\mathbf{v}(z)\mathbf{A}(z) = 0$ so that by (1.9) we would have $\mathbf{v}(z)\mathbf{F}(z) = 0$, contradicting that the coordinates of $\mathbf{F}(z)$ are $\mathbb{C}(z)$-linear independent—they form a basis of V. Thus also

$$\mathbf{A}(z)^{-1}\mathbf{F}(z) = \mathbf{F}(z^k),$$

whence $\Phi(V) \subset V$, showing that $V = \Phi(V)$.

We note that the arguments of the previous two paragraphs also show that since V has dimension d, $F(z) \in V$, and $\Phi(V) \subset V$, the $d+1$ functions $F(z), F(z^k), \dots, F(z^{k^d}) \in V$ are $\mathbb{C}(z)$-linearly dependent, meaning there are polynomials $a_0(z), \dots, a_d(z) \in \mathbb{C}[z]$ such that

$$\sum_{i=0}^{d} a_i(z) F(z^{k^i}) = 0. \tag{1.10}$$

Of course, to prove the theorem, we must show that one has such a relationship with $a_0(z) \neq 0$.

Indeed, as Becker points out [7, p. 273], if one has a functional equation (1.10) with $a_j(z) \neq 0$ with $j > 0$ minimal, then we can just 'shift' it down to one smaller j by applying one of the Cartier operators, since from Lemma 1.22(a) we have for $a = 0, \dots, k-1$ that

$$0 = \Lambda_a \left(\sum_{i=j}^{d} a_i(z) F(z^{k^i}) \right) = \sum_{i=j}^{d} \Lambda_a \left(a_i(z) \right) F(z^{k^{i-1}}),$$

where we are guaranteed from Lemma 1.22(b) that for at least one $a = 0, \dots, k-1$ the polynomial $\Lambda_a(a_j(z))$ is nonzero. □

This argument can be adjusted to prove the following stronger form of Becker's theorem, and so we state it here as a corollary.

Corollary 1.24. *If R is a Noetherian ring and $F(z) \in R[[z]]$ is k-regular for an integer $k \geqslant 2$, then there is an integer $d \geqslant 1$ and polynomials $a_0(z), \dots, a_d(z) \in R[z]$ with $a_0(z)a_d(z) \neq 0$ such that*

$$a_0(z)F(z) + a_1(z)F(z^k) + \cdots + a_d(z)F(z^{k^d}) = 0.$$

That is $F(z)$ is k-Mahler satisfying a Mahler functional equation with coefficients in the ring $R[z]$.

The most important case in the above corollary is the case of $R = \mathbb{Z}$.

Example 1.25. Let s again denote the Stern sequence and set $S(z) := \sum_{n\geqslant 0} s(n)z^n$. Using the definition of s, we have

$$\begin{aligned} zS(z) &= z\sum_{n\geqslant 0} s(2n)z^{2n} + z\sum_{n\geqslant 0} s(2n+1)z^{2n+1} \\ &= z\sum_{n\geqslant 0} s(n)z^{2n} + \sum_{n\geqslant 0} s(n)z^{2n+2} + \sum_{n\geqslant 0} s(n+1)z^{2n+2} \\ &= zS(z^2) + z^2S(z^2) + \sum_{n\geqslant 0} s(n)z^{2n} \\ &= S(z^2)\left(1+z+z^2\right), \end{aligned}$$

which gives that the generating function $S(z)$ satisfies the 2-Mahler equation

$$zS(z) - (z^2+z+1)S(z^2) = 0.$$

1.2.1 *Some comparisons between regular and Mahler functions*

Becker's result, Theorem 1.23 above, shows that every regular function is a Mahler function. The converse of Becker's result is not true, which we can show as a consequence of the following result.

Proposition 1.26. *The sequence $\{a^n\}_{n\geqslant 0}$ is k-regular if and only if $a = 0$ or a is a root of unity.*

Proof. One direction is simple, since if $a = 0$ or a root of unity, the sequence of powers is periodic, and hence k-regular.

For the other direction, assume $\{a^n\}_{n\geqslant 0}$ is k-regular. Then there exist an integer r and integers $\lambda_0, \ldots, \lambda_{r-1}$, not all zero, such that

$$\sum_{j=0}^{r} \lambda_j a^{k^j n} = 0.$$

Now we use the Vandermonde determinant identity, which states that

$$\det \begin{bmatrix} 1 & b_0 & b_0^2 & \cdots & b_0^m \\ 1 & b_1 & b_1^2 & \cdots & b_1^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & b_m & b_m^2 & \cdots & b_m^m \end{bmatrix} = \prod_{0\leqslant i<j\leqslant m} (b_j - b_i).$$

It follows that the sequences $\{b_j^n\}_{n\geqslant 0}$ are linearly independent if and only if the numbers $b_0, b_1, \dots, b_m$ are distinct. Hence the numbers $1, a^k, a^{k^2}, \dots, a^{k^r}$ are not all distinct, and we must have $a^{k^j} = a^{k^l}$ for some $j \neq l$. Thus either $a = 0$ or a is a root of unity. $\square$

Example 1.27. The function $1/(1-2z)$ is k-Mahler for every k, but is not k-regular for any k. To see this, note that inside the disc of radius $1/2$ centred at zero we have that

$$F(z) := \frac{1}{1-2z} = \sum_{n\geqslant 0} 2^n z^n.$$

By Proposition 1.26, the sequence $\{2^n\}_{n\geqslant 0}$ is not k-regular for any k. But it is quite easy to check that $F(z) = 1/(1-2z)$ satisfies the Mahler equation

$$(1-2z)F(z) - (1-2z^k)F(z^k) = 0,$$

for any k, so that $F(z)$ is k-Mahler for each k.

In fact, Example 1.27 suggests the following result concerning the degree of rational Mahler functions.

Proposition 1.28. *If $R(z)$ is a nonzero rational function, then it is a k-Mahler function of degree 1 for every positive integer $k \geqslant 2$.*

Proof. Now write $R(z) = p(z)/q(z)$ for nonzero polynomials $p(z)$ and $q(z)$. Then $R(z)$ satisfies the k-Mahler equation

$$p(z^k)q(z)R(z) - p(z)q(z^k)R(z^k) = 0,$$

which is of degree 1. $\square$

While not all Mahler functions are regular functions, there are some describable families. For example, Becker showed that if $F(z)$ is k-Mahler and the coefficient $a_0(z)$ of $F(z)$ in the functional equation is a nonzero constant, then $F(z)$ is k-regular.

Theorem 1.29 (Becker [7]). *Let $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function satisfying*

$$\sum_{i=0}^{d} a_i(z)F(z^{k^i}) = 0,$$

where $0 \neq a_0(z) \in \mathbb{C}$ and $a_1(z), \dots, a_d(z) \in \mathbb{C}[z]$. Then $F(z)$ is k-regular.

Proof. Without loss of generality, we may assume that $a_0(z) = -1$, since we may just divide by the appropriate complex number if needed. Thus,

$$F(z) = \sum_{i=1}^{d} a_i(z)F(z^{k^i}). \tag{1.11}$$

Set $H := \max\{\deg a_i(z) : i = 1, \ldots, d\}$, and let V be the $\mathbb{C}$-vector space generated by the functions

$$G_{ij}(z) := z^i F(z^{k^j}) \qquad (i = 0, \ldots, H; j = 0, \ldots, d).$$

For $a = 0, \ldots, k-1$ we have $\Lambda_a(G_{ij}(z)) \in V$. To see this, note that if $j = 1, \ldots, d$, then by Lemma 1.22(a) we have

$$\Lambda_a(z^i F(z^{k^j})) = F(z^{k^{j-1}})\Lambda_a(z^i) \in V,$$

since $\Lambda_a(z^i)$ is a monomial (possibly a constant) of degree at most H. If $j = 0$, then we use the functional equation (1.11) and Lemma 1.22(a) to obtain

$$\Lambda_a(z^i F(z)) = \sum_{\ell=1}^{d} \Lambda_a\left(z^i a_\ell(z) F(z^{k^\ell})\right) = \sum_{\ell=1}^{d} \Lambda_a\left(z^i a_\ell(z)\right) F(z^{k^{\ell-1}}).$$

Since for each combination of i and ℓ, $\deg z^i a_\ell(z) \leqslant 2H$, we have $\deg \Lambda_a(z^i a_\ell(z)) \leqslant 2H/k \leqslant H$, so that $\Lambda_a(z^i F(z)) \in V$.

Since $\Lambda_a(V) \subset V$ for each $a = 0, \ldots, k-1$, we have that V is mapped into itself for any element in the semigroup $\Lambda := \langle\{\Lambda_0, \ldots, \Lambda_{k-1}\}\rangle$. Since V is finite-dimensional and $F(z) \in V$, we have that the set $\Lambda(F(z))$ (the semigroup Λ evaluated at $F(z)$ for each element) generates a finite-dimensional $\mathbb{C}$-vector space. But, using the definitions of regularity and the Cartier operators, this is possible if and only if $F(z)$ is k-regular. □

Theorem 1.29 is a simplified version of the following result of Dumas, which we will use in the proof of Theorem 1.32. Its proof can be attained by an argument almost identical to the proof of Theorem 1.29; for details see Dumas's Thesis [26, Theorem 24].

Theorem 1.30 (Dumas). *Let $F(z) \in \mathbb{C}[[z]]$ be a power series satisfying*

$$\sum_{i=0}^{d} a_i(z) F(z^{k^i}) = E(z),$$

where $0 \neq a_0(z) \in \mathbb{C}$, $a_1(z), \ldots, a_d(z) \in \mathbb{C}[z]$, and $E(z)$ is k-regular. Then $F(z)$ is k-regular.

Sometimes functions satisfying a Mahler functional equation with $a_0(z) = 1$ are called k-Becker; for example, see Adamczewski and Bell [1]. Becker conjectured that a result very similar to Theorem 1.29 holds for all regular functions.

Conjecture 1.31 (Becker). If $F(z)$ is a k-regular function, then there exists a k-regular rational function $r(z)$ such that the function $F(z)/r(z)$ satisfies a Mahler functional equation with $a_0(z) = 1$.

Theorem 1.32 (Structure Theorem, Dumas [26]). *A k-Mahler function is the quotient of a series and an infinite product which are k-regular. That is, if $F(z)$ is the* *solution of the Mahler functional equation*

$$a_0(z)F(z)+a_1(z)F(z^k)+\cdots+a_d(z)F(z^{k^d})=0,$$

where $a_0(z)a_d(z)\neq 0$, the $a_i(z)$ are polynomials, then there exists a k-regular series $H(z)$ such that

$$F(z)=\frac{H(z)}{\prod_{j\geqslant 0}\Gamma(z^{k^j})},$$

where $a_0(z)=\rho z^{\delta}\Gamma(z)$, with $\rho\neq 0$ and $\Gamma(0)=1$.

Proof. Suppose that $F(z)=\sum_{n\geqslant 0}f(n)z^n$ satisfies

$$a_0(z)F(z)+a_1(z)F(z^k)+\cdots+a_d(z)F(z^{k^d})=0,$$

where $a_0(z)a_d(z)\neq 0$, the $a_i(z)$ are polynomials and for each $i=0,\ldots,d$ let δ_i be the order of $a_i(z)$ at $z=0$, where we let $\delta_i=0$ if $a_i(z)=0$, and define the polynomials $b_i(z)$ by $a_i(z)=z^{\delta_i}b_i(z)$. Further, let

$$D:=\max\left\{\delta_0,\left\lfloor\frac{k\delta_0-\delta_1}{k-1}\right\rfloor,\left\lfloor\frac{k^2\delta_0-\delta_2}{k^2-1}\right\rfloor,\ldots,\left\lfloor\frac{k^d\delta_0-\delta_d}{k^d-1}\right\rfloor\right\},$$

and define the polynomial

$$p(z):=\sum_{n=0}^{D-\delta_0}f(n)z^n,$$

so that there is a power series $F_D(z)$ such that

$$F(z)=p(z)+z^{D-\delta_0+1}F_D(z). \tag{1.12}$$

Combining this with the Mahler functional equation and separating the $i=0$ term, we have

$$z^{D+1}b_0(z)F_D(z)=-\sum_{i=0}^{d}a_i(z)p(z^{k^i})-\sum_{i=0}^{d}z^{\lambda_i}b_i(z)F_D(z^{k^i}), \tag{1.13}$$

where

$$\lambda_i=\delta_i+k^i(D-\delta_0+1).$$

We claim that $\lambda_i\geqslant D+1$ for each $i=1,\ldots,d$. To see this, note that for each $i=0,\ldots,d$ we have

$$D\geqslant\left\lfloor\frac{k^i\delta_0-\delta_i}{k^i-1}\right\rfloor\geqslant\frac{k^i\delta_0-\delta_i}{k^i-1}+\frac{1}{k^i-1}-1,$$

which gives the desired lower bound on λ_i after some rearrangement.

Since each $\lambda_i\geqslant D+1$ and the lefthand side of (1.13) is divisible by z^{D+1}, we have that the polynomial $\sum_{i=0}^{d}a_i(z)p(z^{k^i})$ is also divisible by z^{D+1}, so we may write

$$\sum_{i=0}^{d} a_i(z)p(z^{k^i}) = z^{D+1}E(z)$$

for some polynomial $E(z)$. Thus we have that

$$b_0(z)F_D(z) = -E(z) - \sum_{i=0}^{d} z^{\lambda_i-(D+1)} b_i(z)F_D(z^{k^i}). \tag{1.14}$$

Now let ρ be the nonzero number such that

$$a_0(z) = z^{\delta_0} b_0(z) = \rho z^{\delta_0} \Gamma(z),$$

with $\Gamma(0) = 1$, and set

$$G(z) := F_D(z) \prod_{j\geqslant 0} \Gamma(z^{k^j}).$$

Thus we may write (1.14) as

$$G(z) = -\rho^{-1} E(z) \prod_{j\geqslant 1} \Gamma(z^{k^j}) - \rho^{-1} \sum_{i=0}^{d} z^{\lambda_i-(D+1)} \left(b_i(z) \prod_{j=0}^{i} \Gamma(z^{k^j}) \right) G(z^{k^i}). \tag{1.15}$$

The infinite product $P(z) := \prod_{j\geqslant 0} \Gamma(z^{k^j})$ is k-regular by Theorem 1.29 since it satisfies the Mahler functional equation

$$P(z) - \Gamma(z)P(z^k) = 0.$$

Combining this with Theorem 1.30, (1.15) gives that $G(z)$ is k-regular.

Using the definition of $G(z)$ and (1.12) we have

$$F(z) = p(z) + z^{D-\delta_0+1} \frac{G(z)}{\prod_{j\geqslant 0} \Gamma(z^{k^j})}.$$

Setting $H(z) := p(z)\prod_{j\geqslant 0} \Gamma(z^{k^j}) + z^{D-\delta_0+1}G(z)$, we have both that $H(z)$ is k-regular, since the set of k-regular functions form a ring, and also that

$$F(z) = \frac{H(z)}{\prod_{j\geqslant 0} \Gamma(z^{k^j})},$$

which is the desired result. □

1.3 Size and growth

The range of automatic sequences is finite, so questions of size and growth concerning automatic sequences are typically uninteresting. Regular sequences can take an infinite number of values. Three immediate questions that arise are: 1) how slow can can an unbounded regular sequence grow?; 2) are there good upper bounds for such sequences?; and, 3) what is the maximum possible growth?

1.3.1 Lower bounds

When considering the question of the growth of a regular sequence, from the lower bound perspective, it is worth noting that any such result will be an 'infinitely often' result at best. For example, there are regular sequences that are unbounded, yet take the value 1 infinitely. The Stern sequence s is a great witness to this property. As we have stated previously, $s(2^n+1)=n+1$, so that the Stern sequence is unbounded, yet also, $s(2^n)=1$ for all n. Similar results hold for the valuation function $\nu_k(n)$, which is the largest integer m such that k^m divides n; ν_k is clearly unbounded, and it takes each nonnegative integer value an infinite number of times.

In 2014, an 'infinitely often' lower bound type result was given by Bell, Coons, and Hare [10]. We present their result with proof here.

Theorem 1.33 (Bell, Coons, and Hare). *Let $k \geqslant 2$. If $f:\mathbb{N}\to\mathbb{Z}$ is an unbounded k-regular sequence, then there exists $c>0$ such that $|f(n)|>c\log n$ infinitely often.*

Lemma 1.34. *Let $k \geqslant 2$ be an integer, let $\mathbf{A}_0,\dots,\mathbf{A}_{k-1}$ be $d\times d$ integer matrices, and let $\mathscr{B}$ be the semigroup generated by $\mathbf{A}_0,\dots,\mathbf{A}_{k-1}$. Then either $\mathscr{B}$ is finite or there is some $\mathbf{S}\in\mathscr{B}$ and fixed vectors $\mathbf{v}$ and $\mathbf{w}\in\mathbb{C}^d$ such that $|\mathbf{w}^T\mathbf{S}^n\mathbf{v}|\geqslant n$ for all sufficiently large n.*

Proof. Suppose that $\mathscr{B}$ is infinite. Then since $\mathscr{B}$ is finitely generated, a result of McNaughton and Zalcstein [43] gives that there is some $\mathbf{S}$ in $\mathscr{B}$ such that the matrices $\mathbf{S},\mathbf{S}^2,\mathbf{S}^3,\dots$ are all distinct. Let $p(x)$ be the characteristic polynomial of $\mathbf{S}$. Then $p(x)$ is a monic integer polynomial. If $p(x)$ has a root λ that is strictly greater than 1 in modulus, then $\mathbf{S}$ has an eigenvector $\mathbf{v}$ such that $\mathbf{S}\mathbf{v}=\lambda\mathbf{v}$. Pick a nonzero vector $\mathbf{w}$ such that $\mathbf{w}^T\mathbf{v}=C\neq 0$. Then $|\mathbf{w}^T\mathbf{S}^n\mathbf{v}|=|C|\cdot|\lambda|^n\geqslant n$ for n sufficiently large.

If, on the other hand, all the roots of $p(x)$ are at most 1 in modulus, then all nonzero eigenvalues of $\mathbf{S}$ are algebraic integers with all conjugates having modulus 1, hence they are roots of unity. Let $\mathbf{Y}$ be a matrix in $\mathrm{GL}_d(\mathbb{C})$ such that $\mathbf{T}:=\mathbf{Y}^{-1}\mathbf{S}\mathbf{Y}$ is in Jordan form, where we take Jordan blocks to be upper-triangular. Then each Jordan block in $\mathbf{T}$ is of the form $\mathbf{J}_i(\lambda)$ with λ either zero or a root of unity and $i\geqslant 1$. Since $\mathbf{S}$ does not generate a finite subsemigroup of $\mathscr{B}$, there is some root of unity ω and some $m>1$ such that $\mathbf{T}$ has a block of the form $\mathbf{J}_m(\omega)$. We may assume, without loss of generality, that $\mathbf{J}_m(\omega)$ is the first block occurring in $\mathbf{T}$. Then the $(1,2)$-entry of $\mathbf{T}^n$ is $n\omega^{n-1}$ and so $|e_1^T\mathbf{T}^n e_2|=n$ for every n. In particular, we have

$$|e_1^T\mathbf{Y}^{-1}\mathbf{S}^n\mathbf{Y}e_2|\geqslant n$$

for every n. Taking $\mathbf{w}^T=e_1^T\mathbf{Y}^{-1}$ and $\mathbf{v}=\mathbf{Y}e_2$ gives the result. □

Proof (of Theorem 1.33). Let $k \geqslant 2$ be an integer, and suppose that $f : \mathbb{N} \to \mathbb{Z}$ is an unbounded k-regular sequence. Given a word $w = i_s \cdots i_0 \in \{0, \ldots, k-1\}^*$, as stated previously, we let $[w]_k$ denote the natural number $n = i_s k^s + \cdots + i_1 k + i_0$. The $\mathbb{Z}$-submodule of all $\mathbb{Z}$-valued sequences spanned by $\mathrm{Ker}_k(f)$ is a finitely generated torsion free module and hence free of finite rank. Let $\left\{\{g_1(n)\}_{n\geqslant 0}, \ldots, \{g_d(n)\}_{n\geqslant 0}\right\}$ be a $\mathbb{Z}$-module basis for the $\mathbb{Z}$-module spanned by $\mathrm{Ker}_k(f)$. Then for each $i \in \{0, 1, \ldots, k-1\}$, the functions $g_1(kn+i), \ldots, g_d(kn+i)$ can be expressed as $\mathbb{Z}$-linear combinations of $g_1(n), \ldots, g_d(n)$ and hence there are $d \times d$ integer matrices $\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}$ such that

$$[g_1(n), \ldots, g_d(n)]\mathbf{A}_i = [g_1(kn+i), \ldots, g_d(kn+i)]$$

for $i = 0, \ldots, k-1$ and all $n \geqslant 0$. In particular, if $i_s \cdots i_0$ is the base-k expansion of n, then $[g_1(0), \ldots, g_d(0)]\mathbf{A}_{i_s} \cdots \mathbf{A}_{i_0} = [g_1(n), \ldots, g_d(n)]$. (We note that this holds even if we pad the base-k expansion of n with zeros at the beginning.) We claim that the $\mathbb{Q}$-span of the vectors $[g_1(i), \ldots, g_d(i)]^T$, as i ranges over all natural numbers, must be all of $\mathbb{Q}^d$. Indeed, if this were not the case, then their span would be a proper subspace of $\mathbb{Q}^d$ and hence the span would have a non-trivial orthogonal complement. In particular, there would exist integers $c_1, \ldots, c_d$, not all zero, such that

$$c_1 g_1(n) + \cdots + c_d g_d(n) = 0$$

for every n, contradicting the fact that $g_1(n), \ldots, g_d(n)$ are linearly independent sequences.

Let $\mathscr{A}$ denote the semigroup generated by $\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}$. Then we have just shown that there exist words $\mathbf{X}_1, \ldots, \mathbf{X}_d$ in $\mathscr{A}$ such that

$$[g_1(0), \ldots, g_d(0)]\mathbf{X}_1, \ldots, [g_1(0), \ldots, g_d(0)]\mathbf{X}_d$$

span $\mathbb{Q}^d$. Now, if $\mathscr{A}$ is finite, then $\{g_1(n)\}_{n\geqslant 0}, \ldots, \{g_d(n)\}_{n\geqslant 0}$ take only finitely many distinct values. Since $\{f(n)\}_{n\geqslant 0}$ is a $\mathbb{Z}$-linear combination of $\{g_1(n)\}_{n\geqslant 0}, \ldots,$ $\{g_d(n)\}_{n\geqslant 0}$, we see that it too takes only finitely many distinct values, which contradicts our assumption that it is unbounded. Thus $\mathscr{A}$ must be infinite. By Lemma 1.34, there exist $\mathbf{Y} \in \mathscr{A}$ and vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^d$ such that $|\mathbf{x}^T\mathbf{Y}^n\mathbf{y}| \geqslant n$ for all n sufficiently large.

By construction, we may write $\mathbf{x}^T = \sum_j \alpha_j [g_1(0), \ldots, g_d(0)]\mathbf{X}_j$ for some complex numbers α_j. Then

$$\mathbf{x}^T\mathbf{Y}^n = \sum_j \alpha_j [g_1(0), \ldots, g_d(0)]\mathbf{X}_j\mathbf{Y}^n.$$

Let u_j be the word in $\{0, 1, \ldots, k-1\}^*$ corresponding to $\mathbf{X}_j$ and let y be the word in $\{0, \ldots, k-1\}^*$ corresponding to $\mathbf{Y}$; that is $u_j = i_s \cdots i_0$ where $\mathbf{X}_j = \mathbf{A}_{i_s} \cdots \mathbf{A}_{i_0}$ and similarly for y. Then we have

$$[g_1(0), \ldots, g_d(0)]\mathbf{X}_j\mathbf{Y}^n = [g_1([u_j y^n]_k), \ldots, g_d([u_j y^n]_k)]^T.$$

Write $\mathbf{y}^T = [\beta_1, \ldots, \beta_d]$. Then

$$\mathbf{x}^T \mathbf{Y}^n \mathbf{y} = \sum_{i,j} \alpha_i \beta_j g_j([u_i y^n]_k).$$

By assumption, each of $\{g_1(n)\}_{n \geqslant 0}, \ldots, \{g_d(n)\}_{n \geqslant 0}$ is in the $\mathbb{Z}$-module generated by $\mathrm{Ker}_k(f)$, and hence there exist natural numbers $p_1, \ldots, p_t$ and $q_1, \ldots, q_t$ with $0 \leqslant q_m < k^{p_m}$ for $m = 1, \ldots, t$ such that for each $s = 1, \ldots, d$ we have $g_s(n) = \sum_{i=1}^t \gamma_{i,s} f(k^{p_i} n + q_i)$ for some integers $\gamma_{i,s}$. Then

$$\mathbf{x}^T \mathbf{Y}^n \mathbf{y} = \sum_{i,j,\ell} \alpha_i \beta_j \gamma_{\ell,j} f([u_i y^n v_\ell]_k),$$

where v_ℓ is the unique word in $\{0, 1, \ldots, k-1\}^*$ of length p_ℓ such that $[v_\ell]_k = q_\ell$. Let $K = \sum_{i,j,\ell} |\alpha_i| \cdot |\beta_j| \cdot |\gamma_{\ell,j}|$. Then since $|\mathbf{x}^T \mathbf{Y}^n \mathbf{y}| \geqslant n$ for all n sufficiently large, there is some $N_0 > 0$ such that for $n > N_0$ some element from

$$\big\{ \{|f([u_i y^n v_j]_k)|\}_{n \geqslant 0} \colon i = 1, \ldots, d, j = 1, \ldots, t \big\}$$

is at least n/K.

We let M denote the maximum of the lengths of $u_1, \ldots, u_d, y, v_1, \ldots, v_t$. Then each of $[u_i y^n v_j]_k < k^{2Mn}$ for $n \geqslant 2$. Hence we have constructed an infinite set of natural numbers $N = N_n := [u_i y^n v_j]_k$ such that $|f(N)| > \log_k(N)/2K$ and so taking $c = (2MK \log k)^{-1}$, we see that $|f(N)| > c \log N$ for infinitely many N. $\square$

The above proof actually shows something a bit more specific. It shows for an unbounded k-regular sequence, that there exist words $u_1, \ldots, u_m, y, v_1, \ldots, v_m \in \{0, 1, \ldots, k-1\}^*$ and a constant $c_0 > 0$ such that for all sufficiently large n there exist an i and j such that $|f([u_i y^n v_j]_k)| \geqslant c_0 n$. Here for a word $w = i_s \cdots i_0 \in \{0, 1, \ldots, k-1\}^*$, we have written $[w]_k = i_s k^s + \cdots + i_0$. This can be thought of as a type of "pumping lemma" for attaining unbounded growth. This argument will prove quite useful when we consider good upper bounds in the next section.

1.3.2 Upper bounds

The question of upper bounds was first addressed by Allouche and Shallit [4, Theorem 2.10] in their original paper introducing regular sequences.

Theorem 1.35 (Allouche and Shallit). *Let f be a k-regular sequence with values in $\mathbb{C}$. Then there is a constant c such that $f(n) = O(n^c)$.*

Proof. We use the matrix version of regular sequences as given by Lemma 1.15. In particular, let d be a positive integer, $\mathscr{A}_f = \{\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}\} \subseteq \mathbb{C}^{d \times d}$, and $\mathbf{v}, \mathbf{w} \in \mathbb{C}^d$ be vectors such that

$$f(n) = \mathbf{w}^T \mathbf{A}_{i_0} \cdots \mathbf{A}_{i_s} \mathbf{v},$$

where $(n)_k = i_s \cdots i_0$ is the base-k expansion of n.

Let $\|\cdot\|$ be a (submultiplivative) matrix norm, $i_0 \cdots i_s$ be the base k expansion of n, and

$$c := \max\{\|\mathbf{v}\|, \|\mathbf{w}\|, \|\mathbf{A}_0\|, \ldots, \|\mathbf{A}_{k-1}\|\}.$$

Then

$$|f(n)| \leqslant \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cdot \prod_{j=0}^{s} \|\mathbf{A}_{i_j}\| \leqslant c^{s+3}.$$

Using the bound $s \leqslant \log_k n$ with some rearrangement gives the result. □

In recent work, Coons [22] determined the optimal constant c for which Theorem 1.35 holds. Its description requires a few definitions, the first of which formalises what is meant by 'optimal' in this context.

Definition 1.36. Let $k \geqslant 1$ be an integer and $f : \mathbb{Z}_{\geqslant 0} \to \mathbb{C}$ be a (not eventually zero) k-regular sequence. We define the *growth exponent of* f, denoted $\mathrm{GrExp}(f)$, by

$$\mathrm{GrExp}(f) := \limsup_{\substack{n\to\infty \\ f(n)\neq 0}} \frac{\log|f(n)|}{\log n}.$$

Definition 1.37. The *spectral radius* of a square matrix is the maximal absolute value of eigenvalues of the matrix. The *joint spectral radius* of a finite set of matrices $\mathscr{A} = \{\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{k-1}\}$, denoted $\rho(\mathscr{A})$, is defined as the real number

$$\rho(\mathscr{A}) = \limsup_{n\to\infty} \max_{0\leqslant i_0, i_1, \ldots, i_{n-1} \leqslant k-1} \left\|\mathbf{A}_{i_0}\mathbf{A}_{i_1} \cdots \mathbf{A}_{i_{n-1}}\right\|^{1/n},$$

where $\|\cdot\|$ is any (submultiplicative) matrix norm.

The joint spectral radius was introduced by Rota and Strang [50] and has a wide range of applications. See Rota and Strang [50] also for details about the independence of the matrix norm in the definition. For an extensive treatment, see Jungers's monograph [32].

Theorem 1.38 (Coons). *Let $k \geqslant 1$ and $d \geqslant 1$ be integers and $f : \mathbb{Z}_{\geqslant 0} \to \mathbb{C}$ be a (not eventually zero) k-regular sequence. If $\mathscr{A}_f$ is any collection of k integer matrices associated to a basis of the $\mathbb{C}$-vector space $\langle \mathrm{Ker}_k(f)\rangle_{\mathbb{C}}$, then*

$$\log_k \rho(\mathscr{A}_f) = \mathrm{GrExp}(f),$$

where $\log_k$ denotes the base-k logarithm.

Before moving on with the needed preliminary results for the proof of this theorem, we describe what it means for a collection of k integer matrices to be associated to a basis of the $\mathbb{C}$-vector space $\langle \mathrm{Ker}_k(f)\rangle_{\mathbb{C}}$. This is all taken in the context of Lemma 1.15 that provides for a set of matrices $\mathscr{A}_f$ coming from Lemma 1.14(e). In particular, given a word $w = i_s \cdots i_0 \in \{0, \ldots, k-1\}^*$, we let $[w]_k$ denote the natural number n such that $(n)_k = w$. Let $\{\{f(n)\}_{n\geqslant 0} = \{g_1(n)\}_{n\geqslant 0}, \ldots, \{g_d(n)\}_{n\geqslant 0}\}$ be a basis for the $\mathbb{C}$-vector space $\langle \mathrm{Ker}_k(f)\rangle_{\mathbb{C}}$. Then for each $i \in \{0, 1, \ldots, k-1\}$, the sequences $\{g_1(kn+i)\}_{n\geqslant 0}, \ldots, \{g_d(kn+i)\}_{n\geqslant 0}$ can be expressed as $\mathbb{C}$-linear combinations of $\{g_1(n)\}_{n\geqslant 0}, \ldots, \{g_d(n)\}_{n\geqslant 0}$ and hence there is a set of $d\times d$ matrices $\mathscr{A}_f = \{\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}\}$ with entries in $\mathbb{C}$ such that

$$\mathbf{A}_i[g_1(n), \ldots, g_d(n)]^T = [g_1(kn+i), \ldots, g_d(kn+i)]^T$$

for $i = 0, \ldots, k-1$ and all $n \geqslant 0$. In particular, if $i_s \cdots i_0$ is the base-k expansion of n, then $\mathbf{A}_{i_0} \cdots \mathbf{A}_{i_s}[g_1(0), \ldots, g_d(0)]^T = [g_1(n), \ldots, g_d(n)]^T$. (We note that this holds even if we pad the base-k expansion of n with zeros at the beginning.)

Definition 1.39. We call a set of matrices $\mathscr{A}_f$, as constructed in the previous paragraph, *a set of matrices associated to a basis of* $\langle \mathrm{Ker}_k(f)\rangle_{\mathbb{C}}$. In general, if $\mathscr{B}_f$ is any set of matrices for which there are vectors $\mathbf{w}$ and $\mathbf{v}$ such that f has linear representation $(\mathbf{w}, \mathscr{B}_f, \mathbf{v})$, then we call the set $\mathscr{B}_f$ a *set of matrices associated to* f.

The first step in the proof of Theorem 1.38 is to modify the proof of Theorem 1.35 to include the notion of the joint spectral radius. This is done by appealing to a result, which we record here as Lemma 1.40; it can be found as Proposition 4 of Blondel et al. [13], though it was first given in the original paper of Rota and Strang [50].

Lemma 1.40. *Let $k \geqslant 1$ be an integer and $\mathscr{A} = \{\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{k-1}\}$ be a finite set of matrices. Given $\varepsilon > 0$ then there is a submultiplicative matrix norm $\|\cdot\|$ such that $\|\mathbf{A}_i\| < \rho(\mathscr{A}) + \varepsilon$ for each $i \in \{0, 1, \ldots, k-1\}$.*

With this lemma in hand, it is quite easy to give a tight upper bound for the optimal constant for Theorem 1.35.

Proposition 1.41. *Let $k \geqslant 2$ be an integer and $f : \mathbb{Z}_{\geqslant 0} \to \mathbb{C}$ be a k-regular function. For any $\varepsilon > 0$, there is a constant $c = c(\varepsilon) > 0$ such that for all $n \geqslant 1$,*

$$\frac{|f(n)|}{n^{\log_k(\rho(\mathscr{A}_f)+\varepsilon)}} \leqslant c,$$

where $\mathscr{A}_f$ is any set of matrices associated to f.

Proof. Let $\varepsilon > 0$ be given and let $\|\cdot\|$ be a matrix norm such that the conclusion of Lemma 1.40 holds. Then

$$|f(n)| \leqslant \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cdot \prod_{j=0}^{s} \|\mathbf{A}_{i_j}\| \leqslant \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cdot (\rho(\mathscr{A}) + \varepsilon)^{s+1},$$

where the base-k expansion of n is $i_s \cdots i_0$. Using the bound $s \leqslant \log_k n$ with some rearrangement gives the result. □

As it turns out, if $\mathscr{B}_f$ is any set of matrices associated to f and $\mathscr{A}_f$ is any set of matrices associated to a basis of $\langle \mathrm{Ker}_k(f)\rangle_{\mathbb{C}}$, then $\rho(\mathscr{A}_f) \leqslant \rho(\mathscr{B}_f)$, though the proof of this statement is only apparent after validating Theorem 1.38.

Lemma 1.42. *Let $k \geqslant 1$ be an integer and $\mathscr{A} = \{\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_{k-1}\}$ be a finite set of matrices. If $\varepsilon > 0$ is a real number, then there is a positive integer m and a matrix $\mathbf{A}_{i_0} \cdots \mathbf{A}_{i_{m-1}}$, such that*

$$(\rho(\mathscr{A}) - \varepsilon)^m < \rho(\mathbf{A}_{i_0} \cdots \mathbf{A}_{i_{m-1}}) < (\rho(\mathscr{A}) + \varepsilon)^m.$$

Proof. By using the properties of limits, this is a direct consequence of the definition of the joint spectral radius. Details are left as an exercise. □

Restricting to a set of matrices associated to a basis of $\langle \mathrm{Ker}_k(f) \rangle_{\mathbb{C}}$ allows us to provide the lower bound analogue of Proposition 1.41.

Proposition 1.43. *Let $k \geqslant 2$ be an integer and $f : \mathbb{Z}_{\geqslant 0} \to \mathbb{C}$ be a k-regular function. For any $\varepsilon > 0$, there is a constant $c = c(\varepsilon) > 0$ such that for infinitely many $n \geqslant 1$,*

$$\frac{|f(n)|}{n^{\log_k(\rho(\mathscr{A}_f) - \varepsilon)}} \geqslant c,$$

where $\mathscr{A}_f$ is any set of matrices associated to a basis of $\langle \mathrm{Ker}_k(f) \rangle_{\mathbb{C}}$.

Proof. As in the proof of Theorem 1.33, we follow the argument of Bell, Coons, and Hare (see p. 198 of [10]).

Let $k \geqslant 2$ be an integer, suppose that $f : \mathbb{Z}_{\geqslant 0} \to \mathbb{C}$ is an unbounded k-regular sequence, and $\mathscr{A}_f = \{\mathbf{A}_0, \ldots, \mathbf{A}_{k-1}\}$ be a set of matrices associated to a basis $\{\{f(n)\}_{n \geqslant 0} = \{g_1(n)\}_{n \geqslant 0}, \ldots, \{g_d(n)\}_{n \geqslant 0}\}$ of the $\mathbb{C}$-vector space $\langle \mathrm{Ker}_k(f) \rangle_{\mathbb{C}}$.

Let $\varepsilon > 0$ be given. Then by Lemma 1.42 there is a positive integer m and a matrix $\mathbf{A} = \mathbf{A}_{i_0} \cdots \mathbf{A}_{i_{m-1}}$ such that $\rho(\mathbf{A}) > (\rho(\mathscr{A}_f) - \varepsilon)^m$. Let λ be an eigenvalue of $\mathbf{A}$ with $|\lambda| = \rho(\mathbf{A})$. Then there is an eigenvector $\mathbf{y}$ such that $\mathbf{A}\mathbf{y} = \lambda \mathbf{y}$. Pick a vector $\mathbf{x}$ such that $\mathbf{x}^T \mathbf{y} = c_1 \neq 0$. Then

$$\left| \mathbf{x}^T \mathbf{A}^n \mathbf{y} \right| = |c_1| \cdot |\lambda|^n = |c_1| \cdot \rho(\mathbf{A})^n > |c_1| \cdot \left(\rho(\mathscr{A}_f) - \varepsilon \right)^{nm}. \tag{1.16}$$

We claim that the $\mathbb{C}$-span of the vectors $[g_1(i), \ldots, g_d(i)]$, as i ranges over all natural numbers, must span all of $\mathbb{C}^d$. If this were not the case, then their span would be a proper subspace of $\mathbb{C}^d$ and hence the span would have a non-trivial orthogonal complement. In particular, there would exist $c_1, \ldots, c_d \in \mathbb{C}$, not all zero, such that

$$c_1 g_1(n) + \cdots + c_d g_d(n) = 0$$

for every n, contradicting the fact that $g_1(n), \ldots, g_d(n)$ are $\mathbb{C}$-linearly independent sequences.

Let $\langle \mathscr{A}_f \rangle$ denote the semigroup generated by the elements of $\mathscr{A}_f$. We have just shown that there exist words $\mathbf{X}_1, \ldots, \mathbf{X}_d$ in $\langle \mathscr{A}_f \rangle$ such that

$$[g_1(0), \ldots, g_d(0)]\mathbf{X}_1, \ldots, [g_1(0), \ldots, g_d(0)]\mathbf{X}_d$$

span $\mathbb{C}^d$.

Now consider $\mathbf{x}^T\mathbf{A}^n\mathbf{y}$ as described in the paragraph ending with (1.16). The following lines are as in the proof of Theorem 1.33. By construction, we may write $\mathbf{x}^T = \sum_j \alpha_j[g_1(0),\ldots,g_d(0)]\mathbf{X}_j$ for some complex numbers α_j. Then

$$\mathbf{x}^T\mathbf{A}^n = \sum_j \alpha_j[g_1(0),\ldots,g_d(0)]\mathbf{X}_j\mathbf{A}^n.$$

Let u_j be the word in $\{0,1,\ldots,k-1\}^*$ corresponding to $\mathbf{X}_j$ and let $y = i_{m-1}\cdots i_0$ be the word in $\{0,\ldots,k-1\}^*$ corresponding to $\mathbf{A}$; that is $y = i_{m-1}\cdots i_0$ where $\mathbf{A} = \mathbf{A}_{i_0}\cdots\mathbf{A}_{i_{m-1}}$ and similarly for u_j. Then we have

$$[g_1(0),\ldots,g_d(0)]\mathbf{X}_j\mathbf{A}^n = [g_1([u_jy^n]_k),\ldots,g_d([u_jy^n]_k)]^T.$$

Write $\mathbf{y}^T = [\beta_1,\ldots,\beta_d]$. Then

$$\mathbf{x}^T\mathbf{A}^n\mathbf{y} = \sum_{i,j}\alpha_i\beta_j g_j([u_iy^n]_k).$$

By assumption, each of $\{g_1(n)\}_{n\geqslant 0},\ldots,\{g_d(n)\}_{n\geqslant 0}$ is in the $\mathbb{C}$-vector space generated by $\mathrm{Ker}_k(f)$, and hence there exist natural numbers $p_1,\ldots,p_t$ and $q_1,\ldots,q_t$ with $0\leqslant q_\ell < k^{p_\ell}$ for $\ell = 1,\ldots,t$ such that each of for $j=1,\ldots,d$, we have $g_j(n) = \sum_{\ell=1}^t \gamma_{\ell,j} f(k^{p_\ell}n+q_\ell)$ for some constants $\gamma_{\ell,j}\in\mathbb{C}$. Then

$$\mathbf{x}^T\mathbf{A}^n\mathbf{y} = \sum_{i,j,\ell}\alpha_i\beta_j\gamma_{\ell,j} f([u_iy^nv_\ell]_k),$$

where v_ℓ is the unique word in $\{0,1,\ldots,k-1\}^*$ of length p_ℓ such that $[v_\ell]_k = q_\ell$. Let $K = \sum_{i,j,\ell}|\alpha_i|\cdot|\beta_j|\cdot|\gamma_{\ell,j}|$. Then since $|\mathbf{x}^T\mathbf{A}^n\mathbf{y}| \geqslant |c_1|\cdot\big(\rho(\mathscr{A}_f)-\varepsilon\big)^{nm}$ for all n, some element from

$$\big\{|f([u_iy^nv_\ell]_k)| : i=1,\ldots,d,\ell=1,\ldots,t\big\}\big\}$$

is at least $(|c_1|/K)\cdot\big(\rho(\mathscr{A}_f)-\varepsilon\big)^{nm}$ for each n. Set $c_2 := |c_1|/K$.

If $M = \max\{|u_i|,|v_\ell| : i=1,\ldots,d,\ \ell=1,\ldots,t\}$, then

$$N = [u_i(i_{m-1}\cdots i_0)^n v_\ell]_k < k^{2M+nm},$$

so that $\log_k(N) - 2M < nm$. Thus, by the finding of the previous paragraph, there are infinitely many N such that

$$\frac{|f(N)|}{N^{\log_k(\rho(\mathscr{A}_f)-\varepsilon)}} = \frac{|f(N)|}{(\rho(\mathscr{A}_f)-\varepsilon)^{\log_k N}} > \frac{c_2}{(\rho(\mathscr{A}_f)-\varepsilon)^{2M}},$$

which is the desired result. □

Proof (of Theorem 1.38). For a given $\varepsilon > 0$, Proposition 1.41 implies that

$$\lim_{n\to\infty}\frac{|f(n)|}{n^{\log_k(\rho(\mathscr{A}_f)+2\varepsilon)}} = 0,$$

and Proposition 1.43 implies that

$$\limsup_{n\to\infty} \frac{|f(n)|}{n^{\log_k(\rho(\mathscr{A}_f)-2\varepsilon)}} = \infty.$$

Taken together these give

$$\log_k(\rho(\mathscr{A}_f)-2\varepsilon) \leqslant \mathrm{GrExp}(f) \leqslant \log_k(\rho(\mathscr{A}_f)+2\varepsilon).$$

Since ε can be taken arbitrarily small, this proves the theorem. □

Example 1.44. For the Stern sequence s, one has

$$\mathrm{GrExp}(s) = \log_2 \varphi,$$

where $\varphi = (1+\sqrt{5})/2$ is the golden ratio. This follows from work of Reznick [49, Theorem 5.13]. See also, Calkin and Wilf [19] and Coons and Tyler [24].

Before moving on, we note the works of Dumas [27, 28] concerning the asymptotic expansion of the summatory functions of regular sequences. Among many results and useful algorithms, his results have the flavour of the following theorem [27, Theorem 1].

Theorem 1.45 (Dumas). *Let f be a k-regular function with linear representation $(\mathbf{w}, \mathscr{A}_f, \mathbf{v})$. Then*

$$s(n) := \sum_{j\leqslant n} f(n) \sim \sum_{\substack{\alpha>\alpha_* \\ \ell\geqslant 0,\vartheta}} n^{\alpha} \binom{\log_k n}{\ell} \exp(i\vartheta \log_k n)\Psi_{\alpha,\ell,\vartheta}(\log_k n) + O(n^{\alpha_*}),$$

where exponents α, angular variables ϑ are real numbers, the numbers ℓ are nonnegative integers, and the functions Ψ are 1-periodic functions. Specific details can be found in Dumas' work [27].

1.3.3 Maximum values and the finiteness property

Determining the maximum values of regular sequences remains a mysterious area, though it is related to a very interesting and important open question regarding the joint spectral radius. As examples and results surrounding this area are sparse, in this section, we will present a motivating extended example—Stern's sequence—as a way to frame some questions.

Recall from Example 1.13 that Stern's diatomic sequence is 2-regular and defined by the relations $s(0) = 0$, $s(1) = 1$, and for $n \geqslant 0$, by

$$s(2n) = s(n), \quad \text{and} \quad s(2n+1) = s(n) + s(n+1).$$

The first few values of the sequence are

$$0,1,1,2,1,3,2,3,1,4,3,5,2,5,3,4,1,5,4,7,3,8,5,7,2,7,5,8,3,7,4,5,1,\ldots.$$

The Stern sequence, like essentially all observed regular sequences, has a limiting distribution between consecutive powers of 2 (powers of k^r for k-regular sequences for some appropriate r). In fact, if one looks at the plot of the points $(n, s(n))$ for n between consecutive powers of 2, the picture seems to have asymptotically stabilised; see Figures 1.2 and 1.3.

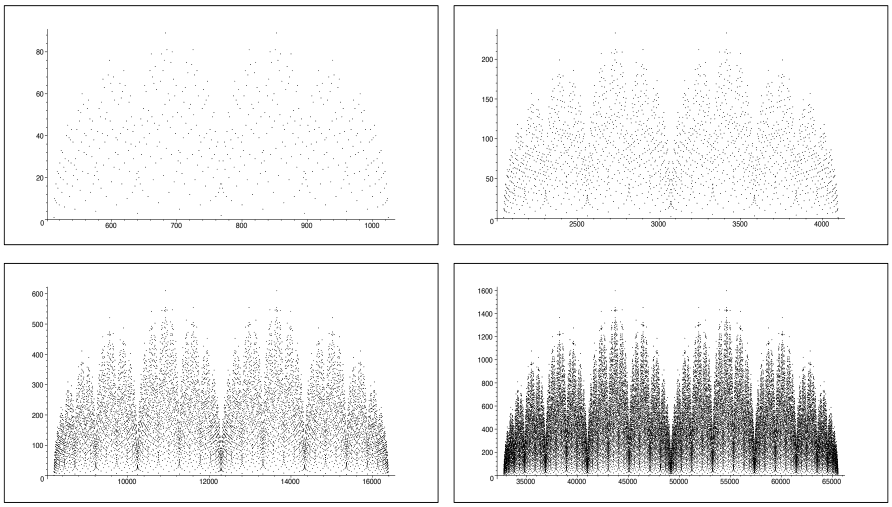

Fig. 1.2 Stern's diatomic sequence in the intervals $\left[2^n, 2^{n+1}\right]$ for $n = 9, 11, 13, 15$.

In particular, note the stabilising two maximums in the each of the plots in Figures 1.2 and 1.3. It is easy to show that the Stern sequence is palindromic between consecutive powers of 2, so we may focus on just the first maximum. (The maximum is in fact attained at at most two points, which we state here without proof.) We will use the defining recursions to classify get a bound on this maximum value.

To this end, for $m \geqslant 0$ define

$$M_m := \max\{s(n) : n \in \left[2^m, 2^{m+1}\right)\}.$$

Then, by observation, we have that $M_0 = 1$, $M_1 = 2$, and $M_2 = 3$.

For $m \geqslant 3$ we note that $s(2n+1) \geqslant s(n) = s(2n)$, so that the maximum value always occurs at an odd index $2n+1 \in \left[2^m, 2^{m+1}\right)$. Of course, like for all numbers, for this value $2n+1$, one of n or $n+1$ is even, so that the recursion for odd indices gives

$$M_m \leqslant M_{m-1} + M_{m-2}.$$

But, combining this inequality with the fact that $M_0 = 1$ and $M_1 = 2$, gives that

$$M_m \leqslant F_{m+2},$$

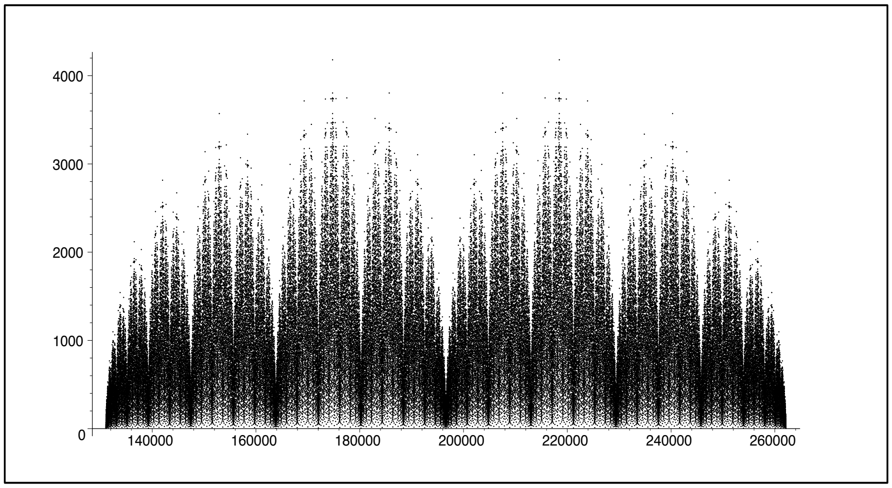

Fig. 1.3 Stern's diatomic sequence in the interval $[2^{17}, 2^{18}]$.

where F_k is the k-th Fibonacci number. This inequality is actually an equality, which we will now show.

Proposition 1.46. *The maximal value of the Stern sequence in the interval* $\left[2^m, 2^{m+1}\right)$ *is the Fibonacci number* F_{m+2} *and this value occurs at* $n = \left(2^{m+2} - (-1)^{m+2}\right)/3$.

Proof. We have already shown above that $M_m \leqslant F_{m+2}$, so it remains only to show that there is an integer $n \in \left[2^m, 2^{m+1}\right)$ such that $s(n) = F_{m+1}$.

To this end, set $\alpha_m := \left(2^{m+2} - (-1)^{m+2}\right)/3$. It is clear that $\alpha_m \in \left[2^m, 2^{m+1}\right)$ and that

$$\alpha_{m+1} = \begin{cases} 2\alpha_m + 1 & \text{if } m \text{ is even;} \\ 2\alpha_m - 1 & \text{if } m \text{ is odd,} \end{cases}$$

therefore by the recurrence for s we have

$$\begin{aligned} s(\alpha_{m+1}) &= \begin{cases} s(\alpha_m) + s(\alpha_m + 1) = s(\alpha_m) + s(2\alpha_{m-1}) & \text{if } m \text{ is even;} \\ s(\alpha_m - 1) + s(\alpha_m) = s(2\alpha_{m-1}) + s(\alpha_m) & \text{if } m \text{ is odd} \end{cases} \\ &= s(\alpha_{m-1}) + s(\alpha_m). \end{aligned}$$

By induction, it follows that $s(\alpha_m) = F_{m+2}$, which is exactly what we set out to show. □

The binary forms

$$\alpha_m := \begin{cases} \left[(\mathtt{10})^{m/2}\mathtt{1}\right]_2 & \text{if } m \text{ is even;} \\ \left[(\mathtt{10})^{(m-1)/2}\mathtt{11}\right]_2 & \text{if } m \text{ is odd} \end{cases}$$

of the integers α_m here are a point of interest. They are of the form $w^k u$ for some words u and w and some integer k. This implies something even more interesting for the normalised graph of the Stern sequence between consecutive powers of two. To be clear, we state the generalisations of these ideas as a series of formal questions.

Question 1.47. Let f be a k-regular sequence. Is there an integer $M \geqslant 1$, such that f (suitably normalised to the box $[0,1]^2$), has a limit when taken between powers of k^M. That is, the normalised picture of the points $(n, f(n))$, where $n \in [k^{Mj}, k^{M(j+1)}]$, converges.

Question 1.48. Let f be an integer-valued k-regular sequence. If f is not an automatic sequence, is there a positive integer M such that

$$\max_{k^{Mm} \leqslant n \leqslant k^{M(m+1)}-1} |f(n)| < \max_{k^{M(m+1)} \leqslant n \leqslant k^{M(m+2)}-1} |f(n)| \;?$$

Question 1.49. Suppose that Question 1.3.3 has a positive answer and that f is an integer-valued k-regular sequence. Is it true that there are words $u, w \in \{0, \ldots, k-1\}^*$ such that one of the maximum values $\alpha_{f,m}$ of $|f(n)|$ in $[k^{Mm}, k^{M(m+1)}]$ satisfies $\alpha_{f,m} = w^{n_m} u$ for some increasing sequence of integers n_m and infinitely many m.

The careful reader will notice that Questions and 1.49 have the added assumption that f is integer-valued. This assumption cannot be removed completely as the questions have negative answers when one looks at general real-valued sequence. In fact, this line of questioning is related to an open question regarding the joint spectral radius (see Definition 1.37) of a finite set of matrices.

Definition 1.50. A finite set of matrices $\mathscr{A}$ is said to have the *finiteness property* provided there is a specific finite product $\mathbf{A}_{i_0} \cdots \mathbf{A}_{i_{m-1}}$ of matrices from $\mathscr{A}$ such that $\rho(\mathbf{A}_{i_0} \cdots \mathbf{A}_{i_{m-1}})^{1/m} = \rho(\mathscr{A})$.

Arising from the work of Daubechies and Lagarias [25], Lagarias and Wang [35] conjectured that the finiteness property holds for all finite sets of real matrices, though this was shown to be false—hence the negative answer to the generalisation of Question 1.49 for real-valued regular sequences. The existence of counterexamples was first shown by Bousch and Mairesse [16] (see also [14, 34]), and a constructive counterexample was recently given by Hare, Morris, Sidorov and Theys [30]. Their counterexample is reminiscent of the Stern sequence, and so we give it here to add a little connective flavour to the questions.

Example 1.51 (Hare, Morris, Sidorov and Theys). Let τ denote the sequence of integers defined by $\tau_0 = 1$, $\tau_1, \tau_2 = 2$, and $\tau_{n+1} = \tau_n \tau_{n-1} - \tau_{n-2}$ for all $n \geqslant 2$, and let F_n be the nth Fibonacci number for $n \geqslant 0$. Define the real number $\alpha_* \in (0,1]$ by

$$\alpha_* := \prod_{n \geqslant 1} \left(1 - \frac{\tau_{n-1}}{\tau_n \tau_{n+1}}\right)^{(-1)^n F_{n+1}}.$$

Then this infinite product converges unconditionally, and the set

$$\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \alpha_* \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right\}$$

does not have the finiteness property.

Note that the number

$$\alpha_* = 0.749326546330367557943961948091344672091327370236064317 3\ldots,$$

and it is unknown if α_* is irrational, though it is suspected.

It is an open and interesting question to determine if all finite sets of rational matrices satisfy the finiteness property. The current best result towards this conjecture is that of Jungers and Blondel [33], who showed that the finiteness property holds for all finite sets of rational matrices provided it holds for all pairs of matrices with entries in $\{-1,0,1\}$. Restricting to the case of nonnegative rational matrices, Jungers and Blondel [33] could reduce $\{-1,0,1\}$ to the set $\{0,1\}$.

As a fact related to Question 1.47, we want to show that there are only few large values of $s(n)$ in the interval $\left[2^m, 2^{m+1}\right)$, compared to the maximal value $M_m = F_{m+2}$. First, we note that the mean value of $s(n)$ in such an interval equals $(3/2)^m$, which can be proved by induction. What we want to show is that there are in fact exponentially few integers n in $\left[2^m, 2^{m+1}\right)$ such that $s(n) \geqslant \varepsilon M_m$, for any $\varepsilon > 0$. By definition of the mean value and the nonnegativity of $s(n)$, the number N of such integers satisfies $N\varepsilon M_m/2^m \leqslant (3/2)^m$, therefore $N \leqslant 3^m/(M_m\varepsilon) \ll (3/\varphi)^m/\varepsilon$, where φ is the golden ratio. Since φ is strictly larger than $3/2$, there are exponentially few integers n such that $s(n)$ is large. This leads us to the following proposition for the graph of $s(n)$ in dyadic intervals $\left[2^m, 2^{m+1}\right)$, normalized to $[0,1]^2$. We define functions f_m from $[0,1]$ to $[0,1]$ by

$$f_m(x) = \frac{1}{F_{m+2}} s\big(2^m + \lfloor 2^m x \rfloor\big).$$

Proposition 1.52. *The sequence $\{f_m\}_{m \geqslant 0}$ of functions converges to zero almost everywhere.*

Proof. By the above considerations there is an $K < 1$ such that

$$\lambda(\{x \in [0,1] : f_m(x) \geqslant \varepsilon\}) \leqslant K^m/\varepsilon,$$

where λ is the Lebesgue measure. It follows that

$$\begin{aligned}
&\lambda\big(\{x \in [0,1] : \exists m \geqslant M \text{ such that } f_m(x) \geqslant \varepsilon\}\big) \\
&\quad = \lambda\left(\bigcup_{m \geqslant M} \{x \in [0,1] : f_m(x) \geqslant \varepsilon\}\right) \leqslant \sum_{m \geqslant M} \lambda\big(\{x \in [0,1] : f_m(x) \geqslant \varepsilon\}\big) \\
&\qquad \leqslant \frac{1}{\varepsilon} \sum_{m \geqslant M} K^m = \frac{1}{\varepsilon} \frac{K^M}{1-K}.
\end{aligned}$$

Setting $A_M(\varepsilon) = \{x \in [0,1] : f_m(x) < \varepsilon \text{ for all } m \geqslant M\}$, we obtain $\lambda(A_M) \geqslant 1 - K^M/(\varepsilon(1-K))$. It follows that

$$1 = \lambda\left(\bigcup_{M\geqslant 1} A_M\right) = \lambda(B_\varepsilon),$$

where

$$B_\varepsilon = \{x \in [0,1] : \exists M \geqslant 1 \text{ such that } f_m(x) < \varepsilon \text{ for all } m \geqslant M\}.$$

Therefore

$$\lambda(\{x \in [0,1] : f_m(x) \to 0 \text{ as } m \to \infty\}) = \lambda\left(\bigcap_{\varepsilon>0} B_\varepsilon\right) = \lambda\left(\bigcap_{n\geqslant 1} B_{1/n}\right) = 1. \quad \square$$

In fact, we conjecture the following more precise statement.

Conjecture 1.53. The sequence $\{f_m\}_{m\geqslant 0}$ of functions converges pointwise and the limit is nonzero if and only if $x \in [0,1]$ is of the form $x = a/(3\cdot 2^s)$ for some integers $a \geqslant 1$ and $s \geqslant 0$.

Another interesting question concerns values of $s(n)$ near the mean value $(3/2)^m$. Lansing [36] studies the quantity

$$H(\lambda,m) = \frac{1}{2^m}\left|\left\{2^m \leqslant n < 2^{m+1} : s(n) \geqslant \lambda(3/2)^m\right\}\right|$$

and notes that the data "suggests that $H(\lambda,m)$ converges to a smooth function, but it is not clear if it actually does." This statement is based on the behaviour for some small values of m. We used randomly chosen integers in the interval $\left[2^m, 2^{m+1}\right)$ for some larger m in order to guess the asymptotic behaviour. Our experiments suggest that $H(\lambda,m)$ converges to zero for all $\lambda > 0$.

We finish this section with a remark concerning the distribution of the values of $s(n)$. Heuristically, the method of obtaining $s(n)$ by a matrix product is (formally) similar to studying the product of independent identically distributed random variables. The question therefore suggests itself: is the distribution of the values $s(n)$ in dyadic intervals $\left[2^m, 2^{m+1}\right)$ close to a log-normal distribution? We leave this as another open question.

1.4 Analytic and algebraic properties of Mahler functions

In this section, we consider the properties of regular functions and Mahler functions viewed as functions of a complex variable. In particular, we will address questions of convergence, analytic behaviour and rationality. In particular, the results will lead to a proof of Bézivin's theorem [12] that an irrational Mahler function is transcendental. The arguments in this section follow closely those of Bell, Coons and Rowland [11], who gave an alternative proof of Bézivin's result.

1.4.1 Analytic properties of Mahler functions

Allouche and Shallit's upper bound on regular sequences, Theorem 1.35, yields the following as an immediate corollary.

Proposition 1.54. *A regular function $F(z)$ converges inside the unit circle.*

This proposition can be used to give an alternative proof that there are Mahler functions that are not regular.

Example 1.55 (Example 1.27 revisited). Recall from Example 1.27, the function $1/(1-2z)$ is k-Mahler for each k. But $z = 1/2$ is a singularity of the function, so it does not converge everywhere inside the unit circle. Hence it is not k-regular for any k by Proposition 1.54.

Dumas' structure theorem, Theorem 1.32, yields the following immediate corollary, which we note here as a proposition.

Proposition 1.56. *Let $k \geqslant 2$ be an integer and let $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. Then $F(z)$ has a positive radius of convergence.*

Proof. Denote by $B(0,r)$ the open ball of radius $r > 0$ centred at the origin. Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function satisfying, say,

$$\sum_{j=0}^{d} a_j(z)F(z^{k^j}) = 0,$$

for $a_j(z) \in \mathbb{C}[z]$, $a_0(z)a_d(z) \neq 0$. Proposition 1.54 states that a k-regular series is analytic in the unit disk, so Theorem 1.32 gives that $F(z)$ converges in $B(0,r)$, where $r \in (0,1)$ is the minimal distance from 0 to a nonzero root of $a_0(z)(z-1)$. □

It is quite easy to see that all polynomials are regular functions, and so they are all Mahler functions as well. As it turns out, polynomials are precisely the set of entire Mahler functions—and so also the set of entire regular functions.

Theorem 1.57. *Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. If $F(z)$ is entire, then $F(z)$ is a polynomial.*

Proof. Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be an entire k-Mahler function satisfying

$$\sum_{j=0}^{d} a_j(z)F(z^{k^j}) = 0,$$

for $a_j(z) \in \mathbb{C}[z]$ with $a_0(z)a_d(z) \neq 0$. Write

$$F(z^{k^d}) = -\sum_{j=0}^{d-1} \frac{a_j(z)}{a_d(z)} F(z^{k^j}). \tag{1.17}$$

Pick $L > 1$ such that all of the zeros of $a_d(z)$ are in the open disk, $B(0,L)$, of radius L centred at the origin. Notice that since the $a_i(z)$ are polynomials, there is an $N > 1$ and a constant $C > 1$ such that for $|z| \geqslant L$, we have

$$\max_{0 \leqslant i \leqslant d-1} \left\{ \left| \frac{a_i(z)}{a_d(z)} \right| \right\} < C|z|^N; \tag{1.18}$$

in particular, the value $N = \max_{0 \leqslant i \leqslant d-1}\{\deg a_i(z), 2\}$ is sufficient.

For $\ell \geqslant 0$ denote

$$M_\ell := \max\left\{ |F(z)| : |z| = L^{k^\ell} \right\},$$

where L is as chosen above. Using (1.17), (1.18), and the Maximum Modulus Theorem, we have for $j \geqslant d$ that

$$M_j \leqslant (d+1)C\big(L^{k^{j-d}}\big)^N M_{j-1} \leqslant C(d+1)L^{Nk^j} M_{j-1}.$$

Thus recursively, we have for each $n \geqslant d$ that

$$M_n \leqslant M_{d-1}(C(d+1))^n L^{Nk^{n+1}}.$$

But since $L > 1$, this implies that there is some constant $b > 0$ such that for $n \geqslant d$ we have

$$M_n \leqslant L^{bk^n}.$$

Now let $m \geqslant b+2$ be a natural number, fix an $\alpha \in \mathbb{C}$ and consider

$$F^{(m-1)}(\alpha) = \frac{1}{2\pi i} \int_{\gamma_n} \frac{F(z)}{(z-\alpha)^m} dz,$$

where γ_n is the circle of radius L^{k^n} with n large enough so that α is inside the circle of radius $L^{k^n}/2$ centred at the origin. Then for all z on γ_n we have that

$$\frac{|z|}{2} \leqslant |z - \alpha|.$$

Thus for n large enough, we have

$$|F^{(m-1)}(\alpha)| \leqslant \frac{1}{2\pi} \cdot 2\pi L^{k^n} \cdot \frac{2^m M_n}{(L^{k^n})^m} = \frac{2^m M_n}{(L^{k^n})^{m-1}} \leqslant 2^m L^{k^n(b-m+1)}.$$

Recall that $m \geqslant b+2$ so that the above gives that

$$|F^{(m-1)}(\alpha)| \leqslant \frac{2^m}{L^{k^n}}.$$

Since n can be taken arbitrarily large, we have that $F^{(m-1)}(\alpha) = 0$. But $\alpha \in \mathbb{C}$ was arbitrary, and so $F^{(m-1)}(z)$ is identically zero; hence $F(z)$ is a polynomial. $\square$

1.4.2 Rational-transcendental dichotomy of Mahler functions

Using Theorem 1.57 one can prove a rational-transcendental dichotomy of Mahler functions; see Bézivin [12].

Theorem 1.58 (Bézivin). *Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. If $F(z)$ is algebraic, then $F(z)$ is a rational function.*

In fact, since algebraic functions have only a finite number of singularities (see Flajolet and Sedgewick [29, Section VII.7.1]), Theorem 1.58 is a consequence of the upcoming Theorem 1.60. First we record a lemma, the proof of which is left as an exercise, though it can be found in the paper of Bell, Coons and Rowland [11].

Lemma 1.59. *Let $k \geqslant 2$ be an integer and let $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. The function $F(z)$ is meromorphic if and only if it has finitely many singularities.*

Theorem 1.60. *Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. If $F(z)$ has only finitely many singularities, then $F(z)$ is a rational function.*

Proof. Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function satisfying

$$\sum_{j=0}^{d} a_j(z) F(z^{k^j}) = 0, \tag{1.19}$$

for $a_j(z) \in \mathbb{C}[z]$ with $a_0(z)a_d(z) \neq 0$. If $F(z)$ has only finitely many singularities, then since by Lemma 1.59 it is meromorphic, there is a non-zero polynomial $q(z) \in \mathbb{C}[z]$ such that $q(z)F(z)$ is entire. For $j \in \{0,\dots,d-1\}$ set

$$q_j(z) := \frac{1}{q(z^{k^j})} \prod_{i=0}^{d} q(z^{k^i}) \in \mathbb{C}[z].$$

Multiplying (1.19) by $\prod_{i=0}^{d} q(z^{k^i}) \in \mathbb{C}[z]$ we then have that

$$\sum_{j=0}^{d} a_j(z) q_j(z) q(z^{k^j}) F(z^{k^j}) = 0,$$

where since $q(z)$ is not identically zero we have that $a_0(z)q_0(z)a_d(z)q_d(z) \neq 0$. Hence $q(z)F(z)$ is an entire k-Mahler function and thus, by the preceding lemma, a polynomial. This proves that $F(z)$ is a rational function. $\square$

One can actually do a lot better as Randé showed in his thesis [48].

Theorem 1.61 (Randé (1992)). *Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. Then $F(z)$ is a rational function or it has the unit circle as a natural boundary.*

Recall that a function is *differentiably finite* (or *D-finite*) provided is satisfies a linear homogeneous differential equation with polynomial coefficient. Since *D*-finite functions can have only a finite number of singularities (see Flajolet and Sedgewick [29, Section VII.9.1]), Randé's result implies the following corollary.

Corollary 1.62. *Let $k \geqslant 2$ be an integer and $F(z) \in \mathbb{C}[[z]]$ be a k-Mahler function. If $F(z)$ is D-finite, then $F(z)$ is a rational function.*

It is a open and very interesting question to determine where Mahler functions fall in the diffeo-algebraic hierarchy. Of particular interest is whether an irrational Mahler function can satisfy an algebraic differential equation. A function that does not satisfy an algebraic differential equation is called *hypertranscendental*.

Question 1.63. Is it true that an irrational Mahler function is hypertranscendental?

For Mahler functions of degree one, this question has been mostly answered by Bundschuh [18], though any sort of general result for other degrees remains open.

1.5 Rational-transcendental dichotomy of regular numbers

While the rational-transcendental dichotomy of regular (and Mahler) functions is more or less straightforward as shown in the previous section, the dichotomy at the level of their special values was much more elusive.

Adamczewski and Bugeaud [2] showed that a real automatic irrational number is transcendental and Bell, Bugeaud and Coons [8] generalised their result to show that if $F(z)$ is a regular function, then the value $F(1/b)$, for any integer $b \geqslant 2$, is either rational or transcendental. In this section, we provide a simplified version of the result of Bell, Bugeaud and Coons.

Theorem 1.64 (Bell, Bugeaud, and Coons). *Let $F(z) \in \mathbb{Z}[[z]]$ be a k-regular power series and $b \geqslant 2$ be a positive integer. Then either $F(1/b)$ is rational or it is transcendental.*

We take as our starting point Equation (1.9). To this end, let $F(z)$ be a k-regular function and let $\mathbf{F}(z) := [F(z) = F_1(z), \ldots, F_d(z)]^T$ be the vector of functions that form a basis for the $\mathbb{Q}(z)$-vector space V in the proof of Theorem 1.23, and recall that (1.9) gives

$$\mathbf{F}(z) = \mathbf{A}(z)\mathbf{F}(z^k), \tag{1.20}$$

where $\mathbf{A}(z) = (a_{i,j}(z)/B)_{1\leqslant i,j\leqslant d} \in \mathbb{Q}[z]^{d\times d}$ is a nonsingular matrix of polynomials $a_{i,j}(z) \in \mathbb{Z}(z)$ of degree at most $k-1$ and B is a nonzero positive integer.

We will require some additional notation. In particular, in this section we take all complex matrix norms $\|\cdot\|$ to be the operator norm; i.e., $\|A\| = \sup_{\|\mathbf{v}\|=1}\|A\mathbf{v}\|$, where the norm of a vector $\mathbf{v}$ is the ordinary Euclidean norm. Also, we let $\nu : \mathbb{Q}((x)) \to \mathbb{Z}\cup\{\infty\}$ be the valuation defined by $\nu(0) = \infty$ and

$$\nu\left(\sum_{n\geqslant -m} c_n x^n\right) := \inf\{i : c_i \neq 0\}$$

when $\sum_{n\geqslant -m} c_n x^n \in \mathbb{Q}((x))$ is a nonzero Laurent power series (this valuation will also be used in further sections).

Lemma 1.65. *Let $\mathbf{F}(z)$ satisfy (1.20) and $H := \max_{1\leqslant i,j\leqslant d}\{\deg a_{i,j}(z)\}$. Then there are $\varepsilon > 0$, polynomials $P_1(z),\ldots,P_d(z),Q(z) \in \mathbb{Z}[z]$ of degree at most $(d-1)(d+2)H$ with $Q(0) = 1$, and a positive constant $C = C(\varepsilon)$ such that for $i \in \{1,\ldots,d\}$ we have*

$$|F_i(t) - P_i(t)/Q(t)| \leqslant Ct^{d(d+2)H}$$

for $t \in (0,\varepsilon)$.

Proof. For $i \in \{1,2,\ldots,d\}$, the theory of simultaneous Padé approximation (see the monograph *Rational Approximations and Orthogonal Polynomials* by Nikishin and Sorokin [46, Chapter 4] for details) provides polynomials $P_i(z)$ and $Q(z)$ of degree each bounded by $(d-1)(d+2)H$, and $Q(0) = 1$, such that

$$\nu\left(Q(z)F_i(z) - P_i(z)\right) \geqslant d(d+2)H.$$

For $i \in \{1,\ldots,d\}$, we thus have

$$\nu\left(F_i(z) - \frac{P_i(z)}{Q(z)}\right) \geqslant d(d+2)H.$$

Since $Q(0) = 1$ and by Proposition 1.54 each of $F_1(z),\ldots,F_d(z)$ converges inside the unit disc, $F_i(z) - P_i(z)/Q(z)$ is analytic inside $B(0,r)$ for $i \in \{1,\ldots,d\}$ for some $r > 0$ since $Q(0) = 1$. Hence there exist power series $G_1(z),\ldots,G_d(z)$ that are analytic inside $B(0,r)$ such that

$$F_i(z) - \frac{P_i(z)}{Q(z)} = z^{d(d+2)H}G_i(z)$$

for $i \in \{1,\ldots,d\}$. Let $\varepsilon \in (0,r)$. Then there is a positive constant C such that

$$|G_1(z)|,\ldots,|G_d(z)| \leqslant C$$

for $|z| \leqslant \varepsilon$. Thus for $i \in \{1,\ldots,d\}$,

$$\left|F_i(t) - \frac{P_i(t)}{Q(t)}\right| \leqslant Ct^{d(d+2)H}$$

whenever $t \in (0, \varepsilon)$. $\square$

Having established the first rational approximations to our vector of regular functions, we now establish a family of good rational approximations, which will be used in the proof of Theorem 1.64.

Lemma 1.66. *Let* $\mathbf{F}(z)$ *satisfy* (1.20) *and* $H := \max_{1 \leqslant i,j \leqslant d}\{\deg a_{i,j}(z)\}$ *and let* $t \in (0,1)$. *Then for each* $n \geqslant 0$ *there are polynomials* $P_{1,n}(z), \ldots, P_{d,n}(z), Q_n(z) \in \mathbb{Z}[z]$ *satisfying:*

(i) $\max_{1 \leqslant i \leqslant d}\{\deg P_{i,n}(z), \deg Q_n(z)\} \leqslant ((d+2)(d-1)+1)Hk^n$;
(ii) $Q_n(z) = B^n Q_0(z^{k^n})$;
(iii) there exists an $\varepsilon > 0$ *and positive constants* $C_0 = C_0(\varepsilon)$ *and* $C_1 = C_1(\varepsilon)$, *not depending on* t, *such that for* $i \in \{1, \ldots, d\}$ *and for all* n *sufficiently large we have* $Q_n(t) \neq 0$ *and*

$$|F_i(t) - P_{i,n}(t)/Q_n(t)| \leqslant C_1 C_0^n t^{d(d+2)Hk^n},$$

whenever $t \in (0, \varepsilon)$ *and in particular the order of vanishing of* $F_i(t) - P_{i,n}(t)/Q(t)$ *at* $t = 0$ *is at least* $d(d+2)Hk^n$.

Proof. By Lemma 1.65, there are $\varepsilon > 0$, polynomials $P_{1,0}(z), \ldots, P_{d,0}(z), Q_0(z) \in \mathbb{Z}[z]$ of degree at most $(d+2)(d-1)H$ with $Q_0(0) = 1$, and a positive constant C such that for $i \in \{1, \ldots, d\}$ we have

$$|F_i(t) - P_i(t)/Q_0(t)| \leqslant Ct^{d(d+2)H}$$

whenever $t \in (0, \varepsilon)$.

We define

$$\mathbf{R}_0(z) := [P_{1,0}(z)/Q_0(z), \ldots, P_{d,0}(z)/Q_0(z)]^T \tag{1.21}$$

and for $n \geqslant 1$, we take

$$\mathbf{R}_n(z) = \mathbf{A}(z)\mathbf{R}_{n-1}(z^k). \tag{1.22}$$

We note that there exist integer polynomials $P_{i,n}(z)$ for $i \in \{1, \ldots, d\}$ and $Q_n(z)$ such that

(a) $\mathbf{R}_n(z) = [P_{1,n}(z)/Q_n(z), \ldots, P_{d,n}(z)/Q_n(z)]^T$;
(b) $Q_n(z) = B \cdot Q_{n-1}(z^k)$ for $n \geqslant 1$.

From (b), we immediately get $Q_n(z) = B^n Q_0(z^{k^n})$. Since the entries of $\mathbf{A}(z)$ are all polynomials of degree at most H, we see that if we define

$$d_n := \max_{1 \leqslant i \leqslant d}\{\deg P_{i,n}(z), \deg Q_n(z)\},$$

then (1.22) gives $d_n \leqslant kd_{n-1} + H$. By induction we see, using the fact that $d_0 \leqslant (d-1)(d+2)H$, that

$$d_n \leqslant d(d+2)H \cdot k^n + H(k^{n-1} + \cdots + k + 1)$$
$$= k^n d_0 + H \cdot \frac{k^n - 1}{k-1} \leqslant ((d-1)(d+2)+1)Hk^n. \quad (1.23)$$

By assumption,

$$\mathbf{F}(x) = \mathbf{A}(x)\mathbf{F}(x^k),$$

and hence for $n \geqslant 1$ we have

$$\mathbf{F}(x) - \mathbf{R}_n(x) = \mathbf{A}(z)\mathbf{A}(z^k)\cdots\mathbf{A}(z^{k^{n-1}})\left(\mathbf{F}(z^{k^n}) - \mathbf{R}_0(z^{k^n})\right).$$

Then for n sufficiently large we have $t^{k^n} < \varepsilon$. Hence if e_i denotes the $d \times 1$ column vector whose i-th coordinate is 1 and all other coordinates are zero, then

$$\begin{aligned}|F_i(t) - P_{i,n}(t)/Q_n(t)| &= \left\| e_i^T \left(\mathbf{F}(t) - \mathbf{R}_n(t)\right)\right\| \\ &= \left\| e_i^T \mathbf{A}(t)\mathbf{A}(t^k)\cdots\mathbf{A}(t^{k^{n-1}})(\mathbf{F}(t^{k^n}) - \mathbf{R}_0(t^{k^n}))\right\| \\ &\leqslant \left\| \left(\mathbf{F}(t^{k^n}) - \mathbf{R}_0(t^{k^n})\right)\right\| \cdot \prod_{\ell=0}^{n-1} \left\|\mathbf{A}(t^{k^\ell})\right\| \\ &\leqslant C\sqrt{d}t^{d(d+2)Hk^n} \cdot \prod_{\ell=0}^{n-1} \left\|\mathbf{A}(t^{k^\ell})\right\|.\end{aligned}$$

Since each of the entries in $\mathbf{A}(z)$ is a polynomial with rational coefficients, there is a positive constant C_0 (independent of t) such that

$$\prod_{\ell=0}^{n-1} \left\|\mathbf{A}(t^{k^\ell})\right\| < C_0^n$$

for all $n \geqslant 1$ and any $t \in (0,1)$. Hence we have

$$|F_i(t) - P_{i,n}(t)/Q_n(t)| < C\sqrt{d}C_0^n t^{d(d+2)Hk^n}$$

for all $i \in \{1, \ldots, d\}$ and all n sufficiently large. To see that this gives the statement about the order of vanishing at $t = 0$, note that if $F_i(t) - P_{i,n}(t)/Q_n(t)$ has a zero of order ℓ at $t = 0$ then we can write $F_i(t) - P_{i,n}(t)/Q_n(t)$ as $t^\ell G(t)$ where $G(0) \neq 0$. It follows that there is a neighbourhood of zero such that $|t^\ell G(t)| > |G(0)||t|^\ell/2$ for t in this neighbourhood. Letting t approach 0 from the right and using the fact that

$$|F_i(t) - P_{i,n}(t)/Q_n(t)| < C\sqrt{d}C_0^n t^{d(d+2)Hk^n}$$

gives $\ell \geqslant d(d+2)Hk^n$ and so $F_i(t) - P_{i,n}(t)/Q_n(t)$ has a zero at $t = 0$ of order at least $d(d+2)Hk^n$. ◻

With these preliminaries in hand, we are ready to proceed with the proof of Theorem 1.64. We will use the following version of the p-adic Schmidt subspace theorem due to Schlickewei [52].

Theorem 1.67 (p-adic Schmidt subspace theorem). *Let $n \geqslant 2$, $\varepsilon > 0$, and let $p_1, \dots, p_s$ be distinct prime numbers. Further, let $L_{1,\infty}, \dots, L_{n,\infty}$ be linearly independent linear forms in $X_1, \dots, X_n$ with algebraic coefficients in $\mathbb{C}$, and for $j = 1, \dots, s$, let $L_{1,p_j}, \dots, L_{n,p_j}$ be linearly independent forms in $X_1, \dots, X_n$ with algebraic coefficients in $\overline{\mathbb{Q}}_{p_j}$. Consider the inequality*

$$|L_{1,\infty}(\mathbf{x}), \dots, L_{n,\infty}(\mathbf{x})| \cdot \prod_{j=1}^{s} |L_{1,p_j}(\mathbf{x}), \dots, L_{n,p_j}(\mathbf{x})|_{p_j} \leqslant \|\mathbf{x}\|^{-\varepsilon}, \tag{1.24}$$

with $\mathbf{x} \in \mathbb{Z}^n$. There are a finite number of proper linear subspaces $T_1, \dots, T_t$ of $\mathbb{Q}^n$ such that all solutions of (1.24) *lie in $T_1 \cup \cdots \cup T_t$.*

Proof (of Theorem 1.64). Let $\mathbf{F}(z)$ satisfy (1.20). By Lemma 1.66, there exist polynomials $P_{1,n}(x), \dots, P_{d,n}(x), Q_n(x) \in \mathbb{Z}[x]$ such that

$$Q_n(x) = B^n Q_0(x^{k^n}), \tag{1.25}$$

and constants $C_1, C_2 > 0$ such that for $i \in \{1, \dots, d\}$ and for sufficiently large n, we have $Q_n(1/b) \neq 0$ and

$$\left| F(1/b) - \frac{P_{d,n}(1/b)}{Q_n(1/b)} \right| \leqslant \frac{C_1 C_0^n}{b^{d(d+2)Hk^n}}.$$

Let D be the smallest positive integer such that

$$p_n := b^{Dk^n} P_{d,n}(1/b) \quad \text{and} \quad q_n := b^{Dk^n} Q_n(1/b)$$

are both integers, and so we have

$$|q_n \cdot F(1/b) - p_n| \leqslant \frac{C_1 C_0^n |q_n|}{b^{d(d+2)Hk^n}}. \tag{1.26}$$

Recall, by Lemma 1.66, we have

$$\deg P_{d,n}(x) \leqslant \deg Q_n(x) \leqslant d(d+1)H$$

so that also $D < d(d+1)H$. Also by (1.25), we have that $\deg Q_0(x^{k^n}) = D$. Write

$$Q_0(x^{k^n}) := \sum_{i=0}^{D} a_i x^{ik^n},$$

and assume, without loss of generality, $a_i \neq 0$ for each i (the general case follows *mutatis mutandis*). Note that by (1.25) we have that

$$|q_n| = B^n \left| \sum_{i=0}^{D} a_i b^{(D-i)k^n} \right| \leqslant B^n \sum_{i=0}^{D} |a_i| b^{(D-i)k^n} \leqslant C_2 B^n b^{Dk^n},$$

where $C_2 \geqslant \sum_{i=0}^{D} |a_i| > 0$ is a positive constant. Thus for n large enough, since $d \geqslant 2$ we have

$$|q_n F(1/b) - p_n| \leqslant \frac{C_1 C_2 (C_1 B)^n b^{d(d+1)Hk^n}}{b^{d(d+2)Hk^n}} = \frac{C_1 C_2 (C_1 B)^n}{b^{dHk^n}} < \frac{1}{b^{Hk^n}}. \tag{1.27}$$

We now setup to apply the p-adic Schmidt subspace theorem, suppose that $\xi := F(1/b)$ is algebraic and for $\mathbf{x} = (x_1, \dots, x_{D+2}) \in \mathbb{Z}^{D+2}$ set

$$L_{i,\infty}(\mathbf{x}) := x_i \qquad (i \in \{1, \dots, D+1\}),$$

and

$$L_{D+2,\infty}(\mathbf{x}) := \xi \sum_{i=1}^{D+1} x_i + x_{D+2}.$$

Also for each prime p dividing b set

$$L_{i,p}(\mathbf{x}) := x_i \qquad (i \in \{1, \dots, D+2\}).$$

For $n \in \mathbb{N}$ denote

$$\mathbf{s}_n := (B^n a_0 b^{Dk^n}, B^n a_1 b^{(D-1)k^n}, \dots, B^n a_D, -p_n) \in \mathbb{Z}^{D+2}.$$

Then (1.27) gives for large enough n that

$$|L_{D+2,\infty}(\mathbf{s}_n)| < \frac{1}{b^{Hk^n}}.$$

Also, we have that

$$|L_{1,\infty}(\mathbf{s}_n) \cdots L_{D+1,\infty}(\mathbf{s}_n)| = \prod_{i=0}^{D} B^n |a_i| b^{ik^n} \leqslant C_3 B^{Dn} b^{\frac{D(D+1)}{2} k^n},$$

where $C_3 := \prod_{i=1}^{D+1} |a_i| + 1 \geqslant 0$ is a positive constant.

For primes p dividing b, we have

$$\begin{aligned}\prod_{i=1}^{D+2} \prod_{p|b} |L_{i,p}(\mathbf{s}_n)|_p &\leqslant \prod_{i=0}^{D} \prod_{p|b} |B^n a_i b^{ik^n}|_p \\ &\leqslant \prod_{i=0}^{D} \prod_{p|b} |b^{ik^n}|_p = \prod_{i=0}^{D} \prod_{p|b} p^{-\nu_p(b) \cdot ik^n} = b^{-\frac{D(D+1)}{2} k^n},\end{aligned}$$

where for $\prod_{p|b} |L_{D+2,p}(\mathbf{s}_n)|_p$ we used the trivial bound of 1.

To bound $\|\mathbf{s}_n\|$, we note first that since $|L_{D+2,\infty}(\mathbf{s}_n)| < b^{-Hk^n}$, we have

$$|p_n| \leqslant |\xi| B^n \left| \sum_{i=0}^{D} a_i b^{(D-i)k^n} \right| + b^{-Hk^n}.$$

Thus

$$\begin{aligned}\|\mathbf{s}_n\|^{D+2} &= \sum_{i=0}^{D}\left|B^n a_i b^{(D-i)k^n}\right|^{D+2} + |p_n|^{D+2} \\ &< \sum_{i=0}^{D}\left|B^n a_i b^{(D-i)k^n}\right|^{D+2} + \left(|\xi|B^n\left|\sum_{i=0}^{D} a_i b^{(D-i)k^n}\right| + b^{-Hk^n}\right)^{D+2} \\ &\leqslant \left(\sum_{i=0}^{D}\left|B^n a_i b^{(D-i)k^n}\right| + |\xi|B^n\left|\sum_{i=0}^{D} a_i b^{(D-i)k^n}\right| + b^{-Hk^n}\right)^{D+2},\end{aligned}$$

and so there is constant $C_4 > 0$ such that $\|\mathbf{s}_n\| \leqslant C_4 B^n b^{Dk^n}$. Thus for a given $\varepsilon > 0$, we have that

$$\frac{1}{C_4^{\varepsilon} B^{\varepsilon n} b^{\varepsilon D k^n}} \leqslant \|\mathbf{s}_n\|^{-\varepsilon}.$$

Now set $\varepsilon = \frac{1}{2D}$. Then putting these bounds together gives for n large enough that

$$\begin{aligned}&|L_{1,\infty}(\mathbf{s}_n),\dots,L_{D+2,\infty}(\mathbf{s}_n)| \cdot \prod_{p|b} |L_{1,p}(\mathbf{s}_n),\dots,L_{D+2,p}(\mathbf{s}_n)|_p \\ &\quad< \frac{C_3 B^{Dn}}{b^{Hk^n}} = \frac{C_3 B^{Dn}}{b^{Hk^n}} \cdot \frac{C_4^{\varepsilon} B^{\varepsilon n} b^{\varepsilon D k^n}}{C_4^{\varepsilon} B^{\varepsilon n} b^{\varepsilon D k^n}} \leqslant \frac{C_3 B^{Dn} C_4^{\varepsilon} B^{\varepsilon n}}{b^{(H-\varepsilon D)k^n}} \cdot \|\mathbf{s}_n\|^{-\varepsilon} \leqslant \|\mathbf{s}_n\|^{-\varepsilon},\end{aligned}$$

for n large enough, since $H \geqslant 1$ as long as $F(x)$ is not identically 1 (in which case $F(1/b)$ is rational and the theorem holds anyway).

Thus for n large enough, the $(D+2)$-tuples $\mathbf{s}_n$ are solutions to the system,

$$|L_{1,\infty}(\mathbf{s}_n),\dots,L_{D+2,\infty}(\mathbf{s}_n)| \cdot \prod_{p|b} |L_{1,p}(\mathbf{s}_n),\dots,L_{D+2,p}(\mathbf{s}_n)|_p \leqslant \|\mathbf{s}_n\|^{-\frac{1}{2D}},$$

which by the p-adic Schmidt subspace theorem, lie in finitely many proper linear subspaces of $\mathbb{Q}^{D+2}$. Hence there exists a nonzero $(D+2)$-tuple $(\alpha_0,\dots,\alpha_{D+1}) \in \mathbb{Q}^{D+2}$, such that for n large enough

$$\alpha_0 B^n a_0 b^{Dk^n} + \sum_{i=1}^{D} \alpha_i B^n a_i b^{(D-i)k^n} - \alpha_{D+1} p_n = 0.$$

Dividing by q_n and taking the limit as $n \to \infty$ we have

$$\alpha_0 - \alpha_{D+1}\xi = 0,$$

so that $\xi = F(1/b) \in \mathbb{Q}$, which completes the proof of the theorem. □

Remark 1.68. In very recent work, Adamczewski and Faverjon [3] have extended the results of Adamczewski and Bugeaud [2] and Bell, Bugeaud and Coons [8] to the best possible. They have shown that an Mahler function evaluated at an algebraic number is either rational or transcendental. Moreover, their proof avoided the use of Schmidt's subspace theorem!

1.6 Diophantine properties of Mahler functions

In our final section, we look at the Diophantine properties of Mahler functions. We first look at how well a Mahler function can be approximated by rational functions. We then use that information to present the Universal transcendence test for Mahler functions due to Bell and Coons [9]. Finally, we focus on the approximation of Mahler functions with algebraic functions.

1.6.1 Rational approximation of Mahler functions

Suppose we have a rational solution to (1.3). Our first result of this section gives bounds on the degrees of the numerator and the denominator of a rational Mahler function. This result can be found in Bell and Coons [9, Proposition 2].

Proposition 1.69. *Let $F(z) = P(z)/Q(z)$ be a rational k-Mahler function satisfying* (1.3) *with* $\gcd(P(z), Q(z)) = 1$ *and set* $H := \max\{\deg a_i(z) : i = 0, \ldots, d\}$. *Then*

$$\deg Q(z) \leqslant \lfloor H(k-1)/(k^{d+1} - 2k^d + 1)\rfloor,$$

and

$$\deg P(z) \leqslant \deg Q(z) + \lfloor H/k^{d-1}(k-1)\rfloor.$$

Proof. Write $F(z) = P(z)/Q(z)$ with $\gcd(P(z), Q(z)) = 1$. Since $F(z)$ is a power series, $Q(0) \neq 0$. Then we have

$$\sum_{i=0}^{d} a_i(z) P(z^{k^i})/Q(z^{k^i}) = 0.$$

In particular, if we multiply both sides by

$$R(z) := \prod_{j=0}^{d-1} Q(z^{k^j}),$$

we see that $Q(z^{k^d})$ must divide $a_d(z)P(z^{k^d})R(z)$. Since $\gcd(P(z), Q(z)) = 1$, we then have that $Q(z^{k^d})$ divides $a_d(z)R(z)$. Let D denote the degree of $Q(z)$. Then considering degrees, we have

$$k^d D \leqslant \deg a_d(z) + \deg R(z) \leqslant H + D + kD + \cdots + k^{d-1}D.$$

In other words, $(k^d - k^{d-1} - \cdots - 1)D \leqslant H$. Since

$$k^d - k^{d-1} - \cdots - 1 = k^d - (k^d - 1)/(k-1) \geqslant k^d(k-2)/(k-1),$$

if $k > 2$, we have

$$D \leqslant H(k-1)/k^d(k-2).$$

If $k = 2$, then all we get is $D \leqslant H$. In any case, setting

$$A(H,k,d) := \lfloor H(k-1)/(k^{d+1} - 2k^d + 1) \rfloor,$$

we have $D = \deg Q(z) \leqslant A(H,k,d)$.

Similarly, we can bound the degree of $P(z)$, but this is slightly more subtle. Suppose that $F(z) = P(z)/Q(z)$ has a pole at $z = \infty$ of order $M > 0$ with $Mk^{d-1} + H < Mk^d$. Since $F(z)$ satisfies (1.3), we have

$$F(z^{k^d})a_d(z) = -\sum_{i=0}^{d-1} a_i(z)F(z^{k^i}). \tag{1.28}$$

Now, the right-hand side of (1.28) has a pole at $z = \infty$ of order at most $k^{d-1}M + H$ and the left-hand side of (1.28) has a pole at $z = \infty$ of order at least $k^d M$. Since the equality (1.28) must hold, we conclude that $Mk^{d-1} + H \geqslant Mk^d$ and so $M \leqslant H/(k^d - k^{d-1})$. In other words,

$$\deg P(z) \leqslant \deg Q(z) + H/k^{d-1}(k-1),$$

which finishes the proof. □

1.6.2 A transcendence test for Mahler functions

While we can bound the degrees of the numerator and the denominator of a rational Mahler function, unfortunately, deciding whether a general power series is a rational function is still not effectively determinable. After all, one can imagine that the function is very close to some rational function and one must go very far out when looking at its coefficients to see that it is irrational. Fortunately, as Bell and Coons showed [9, Lemma 1], deciding whether a Mahler function is a rational function is effective.

Lemma 1.70. *Let $F(z)$ be a Mahler function satisfying* (1.3) *and as before set $H := \max\{\deg a_i(z) : i = 0, \ldots, d\}$. If $P(z)/Q(z)$ is a rational function with $Q(0) \neq 0$ and the degrees of $P(z)$ and $Q(z)$ are strictly less than some positive integer κ, then $F(z) - P(z)/Q(z)$ is either identically zero or it has a nonzero coefficient of z^i for some $i \leqslant H + \kappa \cdot k^{d+1}/(k-1)$.*

Proof. Suppose not. Then $F(z) - P(z)/Q(z) = z^M T(z)$ for some nonzero power series $T(z)$ with $T(0)$ nonzero and some $M > H + \kappa \cdot k^{d+1}/(k-1)$. Then we have

$$\sum_{i=0}^{d} a_i(z)P(z^{k^i})/Q(z^{k^i}) = \sum_{i=0}^{d} a_i(z)z^{Mk^i}T(z^{k^i}). \tag{1.29}$$

Notice the right-hand side of (1.29) has a zero of at least order M at $z = 0$. On the other hand, we can write the left-hand side of (1.29) as a rational function with denominator $Q(z)Q(z^k)\cdots Q(z^{k^d})$ and numerator

$$\sum_{i=0}^{d} a_i(z)P(z^{k^i})R_i(z),$$

where $R_i(z) := \prod_{j\neq i} Q(z^{k^j})$. Thus the numerator of the left-hand side of (1.29) when written in lowest terms has degree at most $H + \kappa(k^d + \cdots + k + 1)$. But this can occur only if the left-hand side of (1.29) is identically zero since $M > H + \kappa(k^{d+1} - 1)/(k-1)$, a contradiction. $\square$

Universal test for transcendence of Mahler functions.

Let $k \geqslant 2$ and $d \geqslant 1$ be integers and $F(z)$ be a k-Mahler function satisfying

$$a_0(z)F(z) + a_1(z)F(z^k) + \cdots + a_d(z)F(z^{k^d}) = 0,$$

for polynomials $a_0(z), \ldots, a_d(z) \in \mathbb{C}[z]$. Set $H := \max\{\deg a_i(z) : i = 0, \ldots, d\}$ and

$$\kappa := \lfloor H(k-1)/(k^{d+1} - 2k^d + 1)\rfloor + \lfloor H/k^{d-1}(k-1)\rfloor + 1.$$

Step 1. Compute the coefficient, $f(i)$, of z^i of $F(z)$ for

$$i = 0, 1, \ldots, \kappa + H + \kappa(k^{d+1} - 1)/(k-1).$$

Step 2. Form the

$$(1+\kappa) \times (1 + H + \kappa(k^{d+1} - 1)/(k-1))$$

matrix $\mathbf{M}$ whose (i, j)-entry is $f(i + j - 2)$.

Step 3. Put this matrix in echelon form and verify whether it has full rank (i.e., rank equal to $1+\kappa$).

Step 4. If it does, then $F(z)$ is transcendental; otherwise it is rational.

Fig. 1.4 Universal test for transcendence of Mahler functions of Bell and Coons.

Proof (of Universal test for transcendence of Mahler functions in Figure 1.4). Let $\mathbf{M}$ be the matrix formed in Step 2 of the universal transcendence test described in Figure 1.4.

Suppose that $\mathbf{M}$ does not have full rank. Then there is a nonzero row vector $\mathbf{q} := [q_0, q_1, \ldots, q_\kappa]$ such that $\mathbf{q} \cdot \mathbf{M} = 0$. In other words,

$$(q_\kappa + q_{\kappa-1}z + \cdots + q_0 z^\kappa)F(z)$$

has the property that 0 is the coefficient of z^i for $i = \kappa, \ldots, \kappa + H + \kappa(k^{d+1} - 1)/(k - 1)$; that is, there is a polynomial $P(z)$ of degree less than κ such that

$$(q_\kappa + q_{\kappa-1} z + \cdots + q_0 z^\kappa) F(z) - P(z)$$

has a zero of order at least $\kappa + H + \kappa(k^{d+1} - 1)/(k - 1)$ at $z = 0$. Then $P(z)$ must have an order of zero at $z = 0$ that is at least as great as the order of zero of $Q(z) := q_\kappa + q_{\kappa-1} z + \cdots + q_0 z^\kappa$ at $z = 0$. This means that $P(z)/Q(z)$ can be reduced to be written as a ratio of polynomials of degree less than κ with the denominator being nonzero at $z = 0$ and such that $F(z) - P(z)/Q(z)$ has a zero at $z = 0$ of order at least $H + \kappa(k^{d+1} - 1)/(k - 1)$. Lemma 1.70 gives then that $F(z) - P(z)/Q(z)$ is identically zero and hence $F(z)$ is rational.

Conversely, if $F(z)$ is rational, then we write $F(z) = P(z)/Q(z)$ with the degree of $P(z)$ and $Q(z)$ less than κ and use $Q(z)$ to provide a nonzero row vector $\mathbf{q}$ as above with $\mathbf{q} \cdot \mathbf{M} = 0$. □

1.6.3 Algebraic approximation of Mahler functions

The main result presented in this subsection is the recent result of Coons [23] concerning a zero order estimate for the difference of a Mahler function with an algebraic function.

As before, let $\nu : \mathbb{C}((z)) \to \mathbb{Z} \cup \{\infty\}$ be the valuation defined by $\nu(0) := \infty$ and

$$\nu\left(\sum c_n z^n\right) := \min\{i : c_i \neq 0\}$$

when $\sum_n c_n z^n$ is nonzero. Also, for $G(z)$ an algebraic function with minimal polynomial $P(z,y) \in \mathbb{C}[z,y]$, we call the value $\deg_y P(z,y)$ the *degree* of $G(z)$ and we call the value $\exp\left(\deg_z P(z,y)\right)$ the *height* of $G(z)$.

Theorem 1.71 (Coons). *If $F(z)$ is an irrational k-Mahler function of degree d_F and height A_F, and $G(z)$ is an algebraic function of degree at most n and height at most H_G, then*

$$\nu\big(F(z) - G(z)\big) \leqslant (d_F + 1) \cdot A_F \cdot n^{d_F + 1} + \frac{k^{d_F + 1} - 1}{k - 1} \cdot \log H_G \cdot n^{d_F}.$$

The order of Coons's bound is very similar to that of previous results on zero estimates of Mahler functions, though those focussed on upper bounds for $\nu(Q(z, F(z)))$ for polynomials $Q(z,y) \in \mathbb{C}[z,y]$ and used quite deep methods, relying on the elimination-theoretic method of Nesterenko [44, 45]; see Becker [6], Nishioka [47], and Töpfer [53]. The approach taken by Coons is quite elementary and easily lends itself to exposition.

The case of rational functions was given by Bell's and Coons's result of the previous section (see Proposition 1.69). It is translated to the language of Theorem 1.71 as the following.

Lemma 1.72 (Bell and Coons). *Let $F(z)$ be an irrational k-Mahler function of degree d_F and height A_F, and let $P(z)/Q(z)$ be any rational function with $Q(0) \neq 0$. Then*

$$\nu\left(F(z) - \frac{P(z)}{Q(z)}\right) \leqslant A_F + \frac{k^{d_F+1}-1}{k-1} \cdot \max\{\deg P(z), \deg Q(z)\}.$$

Theorem 1.71 is the generalisation of this result to approximation by algebraic functions of arbitrary degree. To prove this generalisation, we use a resultant argument.

Lemma 1.73. *Let $f(z)$ and $g(z)$ be two algebraic functions of degrees at least 2 satisfying polynomials of degrees Δ_f and Δ_g with coefficients of degree at most δ_f and δ_g, respectively. Then the algebraic function $f(z)+g(z)$ satisfies a polynomial of degree*

$$\Delta_{f+g} \leqslant \Delta_f \Delta_g$$

with coefficients of degree

$$\delta_{f+g} \leqslant \delta_f \Delta_g + \delta_g \Delta_f.$$

Proof. This result follows by using the Sylvester matrix to calculate a certain resultant. For R a ring and $P, Q \in R[y]$ with

$$P(y) = \sum_{i=0}^{\deg_y P} p_i y^i \quad \text{and} \quad Q(y) = \sum_{i=0}^{\deg_y Q} q_i y^i,$$

the resultant of P and Q with respect to the variable y is denoted by $\operatorname{res}_y(P,Q)$ and may be calculated as the determinant of the $(\deg_y Q + \deg_y P) \times (\deg_y Q + \deg_y P)$ Sylvester matrix; that is

$$\operatorname{res}_y(P,Q) := \det \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_{\deg_y P} & & & & \\ & p_0 & p_1 & p_2 & \cdots & p_{\deg_y P} & & & \\ & & \ddots & \ddots & \ddots & & \ddots & & \\ & & & p_0 & p_1 & & p_2 & \cdots & p_{\deg_y P} \\ q_0 & q_1 & q_2 & \cdots & q_{\deg_y Q} & & & & \\ & q_0 & q_1 & q_2 & \cdots & q_{\deg_y Q} & & & \\ & & \ddots & \ddots & \ddots & & \ddots & & \\ & & & q_0 & q_1 & & q_2 & \cdots & q_{\deg_y Q} \end{pmatrix},$$

where there are $\deg_y Q$ rows of the coefficients of P and $\deg_y P$ rows of the coefficients of Q. Now suppose $R = \mathbb{C}[z,x]$, so that the entries of the above Sylvester matrix are polynomials in the variables z and x, and set $D(x,z) := \operatorname{res}_y(P,Q)$. Since polynomial degrees are additive, using the Leibniz formula for the determinant, we have immediately that

$$\deg_z D(x,z) \leqslant \deg_y Q \deg_z P + \deg_y P \deg_z Q \tag{1.30}$$

and

$$\deg_x D(x,z) \leqslant \deg_y Q \deg_x P + \deg_y P \deg_x Q. \tag{1.31}$$

The lemma now follows immediately by combining (1.30) and (1.31) with the fact that given algebraic functions $f(z), g(z) \in \mathbb{C}[[z]]$ and polynomials $P_f(z,y)$, $P_g(z,y) \in \mathbb{C}[z,y]$ with $P_f(z,f) = P_g(z,g) = 0$, the algebraic function $f(z)+g(z)$ is a root of the polynomial $\mathrm{res}_y(P_f(z,y), P_g(z,x-y))$ viewed as a polynomial in x. □

Using Lemma 1.72 as the result for algebraic functions of degree 1, we now focus on algebraic functions of degree at least 2.

Lemma 1.74. *Let $a_0(z),\ldots,a_d(z)$ be polynomials of degree at most A. If $G(z) \in \mathbb{C}[[z]]$ is an algebraic function of degree $\Delta_G \geqslant 2$ satisfying a minimal polynomial with coefficients of degree at most δ_g, then the function*

$$M_G(z) := \sum_{i=0}^{d} a_i(z) G(z^{k^i})$$

is an algebraic function satisfying a polynomial of degree $\Delta_{M_G} \leqslant \Delta_G^{d+1}$ whose coefficients have degree

$$\delta_{M_G} \leqslant (d+1)A \cdot \Delta_G^{d+1} + \frac{k^{d+1}-1}{k-1} \cdot \delta_G \cdot \Delta_G^{d}.$$

Proof. Since $G(z)$ is an algebraic function, so is $\sum_{i=0}^{d} a_i(z) G(z^{k^i})$. One can easily gain information about the sum using the theory of resultants.

To get an upper bound on $\nu(M_G(z))$, we apply the idea of the previous paragraph by including the terms $G_i(z) := a_i(z) G(z^{k^i})$ one at a time. To do this, let

$$P_G(z,y) := g_{\Delta_G} y^{\Delta_G} + \cdots + g_1 y + g_0$$

be the minimal polynomial of $G(z)$. Here we have denoted the degree of $G(z)$ by Δ_G. Set $\delta_G := \deg_z P_G(z,y)$. Then

$$P_{G_i}(z,y) = a_i(z)^{\Delta_G} P_G(z^{k^i}, y/a_i(z))$$

is a polynomial with $P_{G_i}(z, G_i(z)) = 0$, where, of course, we only form this polynomial when $a_i(z) \neq 0$. Here, we have that $P_{G_i}(z,y)$ is still minimal with respect to the degree of y, but there is no guarantee that it is minimal with respect to the degree of z for this degree of y. However, we do have that the minimal polynomial of $G_i(z)$ divides $P_{G_i}(z,y)$ and the quotient is just a polynomial in z. In any case, the above gives that

$$\Delta_{G_i} := \deg_y P_{G_i}(z,y) = \deg_y P_G(z,y) = \Delta_G \tag{1.32}$$

and

$$\delta_{G_i} := \deg_z P_{G_i}(z,y) \leqslant A\Delta_G + k^i \delta_G. \tag{1.33}$$

The lemma now follows by combining (1.32) and (1.33) with Lemma 1.73. □

Lemma 1.75. *Let $G(z) \in \mathbb{C}[[z]]$ be an algebraic function of degree at least 2 satisfying the polynomial $P_G(z,y) = a_n(z)y^n + a_{n-1}(z)y^{n-1} + \cdots + a_1(z)y + a_0(z)$, with $a_0(z) \neq 0$. Then $\nu(G(z)) \leqslant \nu(a_0(z))$. In particular, $\nu(G(z)) \leqslant \deg_z P_G(z,y)$.*

Proof. Since $P_G(z,y)$ is a minimal polynomial, we have $a_0(z) \neq 0$. We thus have, identically,

$$\left(a_n(z)G(z)^{n-1} + a_{n-1}(z)G(z)^{n-2} + \cdots + a_1(z)\right) G(z) = -a_0(z).$$

The fact $G(z), a_n(z), \ldots, a_0(z) \in \mathbb{C}[[z]]$ then gives

$$\nu\left(a_n(z)G(z)^{n-1} + a_{n-1}(z)G(z)^{n-2} + \cdots + a_1(z)\right) + \nu(G(z)) = \nu(a_0(z)),$$

which proves the lemma, since each of the terms is a nonnegative integer. □

Proof (of Theorem 1.71). Let $F(z)$ be a k-Mahler function satisfying (1.3) of degree d_F and height A_F and let $G(z)$ be an algebraic function of degree at most n and height at most H_G. Since by Lemma 1.72, the theorem holds for $n = 1$, we may assume without loss of generality that $n \geqslant 2$.

Set $M := \nu(F(z) - G(z))$, and write

$$F(z) - G(z) = z^M T(z),$$

where $T(z) \in \mathbb{C}[[z]]$ with $T(0) \neq 0$. Then also

$$\sum_{i=0}^{d} a_i(z)F(z^{k^i}) - \sum_{i=0}^{d} a_i(z)G(z^{k^i}) = \sum_{i=0}^{d} a_i(z)z^{k^i M}T(z^{k^i}),$$

which since $F(z)$ satisfies (1.3) reduces to

$$M_G(z) := \sum_{i=0}^{d} a_i(z)G(z^{k^i}) = -\sum_{i=0}^{d} a_i(z)z^{k^i M}T(z^{k^i}).$$

This immediately implies that

$$\nu(F(z) - G(z)) = M \leqslant \nu\left(M_G(z)\right) \leqslant \delta_{M_G},$$

where the last inequality follows from Lemma 1.75. By definition, $\delta_G = \log H_G$, hence applying Lemma 1.74 proves the theorem. □

The most important term in the inequality of Theorem 1.71 is the rightmost term. One the most important questions in the algebraic approximation of Mahler functions concerns the degree of n in this term. The current best known upper bound is d_F, but a lower value may be true. In particular, one may expect a ‘Roth-type’ upper bound.

Question 1.76. If $F(z)$ is an irrational Mahler function and $G(z)$ is an algebraic function of degree at most n and height at most H_G where $\log H_G \geqslant n \geqslant 1$, then is there a constant $c > 0$ such that $\nu(F(z) - G(z)) \leqslant c \cdot \log H_G \cdot n$?

Index

References

1. Adamczewski, B., Bell, J.P.: A problem about Mahler functions. preprint available at http://arxiv.org/abs/1303.2019
2. Adamczewski, B., Bugeaud, Y.: On the complexity of algebraic numbers. I. Expansions in integer bases. Ann. of Math. (2) **165**(2), 547–565 (2007). DOI 10.4007/annals.2007.165.547. URL http://dx.doi.org/10.4007/annals.2007.165.547
3. Adamczewski, B., Faverjon, C.: Méthode de Mahler: relations linéaires, transcendance et application aux nombres automatiques. https://arxiv.org/pdf/1508.07158v2.pdf p. preprint
4. Allouche, J.P., Shallit, J.: The ring of *k*-regular sequences. Theoret. Comput. Sci. **98**(2), 163–197 (1992). DOI 10.1016/0304-3975(92)90001-V. URL http://dx.doi.org/10.1016/0304-3975(92)90001-V
5. Allouche, J.P., Shallit, J.: Automatic sequences. Cambridge University Press, Cambridge (2003). DOI 10.1017/CBO9780511546563. URL http://dx.doi.org/10.1017/CBO9780511546563. Theory, applications, generalizations
6. Becker, P.G.: Effective measures for algebraic independence of the values of Mahler type functions. Acta Arith. **58**(3), 239–250 (1991)
7. Becker, P.G.: *k*-regular power series and Mahler-type functional equations. J. Number Theory **49**(3), 269–286 (1994). DOI 10.1006/jnth.1994.1093. URL http://dx.doi.org/10.1006/jnth.1994.1093
8. Bell, J.P., Bugeaud, Y., Coons, M.: Diophantine approximation of Mahler numbers. Proc. Lond. Math. Soc. (3) **110**(5), 1157–1206 (2015). DOI 10.1112/plms/pdv016. URL http://dx.doi.org/10.1112/plms/pdv016
9. Bell, J.P., Coons, M.: Transcendence tests for Mahler functions. Proc. Amer. Math. Soc. **145**(3), 1061–1070 (2017). DOI 10.1090/proc/13297. URL http://dx.doi.org/10.1090/proc/13297
10. Bell, J.P., Coons, M., Hare, K.G.: The minimal growth of a *k*-regular sequence. Bull. Aust. Math. Soc. **90**(2), 195–203 (2014). DOI 10.1017/S0004972714000197. URL http://dx.doi.org/10.1017/S0004972714000197
11. Bell, J.P., Coons, M., Rowland, E.: The rational-transcendental dichotomy of Mahler functions. J. Integer Seq. **16**(2), Article 13.2.10, 11 (2013)
12. Bézivin, J.P.: Sur une classe d'équations fonctionnelles non linéaires. Funkcial. Ekvac. **37**(2), 263–271 (1994). URL http://www.math.kobe-u.ac.jp/~fe/xml/mr1299865.xml
13. Blondel, V.D., Nesterov, Y., Theys, J.: On the accuracy of the ellipsoid norm approximation of the joint spectral radius. Linear Algebra Appl. **394**, 91–107 (2005). DOI 10.1016/j.laa.2004.06.024. URL http://dx.doi.org/10.1016/j.laa.2004.06.024

14. Blondel, V.D., Theys, J., Vladimirov, A.A.: An elementary counterexample to the finiteness conjecture. SIAM J. Matrix Anal. Appl. **24**(4), 963–970 (electronic) (2003). DOI 10.1137/S0895479801397846. URL `http://dx.doi.org/10.1137/S0895479801397846`
15. Borel, E.: Les probabilités dénombrables et leurs applications arithmétiques. Palermo Rend. **27**, 247–271 (1909)
16. Bousch, T., Mairesse, J.: Asymptotic height optimization for topical IFS, Tetris heaps, and the finiteness conjecture. J. Amer. Math. Soc. **15**(1), 77–111 (electronic) (2002). DOI 10.1090/S0894-0347-01-00378-2. URL `http://dx.doi.org/10.1090/S0894-0347-01-00378-2`
17. Bugeaud, Y.: Expansions of algebraic numbers. In: Four faces of number theory, EMS Ser. Lect. Math., pp. 31–75. Eur. Math. Soc., Zürich (2015)
18. Bundschuh, P.: Algebraic independence of infinite products and their derivatives. In: Number theory and related fields, *Springer Proc. Math. Stat.*, vol. 43, pp. 143–156. Springer, New York (2013). DOI 10.1007/978-1-4614-6642-0_6. URL `http://dx.doi.org/10.1007/978-1-4614-6642-0_6`
19. Calkin, N.J., Wilf, H.S.: Binary partitions of integers and Stern-Brocot-like trees. unpublished pp. updated version August 5, 2009, 19 pages (1998)
20. Cobham, A.: On the base-dependence of sets of numbers recognizable by finite automata. Math. Systems Theory **3**, 186–192 (1969)
21. Cobham, A.: Uniform tag sequences. Math. Systems Theory **6**, 164–192 (1972)
22. Coons, M.: Regular Sequences and the Joint Spectral Radius. Internat. J. Found. Comput. Sci. p. to appear
23. Coons, M.: Zero order estimate for Mahler functions. New Zealand J. Math. **46**, 83–88 (2016)
24. Coons, M., Tyler, J.: The maximal order of Stern's diatomic sequence. Mosc. J. Comb. Number Theory **4**(3), 3–14 (2014)
25. Daubechies, I., Lagarias, J.C.: Sets of matrices all infinite products of which converge. Linear Algebra Appl. **161**, 227–263 (1992). DOI 10.1016/0024-3795(92)90012-Y. URL `http://dx.doi.org/10.1016/0024-3795(92)90012-Y`
26. Dumas, P.: Récurrences mahlériennes, suites automatiques, études asymptotiques. Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt (1993). Thèse, Université de Bordeaux I, Talence, 1993
27. Dumas, P.: Joint spectral radius, dilation equations, and asymptotic behavior of radix-rational sequences. Linear Algebra Appl. **438**(5), 2107–2126 (2013). DOI 10.1016/j.laa.2012.10.013. URL `http://dx.doi.org/10.1016/j.laa.2012.10.013`
28. Dumas, P.: Asymptotic expansions for linear homogeneous divide-and-conquer recurrences: algebraic and analytic approaches collated. Theoret. Comput. Sci. **548**, 25–53 (2014). DOI 10.1016/j.tcs.2014.06.036. URL `http://dx.doi.org/10.1016/j.tcs.2014.06.036`
29. Flajolet, P., Sedgewick, R.: Analytic combinatorics. Cambridge University Press, Cambridge (2009). DOI 10.1017/CBO9780511801655. URL `http://dx.doi.org/10.1017/CBO9780511801655`
30. Hare, K.G., Morris, I.D., Sidorov, N., Theys, J.: An explicit counterexample to the Lagarias-Wang finiteness conjecture. Adv. Math. **226**(6), 4667–4701 (2011). DOI 10.1016/j.aim.2010.12.012. URL `http://dx.doi.org/10.1016/j.aim.2010.12.012`
31. Hartmanis, J., Stearns, R.E.: On the computational complexity of algorithms. Trans. Amer. Math. Soc. **117**, 285–306 (1965)
32. Jungers, R.: The joint spectral radius, *Lecture Notes in Control and Information Sciences*, vol. 385. Springer-Verlag, Berlin (2009). DOI 10.1007/978-3-540-95980-9. URL `http://dx.doi.org/10.1007/978-3-540-95980-9`. Theory and applications
33. Jungers, R.M., Blondel, V.D.: On the finiteness property for rational matrices. Linear Algebra Appl. **428**(10), 2283–2295 (2008). DOI 10.1016/j.laa.2007.07.007. URL `http://dx.doi.org/10.1016/j.laa.2007.07.007`
34. Kozyakin, V.S.: A dynamical systems construction of a counterexample to the finiteness conjecture. In: Proceedings of the 44th IEEE Conference on Decision and Control, European Control Conference, pp. 2338–2343 (2005)

35. Lagarias, J.C., Wang, Y.: The finiteness conjecture for the generalized spectral radius of a set of matrices. Linear Algebra Appl. **214**, 17–42 (1995). DOI 10.1016/0024-3795(93)00052-2. URL http://dx.doi.org/10.1016/0024-3795(93)00052-2
36. Lansing, J.: Distribution of values of the binomial coefficients and the Stern sequence. J. Integer Seq. **16**(3), Article 13.3.7, 10 (2013)
37. Loxton, J.H.: A method of Mahler in transcendence theory and some of its applications. Bull. Austral. Math. Soc. **29**(1), 127–136 (1984). DOI 10.1017/S0004972700021341. URL http://dx.doi.org/10.1017/S0004972700021341
38. Loxton, J.H., van der Poorten, A.J.: Arithmetic properties of automata: regular sequences. J. Reine Angew. Math. **392**, 57–69 (1988)
39. Mahler, K.: Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen. Math. Ann. **101**(1), 342–366 (1929)
40. Mahler, K.: Arithmetische Eigenschaften einer Klasse transzendental-transzendenter Funktionen. Math. Z. **32**(1), 545–585 (1930)
41. Mahler, K.: Uber das Verschwinden von Potenzreihen mehrerer Veränderlichen in speziellen Punktfolgen. Math. Ann. **103**(1), 573–587 (1930)
42. Mahler, K.: Remarks on a paper by W. Schwarz. J. Number Theory **1**, 512–521 (1969)
43. McNaughton, R., Zalcstein, Y.: The Burnside problem for semigroups. J. Algebra **34**, 292–299 (1975)
44. Nesterenko, Y.V.: Estimate of the orders of the zeroes of functions of a certain class, and their application in the theory of transcendental numbers. Izv. Akad. Nauk SSSR Ser. Mat. **41**(2), 253–284, 477 (1977)
45. Nesterenko, Y.V.: Algebraic independence of algebraic powers of algebraic numbers. Mat. Sb. (N.S.) **123(165)**(4), 435–459 (1984)
46. Nikishin, E.M., Sorokin, V.N.: Rational approximations and orthogonality, *Translations of Mathematical Monographs*, vol. 92. American Mathematical Society, Providence, RI (1991). Translated from the Russian by Ralph P. Boas
47. Nishioka, K.: Algebraic independence measures of the values of Mahler functions. J. Reine Angew. Math. **420**, 203–214 (1991). DOI 10.1515/crll.1991.420.203. URL http://dx.doi.org/10.1515/crll.1991.420.203
48. Randé, B.: Équations fonctionnelles de Mahler et applications aux suites p-régulières. Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt (1992). Thèse, Université de Bordeaux I, Talence, 1992
49. Reznick, B.: Some binary partition functions. In: Analytic number theory (Allerton Park, IL, 1989), *Progr. Math.*, vol. 85, pp. 451–477. Birkhäuser Boston, Boston, MA (1990)
50. Rota, G.C., Strang, G.: A note on the joint spectral radius. Nederl. Akad. Wetensch. Proc. Ser. A 63 = Indag. Math. **22**, 379–381 (1960)
51. Roth, K.F.: Rational approximations to algebraic numbers. Mathematika **2**, 1–20; corrigendum, 168 (1955)
52. Schlickewei, H.P.: The $\mathfrak{p}$-adic Thue-Siegel-Roth-Schmidt theorem. Arch. Math. (Basel) **29**(3), 267–270 (1977). DOI 10.1007/BF01220404. URL http://dx.doi.org/10.1007/BF01220404
53. Töpfer, T.: Zero order estimates for functions satisfying generalized functional equations of Mahler type. Acta Arith. **85**(1), 1–12 (1998)