Combinatorics on words and generating Dirichlet series of automatic sequences

Jean-Paul Allouche[1][0000−0002−9060−0784], Jeffrey Shallit[2][0000−0003−1197−3820], and Manon Stipulanti[3][0000−0002−2805−2465]

[1] CNRS, IMJ-PRG, Sorbonne, Paris, France
jean-paul.allouche@imj-prg.fr
[2] School of Computer Science, University of Waterloo, Waterloo, Canada
shallit@uwaterloo.ca
[3] Department of Mathematics, ULiège, Liège, Belgium
m.stipulanti@uliege.be

Abstract. Generating series are crucial in enumerative combinatorics, analytic combinatorics, and combinatorics on words. Though it might seem at first view that generating Dirichlet series are less used in these fields than ordinary and exponential generating series, there are many notable papers where they play a fundamental role, as can be seen in particular in the work of Flajolet and several of his co-authors. In this paper, we study Dirichlet series of integers with missing digits or blocks of digits in some integer base b, i.e., where the summation ranges over the integers whose expansions form some language strictly included in the set of all words on the alphabet $\{0, 1, \ldots, b-1\}$ that do not begin with a 0. We show how to unify and extend results proved by Nathanson in 2021 and by Köhler and Spilker in 2009. En route, we encounter several sequences from Sloane's On-Line Encyclopedia of Integer Sequences, as well as some famous q-automatic sequences or q-regular sequences.

Keywords: Combinatorics on words · Generating Dirichlet series · Automatic sequences · Missing digits · Restricted words.

2020 Mathematics Subject Classification: 05A15, 11B85, 11M41, 40A05, 68R15

1 Introduction

Combinatorics (and in particular combinatorics on words) frequently uses generating functions. The point of generating functions is to "synthesize" or "concentrate" in a single function the properties of the sequence of its coefficients. There are three main types of generating functions associated with a sequence $(a_n)_{n\geq 0}$:

- ⋆ the ordinary generating function: $\displaystyle\sum_{n\geq 0} a_n x^n$;

⋆ the exponential generating function: $\displaystyle\sum_{n\geq 0} \frac{a_n}{n!} x^n$;

⋆ the Dirichlet generating function: $\displaystyle\sum_{n\geq 1} \frac{a_n}{n^s}$, where $s \in \mathbb{C}$.

Note that these series can be considered either as "formal" series or as complex (or real) functions only defined in some *ad hoc* domain. An excellent source for generating functions is Wilf's book [37] (also available online). At a first glance, it might seem that ordinary and exponential generating series are more commonly used in (enumerative) combinatorics, while Dirichlet series are more used in number theory. However, this impression is certainly biased due, in particular, to the most famous such series, namely the Riemann zeta function $\zeta : s \mapsto \sum_{n\geq 1} 1/n^s$. To get easily convinced that generating Dirichlet series are *also* used in combinatorics, it suffices to mention the strong link between ordinary generating functions and generating Dirichlet functions through the Mellin transform (see, for instance, the paper of Wintner [38], who cites the 1859 pioneering paper of Riemann[4]), or the many papers of the *virtuosi* Flajolet and co-authors, who make use of the Mellin transform. In particular one can consult the "bible" by Flajolet and Sedgewick [12], from which we do not resist quoting the following lines:

> *[...] It is possible to go much further as first shown by De Bruijn, Knuth, and Rice in a landmark paper [7], which also constitutes a historic application of Mellin transforms in analytic combinatorics. (We refer to this paper for historical context and references.)*

Playing around with harmonic series and Dirichlet series with "missing terms" in the summation range, we revisited the 1914 paper of Kempner [24], where he studied the series $\sum' 1/n$, where $\sum'$ means that the summation ranges over integers with no occurrence of the digit 9 in their base-10-expansion: see Theorem 1 presented later on. Thus the *abscissa of convergence* (whose definition is recalled in Section 1.2) of the Dirichlet series $\sum' 1/n^s$ is less than or equal to 1. This abscissa of convergence is worth determining precisely, not only for its own sake, but also because is related to the singularities of the series. Actually, it is natural to look at series with a more general flavor. Following this, we came across two papers: namely, a 2009 paper by Köhler and Spilker [26], and a 2021 paper by Nathanson [30], where the abscissa of convergence of "restricted" Dirichlet series is determined. Here, "restricted" means that integers having some digits or some blocks of digits in their b-ary expansion are excluded from the summation range. Within this topic, our contribution is twofold. First, it appears that a classical result about the abscissa of convergence of Dirichlet series used in [26] provides a (shorter?) proof of results in [30]. We give this proof in Section 2. Second, it turns out that results of [26] can be extended to generalize those of [30]. So, in Section 3, we prove results giving the abscissa of convergence of restricted Dirichlet series (see Theorems 4 and 5 below). Namely, the abscissa of

[4] See a transcription at https://www.claymath.org/sites/default/files/zeta.pdf.

convergence is equal to $\log_b \lambda$, where λ is the dominant root of a polynomial with integer coefficients, and b is the base of the considered numeration system.

Further, since the sequences of coefficients of the series studied in the papers [26,30] are binary b-*automatic* sequences (for more on automatic sequences, see, for instance, the books [3,13,15]), this work opens the way to determining the abscissa of convergence of generating Dirichlet series for more general automatic sequences (see Conclusion).

As a final comment to this introduction, we would also like to add that related work is contained in the papers [2,5,11,22,23,29], in the papers of Janjić cited at the beginning of Section 3, and in the paper of Noonan and Zeilberger [31].

1.1 Notation and definitions

We let $\mathbb{N}$ denote the set of non-negative integers $\{0, 1, 2, \dots\}$. A subset B of a set A is called *proper* if $B \neq A$. We let $|A|$ denote the *cardinality* of the finite set A. (Also note that we will use the notation $|x|$ to denote the absolute value of the real number x. There should be no confusion between the two uses of $|\ |$.)

Let $(a_n)_{n\geq 1}$ be a sequence of real numbers. Its *summatory function* is the sequence $(A(n))_{n\geq 1}$ of its *partial sums* defined by $A(n) = \sum_{i=1}^{n} a_i$ for each $n \geq 1$. (Following the usual convention for a summation over an empty set of indices, we set $A(0) = 0$.) We also warn the reader that most sequences in this text will be indexed starting at 1 (except b-ary expansions as mentioned below).

An *alphabet* is a finite set. A (finite) *word* on an alphabet A is a finite sequence of letters from A. The *length* of a word is the number of letters it is made of. The *empty word* is the only word with 0-length. For all $n \geq 0$, we let A^n denote the set of all length-n words over A. We let A^* denote the set of words on A (including the empty word). It is equipped with the concatenation of words, which makes it a monoid (i.e., called the *free monoid* on A).

Let $b \geq 2$ be an integer. If the expansion of an integer n in the base-b numeration system is $n = \sum_{0\leq j\leq k} w_j b^j$, with $w_j \in [0, b-1] = \{0, 1, \dots, b-1\}$ and $w_k \neq 0$, then the *base-b* or *b-ary representation* of n, written $\mathrm{rep}_b(n)$, is the word $\mathrm{rep}_b(n) = w_k w_{k-1} \dots w_0$ on the alphabet $[0, b-1]$ (we warn the reader to pay attention to the order of indices and to the fact that they end at 0). The base-b representation of 0 is the empty word.

For two sequences $(u_n)_{n\geq 0}, (v_n)_{n\geq 0}$ of integers, we recall the classical notation $u_n \sim v_n$, which means that there exists an integer n_0 such that $v_n \neq 0$ for $n \geq n_0$ and $\lim_{n\to+\infty} \frac{u_n}{v_n} = 1$.

1.2 Dirichlet series, abscissa of convergence, and restrictions

Given a sequence $(a_n)_{n\geq 1}$ of real numbers, its associated *Dirichlet (generating) series* is defined by

$$F_a(s) = \sum_{n\geq 1} \frac{a_n}{n^s}, \text{ where } s \in \mathbb{C}.$$

It is well known that either there exists a real number s_0 such that $F_a(s)$ converges for all s with $\mathfrak{Re}(s) > s_0$ and diverges for all s with $\mathfrak{Re}(s) < s_0$, or $F_a(s)$ converges everywhere, or else $F_a(s)$ converges nowhere. The real number s_0 is called the *abscissa of convergence* of the series $F_a(s)$. If $F_a(s)$ does not converge anywhere (resp., converges everywhere), we set $s_0 = +\infty$ (resp., $s_0 = -\infty$).

Example 1. The most famous example of Dirichlet series is when all coefficients $a_n = 1$: the series is the *Riemann zeta function*. Its abscissas of convergence and of absolute convergence are both equal to $+1$. Another well-studied example is when $a_n := \pm 1$: the abscissa of convergence of the Dirichlet series $\sum(\pm 1)^n/n^s$ is equal to 0, while its abscissa of absolute convergence is $+1$ (this is well known, see, e.g., [28, p. 10]).

In the following, we will consider Dirichlet series in which the summation range is restricted or limited to some integers sharing a similar property. Let $b \geq 2$ be an integer and let L be a language on the alphabet $[0, b-1]$, i.e., a set L of words on $[0, b-1]$. We define the *restricted* Dirichlet series $F_L(s)$ by

$$F_L(s) = \sum_{\substack{n \geq 1 \\ \mathrm{rep}_b(n) \in L}} \frac{1}{n^s},$$

i.e., the series $F_a(s)$ where, for all $n \geq 1$, the coefficient a_n is 1 or 0 depending on whether $\mathrm{rep}_b(n)$ belongs to L or not.

At the beginning of the 20th century, Kempner [24] considered the set of positive integers whose base-10 representation contains no 9.

Theorem 1 (Kempner, [24]). *Consider the language $L = [0, 8]^*$ on the alphabet $[0, 9]$. Then the restricted harmonic series $F_L(1)$ converges.*

Remark 1. This result has now become classical, but it might seem strange since the "whole" series $\sum_{n \geq 1} 1/n$ diverges. Actually it is not hard to get convinced that "many" integers are suppressed in the restricted series (think, for instance, of all the integers belonging to the interval $[9 \cdot 10^k, 10^{k+1} - 1]$ for a fixed k).

Of course Kempner's result implies that the abscissa of convergence of the restricted Dirichlet series $F_L(s)$ is less than or equal to 1. It is then natural to try to determine the (exact) abscissa of convergence of this series and similar ones.

2 A shorter proof of Nathanson's result using Köhler and Spilker's method

In their 2009 paper, Köhler and Spilker studied the abscissa of convergence of some restricted Dirichlet series. For this particular restriction, they considered base-b representations that avoid some specific set of digits.

Theorem 2 (Köhler and Spilker, [26, Satz 2]). *Let $b \geq 2$ be an integer. Consider a non-empty subset D of $[0, b-1]$ with $D \neq \{0\}$. Then the abscissa of convergence of the restricted Dirichlet series $F_{D^*}(s)$ is equal to $\frac{\log |D|}{\log b}$.*

In particular, when the set D in Theorem 2 contains all but one element, we obtain a refinement of Theorem 1.

Corollary 1. *Let $b \geq 2$ be an integer. Consider the set $D = [0, b-1] \setminus \{a\}$ where $a \in [1, b-1]$. Then the abscissa of convergence of the restricted Dirichlet series $F_{D^*}(s)$ is equal to $\log_b(b-1)$.*

In 2021, Nathanson independently considered a different kind of restrictions on Dirichlet series [30]. For an integer base $b \geq 2$, the digits in a given position in the base-b representations are constrained to some rules. Depending on the position, the rules may vary: this is a fundamental difference between Köhler and Spilker's work and Nathanson's.

Theorem 3 (Nathanson, [30, Theorem 1]). *Let $b \geq 2$ be an integer. For all $i \geq 0$, consider a proper subset D_i of $[0, b-1]$ and let $\mathcal{D}$ denote the sequence $(D_i)_{i \geq 0}$. Define the language*

$$L_{\mathrm{forb}(\mathcal{D})} = \{w_n w_{n-1} \cdots w_0 : \ w_n \neq 0 \text{ and } w_i \notin D_i, \ \forall\, i \in [0, n]\}.$$

Suppose furthermore that the set $\mathcal{M} = \{m \geq 1 : \ D_{m-1} \neq [1, b-1]\}$ is infinite and that there exist non-negative real numbers $\alpha_0, \alpha_1, \ldots, \alpha_{b-1}$ such that for all $m \geq 1$ and for all $\ell \in [0, b-1]$, the limit

$$\lim_{m \to +\infty} \frac{1}{m} |\{i \in [0, m-1] \ : \ |D_i| = \ell\}|$$

exists and is equal to α_ℓ. Then the restricted Dirichlet series $F_{L_{\mathrm{forb}(\mathcal{D})}}(s)$ has abscissa of convergence

$$\Theta_{\mathcal{D}} = \frac{1}{\log b} \sum_{\ell=0}^{b-1} \alpha_\ell \log(b - \ell).$$

Observe that Theorem 3 clearly implies Theorem 2. However, we propose to give a shorter proof of the first theorem using the proof and methods from the second theorem. As Köhler and Spilker did in the proof of Theorem 2, we require [26, Lemma 1], which is (partially) recalled below and for which Köhler and Spilker give a standard reference, namely [36, p. 292].

Lemma 1. *Let $(a_n)_{n \geq 1}$ be a sequence. Assume that the series $A(n) = \sum_{i=1}^n a_i$ of partial sums is positive and tends to infinity. Then the abscissa of convergence of the Dirichlet series $F_a(s)$ is*

$$\limsup_{n \to +\infty} \frac{\log A(n)}{\log n}.$$

Proof (of Theorem 3). First of all, the hypothesis means that, for all $\ell \in [0, b-1]$,

$$|\{i \in [0, m-1] : |D_i| = \ell\}| = m\alpha_\ell + o_\ell(m), \tag{1}$$

where

$$\lim_{m \to +\infty} \frac{o_\ell(m)}{m} = 0. \tag{2}$$

As already mentioned, our strategy to prove the statement is inspired by the proof of Theorem 2 and makes use of Lemma 1. To that aim, we define the sequences $(a_n)_{n\geq 1}$ and $(A(n))_{n\geq 1}$ by

$$a_n = \begin{cases} 1, & \text{if } \mathrm{rep}_b(n) \in L_{\mathrm{forb}(\mathcal{D})}; \\ 0, & \text{otherwise,} \end{cases}$$

and $A(n) = |\{m \in [1, n] : \mathrm{rep}_b(m) \in L_{\mathrm{forb}(\mathcal{D})}\}|$ for all $n \geq 1$. Note that the nth partial sum of the sequence $(a_n)_{n\geq 1}$ is $A(n)$. Therefore, using the notation from Lemma 1, we have $F_a(s) = F_{L_{\mathrm{forb}(\mathcal{D})}}(s)$. To prove that the abscissa of convergence of the restricted Dirichlet series $F_{L_{\mathrm{forb}(\mathcal{D})}}(s)$ is equal to $\Theta_{\mathcal{D}}$, it suffices, by Lemma 1, to prove that

$$\limsup_{n \to +\infty} \frac{\log A(n)}{\log n} = \Theta_{\mathcal{D}}. \tag{3}$$

To evaluate the quantity $A(n)$ for large enough n, we define k by $b^{k-1} \leq n < b^k$, i.e., $k = \lfloor \frac{\log n}{\log b} \rfloor + 1$. Of course, we have

$$A(b^{k-1}) \leq A(n) \leq A(b^k). \tag{4}$$

We divide the rest of the proof into two steps. Firstly, we prove that the lim sup in (3) is bounded above by $\Theta_{\mathcal{D}}$. Secondly, we build up an increasing sequence of integers such that the corresponding lim sup is bounded below by $\Theta_{\mathcal{D}}$.

Step 1: First, we prove that the lim sup in (3) is less than $\Theta_{\mathcal{D}}$. We note that $\mathrm{rep}_b(b^\ell) = 10^\ell$ for all $\ell \geq 0$. So, up to padding representations with leading zeroes, we can consider that all integers in the interval $[1, b^k - 1]$ can be represented by words of length k on $[0, b-1]$. Hence $A(b^k - 1)$ can be bounded from above by the number of words of length k on $[0, b-1]$ such that their i-th digits belong to $[0, b-1] \setminus D_i$, i.e.,

$$A(b^k - 1) \leq \left(\prod_{i=0}^{k-1} (b - |D_i|) \right) - 1.$$

Note that we subtracted 1 because the word 0^k does not represent an integer in $[1, b^k]$ (actually, it doesn't represent any integer, since the number 0 is represented by the empty word). Thus

$$A(b^k) \leq A(b^k - 1) + 1 \leq \prod_{i=0}^{k-1} (b - |D_i|).$$

By (1), we obtain

$$A(b^k) \leq \prod_{i=0}^{k-1}(b-|D_i|) = \prod_{\ell=0}^{b-1}(b-\ell)^{\alpha_\ell k + o_\ell(k)} \leq b^{\sum_{\ell=0}^{b-1}|o_\ell(k)|}\prod_{\ell=0}^{b-1}(b-\ell)^{\alpha_\ell k}. \quad (5)$$

Putting (4) and (5) together gives

$$\log A(n) \leq (k\log b)\left(\sum_{\ell=0}^{b-1}\frac{|o_\ell(k)|}{k} + \frac{1}{\log b}\sum_{\ell=0}^{b-1}\alpha_\ell \log(b-\ell)\right).$$

Since $k = \lfloor \frac{\log n}{\log b} \rfloor + 1$ implies that $k\log b \sim \log n$ when n goes to infinity, we get

$$\limsup_{n\to+\infty}\frac{\log A(n)}{\log n} \leq \lim_{k\to\infty}\left(\sum_{\ell=0}^{b-1}\frac{|o_\ell(k)|}{k} + \frac{1}{\log b}\sum_{\ell=0}^{b-1}\alpha_\ell \log(b-\ell)\right) = \Theta_{\mathcal{D}},$$

where the last equality follows using (2). This ends the first step. ■

Step 2: Now we prove that there exists an increasing sequence $(m_t)_{t\geq 0}$ of integers such that $\limsup_{m_t\to+\infty} \frac{\log A(m_t)}{\log m_t} \geq \Theta_{\mathcal{D}}$. We will in fact use the sequence $(m_t)_{t\geq 0}$ consisting of the elements of $\mathcal{M}$ in increasing order (recall that $\mathcal{M}$ is defined as in Theorem 3 above and that $\mathcal{M}$ is infinite). In short, for all $t \geq 0$, m_t is the $(t+1)$st element of $\mathcal{M}$.

First, for each $t \geq 0$, we consider the set of integers in $[b^{m_{t-1}}, b^{m_t}-1]$, whose base-b representations have length (exactly) m_t and are such that their ith digit is not in D_i, i.e., the set

$$\left\{n \geq 1 : n \in [b^{m_{t-1}}, b^{m_t}-1] \text{ and } \operatorname{rep}_b(n) \in L_{\mathrm{forb}(\mathcal{D})}\right\}.$$

We let P_{m_t} denote the cardinality of this set. Observe that $A(b^{m_t}) \geq P_{m_t}$, so in order to bound the lim sup above, we will evaluate P_{m_t} in the following. Since base-b representations do not start with 0, we have

$$P_{m_t} = \begin{cases} \displaystyle\prod_{i=0}^{m_t-1}(b-|D_i|), & \text{if } 0 \in D_{m_t-1}; \\ \displaystyle\left(\frac{b-|D_{m_t-1}|-1}{b-|D_{m_t-1}|}\right)\prod_{i=0}^{m_t-1}(b-|D_i|), & \text{if } 0 \notin D_{m_t-1}. \end{cases} \quad (6)$$

(Note that, if $0 \notin D_{m_t-1}$, then there are $b-|D_{m_t-1}|-1$ possible choices for the digit in position $m_t - 1$ as leading zeroes are forbidden.) Following [30], we distinguish two cases to evaluate P_{m_t}, depending on the cardinality $|D_{m_t-1}|$.

Case 1: If $0 \leq |D_{m_t-1}| \leq b-2$, then

$$\frac{b-|D_{m_t-1}|-1}{b-|D_{m_t-1}|} \geq \frac{1}{2},$$

thus

$$P_{m_t} \geq \frac{1}{2} \prod_{i=0}^{m_t-1} (b - |D_i|) = \frac{1}{2} \prod_{\ell=0}^{b-1} (b-\ell)^{\alpha_\ell m_t + o_\ell(m_t)}.$$

where we used (1) in the last equality.

Case 2: If $|D_{m_t-1}| = b-1$, then $0 \in D_{m_t-1}$ (recall that $m_t \in \mathcal{M}$). Thus, using the first equality in Eq. (6), we have

$$P_{m_t} = \prod_{i=0}^{m_t-1} (b - |D_i|) = \prod_{\ell=0}^{b-1} (b-\ell)^{\alpha_\ell m_t + o_\ell(m_t)}$$

by using (1) again.

Hence, in both cases, we obtain

$$P_{m_t} \geq \frac{1}{2} \prod_{\ell=0}^{b-1} (b-\ell)^{\alpha_\ell m_t + o_\ell(m_t)} \geq \frac{1}{2}\, b^{-\sum_{\ell=0}^{b-1} |o_\ell(m_t)|} \prod_{\ell=0}^{b-1} (b-\ell)^{\alpha_\ell m_t},$$

since (distinguishing the cases $o_\ell(m_t) \geq 0$ and $o_\ell(m_t) < 0$)

$$o_\ell(m_t) \log(b-\ell) \geq -|o_\ell(m_t)| \log b.$$

This in turn yields

$$\log A(b^{m_t}) \geq \log P_{m_t} \geq -\log 2 - \log b \sum_{\ell=0}^{b-1} |o_\ell(m_t)| + m_t \sum_{\ell=0}^{b-1} \alpha_\ell \log(b-\ell),$$

and thus

$$\frac{\log A(b^{m_t})}{\log b^{m_t}} \geq \frac{\log P_{m_t}}{m_t \log b} \geq -\frac{\log 2}{m_t \log b} - \sum_{\ell=0}^{b-1} \frac{|o_\ell(m_t)|}{m_t} + \frac{1}{\log b} \sum_{\ell=0}^{b-1} \alpha_\ell \log(b-\ell).$$

Finally, by using (2), we obtain

$$\limsup_{m_t \to +\infty} \frac{\log A(b^{m_t})}{\log b^{m_t}} \geq \frac{1}{\log b} \sum_{\ell=0}^{b-1} \alpha_\ell \log(b-\ell) = \Theta_{\mathcal{D}}.$$

This finishes up the two steps of the proof. The proof is now complete. □

3 Opening the door to a myriad of cases

In this section, we show how Köhler and Spilker's strategy using Lemma 1 may open the way to generalizations of Theorems 2 and 3 (note that none of these results may be applied to the cases considered in Section 3). We focus on words that periodically avoid some blocks. Counting words having a prescribed structure is notably done in several papers of Janjić [17,18,19,20,21]. An important paper on the subject is of course the 1999 Noonan and Zeilberger paper [31], that describes and implements the “Goulden-Jackson cluster method”.

We start with a lemma that will prove itself useful.

Lemma 2. *Let $(a_n)_{n\geq 1}$ be a sequence of non-negative numbers and let $(A(n))_{n\geq 1}$ be its summatory function, i.e., $A(n) = \sum_{1\leq i\leq n} a_i$. Suppose that there exist an integer $b \geq 2$ and a real $\lambda > 1$ such that $A(b^k) = \lambda^{k(1+o(1))}$ when k goes to infinity. Then, $\frac{\log A(n)}{\log n} \sim \frac{\log \lambda}{\log b}$ when n goes to infinity.*

Note that the assumption in the previous lemma is equivalent to saying that $\log A(b^k) = k(1 + o(1)) \log \lambda$, i.e., $\log A(b^k) \sim k \log \lambda$.

Proof (of Lemma 2). For an integer $n \geq 1$, let k be such that $b^{k-1} \leq n < b^k$ (i.e., $k-1 = \lfloor \frac{\log n}{\log b} \rfloor$). We have $k \sim \frac{\log n}{\log b}$ when n goes to infinity. Since $A(b^{k-1}) \leq A(n) \leq A(b^k)$, we have

$$\frac{\log A(b^{k-1})}{\log n} \leq \frac{\log A(n)}{\log n} \leq \frac{\log A(b^k)}{\log n}$$

for $n \geq 2$. By assumption both $\log A(b^k)$ and $\log A(b^{k-1})$ behave like $k \log \lambda$, thus

$$\lim_{n\to+\infty} \frac{\log A(b^{k-1})}{\log n} = \frac{\log \lambda}{\log b} \quad \text{and} \quad \lim_{n\to+\infty} \frac{\log A(b^k)}{\log n} = \frac{\log \lambda}{\log b}$$

hence, the squeeze theorem finally gives $\lim_{n\to+\infty} \frac{\log A(n)}{\log n} = \frac{\log \lambda}{\log b}$, as desired. □

3.1 All-distinct-letter blocks

We play with a specific example in base 10 where representations periodically avoid two blocks that do not share a letter. More specifically, define L to be the language of all the words $w_k w_{k-1} \cdots w_0$ over $[0, 9]$ such that $w_k \neq 0$ and, for all $i \in [0, k-1]$,

$$w_{i+1} w_i \neq \begin{cases} 12, & \text{if } i \text{ is even;} \\ 89, & \text{if } i \text{ is odd;} \end{cases}$$

Observe that L contains base-10 representations that periodically avoid the blocks 12 and 89. Note that the blocks 12 and 89 do not have a letter in common. Also note that if words in L contain the block 12 (resp., 89), then it must be at an odd (resp., even) position; it can also be the case that they do not contain them at all.

The goal of this section is to determine the abscissa of convergence of the restricted Dirichlet series $F_L(s)$. To this aim, for all $n \geq 0$, we let z_n denote the cardinality of the sub-language $L \cap [0, 9]^n$, i.e., z_n gives the number of length-n words in L. One can also compute the first few elements of the sequence $(z_n)_{n\geq 0}$ by hand and obtain $1, 9, 89, 881, 8721$. Plugging them in Sloane's On-Line Encyclopedia of Integer Sequences (OEIS), we find the related sequences [34, A072256] and [34, A138288]. In the description of the first, we read that the nth term gives the number of "01-avoiding words of length n on the alphabet $[0, 9]$ which do not end in 0". Quite clearly, by reversing the reading order, it also gives the number of length-n words over $[0, 9]$ not starting with 0 and avoiding the pattern 10. If we denote the latter sequence by $(x_n)_{n\geq 0}$, Sloane's OEIS provides a recurrence relation given by $x_0 = 1$, $x_1 = 9$, and, for all $n \geq 0$,

$$x_{n+2} = 10x_{n+1} - x_n. \tag{7}$$

We now establish a bijection between the language L and that containing words satisfying the condition described above. We start off with the following result.

Proposition 1. *Let M_1 be the set of words over the alphabet $[0,9]$ containing at least one of the blocks $12, 89$ and such that every block 12 (resp., 89) necessarily occurs at even (resp., odd) positions. Let M_2 be the set of words over the alphabet $[0,9]$ containing the block 12. Define the map $f : [0,9]^* \to [0,9]^*$ that sends a word to that obtained by replacing all occurrences of the blocks 89 (resp., 12) in odd positions by the block 12 (resp., 89). Then the map $f_1 : M_1 \to M_2, u \mapsto f(u)$ is a bijection.*

Proof. First, note that for all $u \in M_1$ we have $f_1(u) \in M_2$. Indeed, if u contains a block 12 occurring at an even position, then 12 is clearly a factor of $f_1(u)$, so $f_1(u) \in M_2$. Otherwise, u necessarily contains a block 89 at an odd position and similarly, $f_1(u) \in M_2$.

Now consider the map $f_2 : M_2 \to M_1, v \mapsto f(v)$. Proving that $f_2(M_2) \subseteq M_1$ follows the same lines as above. Now observe that we have $f_2(f_1(u)) = u$ for all $u \in M_1$ and $f_1(f_2(v)) = v$ for all $v \in M_2$. □

Using Proposition 1, complementing the languages, and getting rid of words starting with 0, we link our sequence of interest and that of Sloane's.

Corollary 2. *For all $n \geq 0$, we have $z_n = x_n$.*

Corollary 2 and the work in [26, §3] imply the next result. Indeed, in that section, Köhler and Spilker consider the language of base-b representations avoiding a fixed block of digits and determine the abscissa of convergence of the corresponding restricted Dirichlet series.

Theorem 4. *Consider the dominant root $\lambda = 5 + 2\sqrt{6}$ of the characteristic polynomial $P(x) = x^2 - 10x + 1$ of the recurrence in (7). Then the abscissa of convergence of the restricted Dirichlet series $F_L(s)$ is $\frac{\log \lambda}{\log 10}$.*

An alternative approach to computing the number z_n of length-n words in L uses the Goulden-Jackson cluster method; see, for example, [31]. This general technique provides an algorithm to determine the generating function for the number of length-n words over a finite alphabet that avoid some given finite set of patterns. At first glance L is not defined in terms a finite list of avoided patterns (since it depends on the parity of the position), but we can nevertheless resort to the following trick: expand the alphabet from $\{0, 1, \ldots, 9\}$ to an alphabet of twice the size, containing both primed and unprimed digits. Let us agree that an unprimed digit corresponds to an even position, and an primed digit corresponds to an odd position. Define d_n to be the number of length-n words over this larger alphabet that avoid ab and $a'b'$ for $0 \le a, b < 10$ and also $12'$ and $8'9$. Then, the Goulden-Jackson cluster method, as implemented in Maple by the `DAVID_IAN` package, demonstrates that $\sum_{n\ge 0} d_n x^n = (1+10x-x^2)/(1-10x+x^2)$. It now follows that $(d_{n+1}-d_n)/2$ is equal to z_n, where the division by 2 arises because we could begin with either 0 or $0'$.

3.2 Blocks sharing letters

We now turn to the case where base-10 representations periodically avoid two blocks sharing common letters. It is a little trickier to handle and is *a priori* neither a consequence of Köhler and Spilker's results [26] nor Nathanson's [30]. As in the previous section, we define L to be the language of all words $w_k w_{k-1} \cdots w_0$ over $[0,9]$ such that $w_k \neq 0$ and, for all $i \in [0, k-1]$,

$$w_{i+1} w_i \neq \begin{cases} 12, & \text{if } i \text{ is even;} \\ 21, & \text{if } i \text{ is odd;} \end{cases}$$

and also define L' to be the same language without the condition $w_k \neq 0$. Observe that L contains base-10 representations that periodically avoid the blocks 12 and 21.

Similarly, we want to determine the abscissa of convergence of the restricted Dirichlet series $F_L(s)$. So, for all $n \ge 0$, we let z_n denote the cardinality of the sub-language $L \cap [0,9]^n$ and, for all $n \ge 1$, we also define the sets

$$\begin{aligned} S_n &= \{w_{n-1} \cdots w_0 \in L : w_{n-1} \neq 2 \text{ if } n \text{ is even and } w_{n-1} \neq 1 \text{ otherwise}\}, \\ P_n &= \{w_{n-1} \cdots w_0 \in L : w_{n-1} = 2 \text{ if } n \text{ is even and } w_{n-1} = 1 \text{ otherwise}\}, \\ Q_n &= \{w_{n-1} \cdots w_0 \in L' : w_{n-1} = 0\}. \end{aligned}$$

We have $|S_1| = 8$, $|P_1| = 1 = |Q_1|$, and

$$\begin{aligned} |S_{n+1}| &= 8(|S_n| + |P_n| + |Q_n|), \\ |P_{n+1}| &= |S_n| + |Q_n|, \\ |Q_{n+1}| &= |S_n| + |P_n| + |Q_n|, \end{aligned}$$

for all $n \ge 1$. For instance, let us explain the first equality. To build up a word in S_{n+1}, we may prepend a letter $a \in [0,9]$ to a word w of length n, according to one of the next three cases: if $w \in S_n$ (resp., P_n; resp., Q_n), then $a \in [0,9]$ except 0, and either 1 or 2, depending on the parity; this gives 8 possibilities.

As we have $(z_n)_{n\ge 1} = (|S_n| + |P_n|)_{n \ge 1}$, the first few values of $(z_n)_{n\ge 0}$ are

$$1, 9, 89, 882, 8739, 86589, 857952, 8500869, 84229389, 834572322.$$

Contrarily to the previous section, this sequence does not *directly* belong to the OEIS. A fortunate error made by the authors nevertheless lead them to the sequence [34, A322054]. For all $n \ge 0$, its nth term gives the number of "decimal strings of length n that do not contain a specific string aa (where a is a fixed single digit)". If we denote the latter by $(x_n)_{n\geq 0}$, the OEIS provides us with a recurrence relation given by $x_0 = 1$, $x_1 = 10$, and

$$x_{n+2} = 9(x_{n+1} + x_n) \tag{8}$$

for all $n \geq 0$. Its first few terms are

$$1, 10, 99, 981, 9720, 96309, 954261, 9455130, 93684519.$$

We will establish that the sequence $(z_n)_{n\geq 0}$ can be obtained as the first difference of the sequence $(x_n)_{n\geq 0}$. We start with a result analoguous to Proposition 1.

Proposition 2. *Let M_1 be the set of words over the alphabet $[0,9]$ containing at least one of the blocks $12, 21$ and such that every block 12 (resp., 21) necessarily occurs at even (resp., odd) positions. Let M_2 be the set of words over the alphabet $[0,9]$ containing the block aa, for a fixed letter $a \in [0,9]$. Define the map $f_a : [0,9]^* \to [0,9]^*$ that sends every word to a word obtained by replacing every 1 (resp., 2) in an even (resp., odd) position by the letter a and every letter a in an even (resp., odd) position by the letter 1 (resp., 2). Then the map $f_{a,1} : M_1 \to M_2, u \mapsto f_a(u)$ is a bijection. Furthermore, if the letter a is non-zero, then $z_{n+1} = x_{n+1} - x_n$ for all $n \geq 0$.*

Proof. The proof of the first part of the statement follows the same lines as that of Proposition 1 by considering the map $f_{a,2} : M_2 \to M_1, v \mapsto f_a(v)$ instead.

Let us prove the second part. The number of decimal length-$(n+1)$ words avoiding aa, which is x_{n+1}, is equal to the sum of the number of decimal length-n words avoiding aa with the number of decimal length-$(n+1)$ words avoiding aa and not starting with 0. The first number is x_n by definition and the second is z_{n+1} by the first part of the statement. Indeed, the first part of the statement implies that the number of length-n words in M_1 is equal to the number of decimal length-n words containing aa. Now complementing the two languages, getting rid of words starting with 0 and since $a \neq 0$, we obtain $x_{n+1} = z_{n+1} + x_n$, as desired. □

Theorem 5. *Consider the dominant root $\lambda = \frac{3}{2}(3 + \sqrt{13})$ of the characteristic polynomial $P(x) = x^2 - 9x - 9$ of the recurrence in (8). Then the abscissa of convergence of the restricted Dirichlet series $F_L(s)$ is $\frac{\log \lambda}{\log 10}$.*

Proof. Let a_n be equal to 1 if $\text{rep}_{10}(n) \in L$, 0 otherwise, and let $A(n)$ denote the summatory function of the sequence $(a_n)_{n\geq 0}$. By Lemma 1, the abscissa of convergence of the restricted Dirichlet series $F_L(s)$ is governed by the asymptotics of $(\log |A(n)|/\log n)_{n\geq 0}$. Observe that the behavior of $A(10^n)$ is determined by that of $z_n = x_{n+1} - x_n$ by Proposition 2. Now note that the sequence $(x_n)_{n\geq 1}$ satisfies $x_n \sim \lambda^n$ when n goes to infinity. Thus $z_n \sim C\lambda^n$ with $C = (\lambda - 1)/\lambda$ by Proposition 2. So Lemma 2 implies that the abscissa of convergence of $F_L(s)$ is $\frac{\log \lambda}{\log 10}$. □

We can also count the number of words in L using the Goulden-Jackson cluster method, using the same idea as in Section 3.1. Define e_n to be the number of length-n words over this larger alphabet that avoid ab and $a'b'$ for $0 \leq a, b < 10$ and also $12'$ and $2'1$. Then, the Goulden-Jackson cluster method, as implemented in Maple by the `DAVID_IAN` package, demonstrates that $\sum_{n\geq 0} e_n x^n = (1+11x+9x^2)/(1-9x-9x^2)$. It now follows that $(e_{n+1} - e_n)/2$ is equal to z_n, where again the division by 2 arises because we could begin with either 0 or $0'$.

3.3 *To infinity and beyond!*

Other results, involving in particular several sequences in the OEIS, can be worked out using similar methods, but also by using the method in [31]. Here, we give examples that do not use [31]. For instance, inspired by the previous section, we fix two integers $b, k \geq 2$ and we define a sequence $(y_n)_{n\geq 0}$ analogous to $(x_n)_{n\geq 0}$ from Section 3.2: we set $y_n = b^n$ for all $n \in [0, k-1]$ and, for all $n \geq k$,

$$y_n = (b-1) \cdot \sum_{i=1}^{k} y_{n-i}. \tag{9}$$

Proposition 3. *The sequence $(y_{n+1} - y_n)_{n\geq 0}$ counts the number of length-$(n+1)$ words over $[0, b-1]$ that do not start with 0 and that avoid the k-th power a^k, for some fixed letter $a \in [0, b-1]$.*

Proof. For all $m \geq 0$, y_m gives the number of length-m words over $[0, b-1]$ that do not contain the kth power a^k, so $y_{n+1} - y_n$ gives the expected number. $\square$

Example 2. For all $k \geq 2$, the sequence $(y_n)_{n\geq 0}$ corresponding to $b = 2$ gives the k-bonacci numbers: indeed, binary k-bonacci words avoid the block 1^k. For $k = 2$ and $b = 10$, we find back the sequence $(x_n)_{n\geq 0}$ defined by Equation (8). Other cases are displayed in Table 1.

Table 1. Depending on the values of the parameters k and b, the entry number in [34] of the sequence $(y_n)_{n\geq 0}$ is given (sometimes, one needs to crop the first few terms).

k	b	entry number of $(y_n)_{n\geq 0}$ in [34]
2	3	A028859, A155020
2	4	A125145
2	5	A086347
2	6	A180033
2	7	A180167
2	10	A322054
3	3	A119826
3	4	A282310

We study the characteristic polynomial associated with the sequence $(y_n)_{n\geq 0}$. Recall that a real number is *Pisot* if it is an algebraic integer whose (other) algebraic conjugates have modulus less than 1.

Proposition 4. *The characteristic polynomial $P(x) = x^k - (b-1)\sum_{i=0}^{k-1} x^i$ of the recurrence in* (9) *has a dominant root that is a Pisot number and belongs to the interval* $(1, b)$.

Proof. From the proof of [6, Theorem 2 (p. 254)], we deduce that the polynomial $P(x)$ has one real root β with modulus larger than 1 and all other roots with modulus less than 1. Next, observe that $P(1) = 1 - k(b-1) < 0$ and $P(b) = b^k - (b-1)\frac{b^k-1}{b-1} = 1 > 0$, so $\beta \in (1, b)$. All in all, this proves the statement. □

Let us come back to our Dirichlet series of interest.

Theorem 6. *Let L be the set of words over $[0, b-1]$ that do not start with 0 and that avoid the k-th power a^k, for some fixed letter $a \in [0, b-1]$. Let λ be the dominant root of the characteristic polynomial of the recurrence in* (9)*. Then the abscissa of convergence of the series $F_L(s)$ is equal to $\frac{\log \lambda}{\log b}$.*

Proof (Sketch). The proof is similar to that of Theorem 5. By Proposition 4, the growth rate of the sequence $(y_n)_{n\geq 0}$ is λ^n, so that of the sequence of first differences $(y_{n+1} - y_n)_{n\geq 1}$ is $C\lambda^n$ with $C = (\lambda-1)/\lambda$. We conclude by using Lemmas 1 and 2 together with Proposition 3. □

4 An alternative approach

As indicated in Lemma 1, the abscissa of convergence of the Dirichlet series $F_a(s) = \sum_{n\geq 1} a_n/n^s$ can be deduced from the asymptotic behavior of the summatory function $A(n) = \sum_{1\leq i\leq n} a_i$. For the case of a q-automatic (or q-regular) sequence $(a_n)_{n\geq 0}$, an abundant literature studies this behavior: we will restrict ourselves to cite one of the papers of Dumont and Thomas [10], the chapter of Drmota and Grabner [9] (in [4]), and the paper of Heuberger and Krenn [16]. In this section, we will take advantage of the fact that the sequences $(a_n)_{n\geq 0}$ considered so far are automatic. Nevertheless, as will be seen in Section 5 (see Remark 2), our previous method can be applied for certain non-automatic sequences.

Before giving an alternative proof of the main results in Sections 3.1 and 3.2, we show how the following result by Drmota and Grabner handles the simple case of avoiding one letter.

Theorem 7 (Drmota and Grabner, [4, Theorem 9.2.15]). *Let $b \geq 2$ be an integer. Let $(s_n)_{n\geq 0}$ be a b-regular sequence whose linear representation is given by the vectors L, R and the matrices $M_0, \ldots, M_{b-1}$. Write $M = \sum_{i=0}^{b-1} M_i$ and assume that M has a unique eigenvalue $\lambda > 0$ of maximal modulus and that λ has algebraic multiplicity 1. Assume further that $\lambda > \max_i ||M_i||$ for some* *matrix norm* $||\cdot||$. *Let* μ *be the modulus of the second largest eigenvalue. Then there exists a periodic continuous function* Φ *such that*

$$\sum_{i<n} s_i = n^{\log_b \lambda}\Phi(\log_b n) + O(n^{\log_b \mu}) + O(\log n).$$

We obtain a result similar to Corollary 1.

Corollary 3. *Let* $b \geq 2$ *be an integer and let* a *be a non-zero letter in* $[1, b-1]$. *Let* L *be the language over* $[0, b-1]$ *of base-*b *representations that avoid the letter* a. *Then the abscissa of convergence of the series* $F_L(s)$ *is equal to* $\log_b(b-1)$.

Proof (Sketch). We use the notation of Theorem 7. For all $n \geq 0$, let $s_n = 1$ if $\text{rep}_b(n) \in L$, $s_n = 0$ otherwise. By definition, we have $s_{bn+c} = s_n$ for all $c \neq a$ and $s_{bn+a} = 0$. The sequence $(s_n)_{n\geq 0}$ is b-regular with $M_c = 1$ for all $c \neq a$ and $M_a = 0$. Thus $\lambda = b - 1$. Using Lemma 1 and Theorem 7, the abscissa of convergence of the Dirichlet series $F_L(s)$ is $\log_b(b-1)$. □

In the following, we let L_1 (resp., L_2) denote the language L from Section 3.1 (resp., Section 3.2). As before, we stick to Drmota and Grabner's notation and we define the sequences $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$ by $s_n = 1$ if $\text{rep}_{10}(n) \in L_1$, $s_n = 0$ otherwise and similarly, $t_n = 1$ if $\text{rep}_{10}(n) \in L_2$, $t_n = 0$ otherwise. The first few terms of $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$ are $1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1$. We have $|L_1 \cap [0, 9]^n| = \sum_{i=10^n}^{10^{n+1}} s_i$ and $|L_2 \cap [0, 9]^n| = \sum_{i=10^n}^{10^{n+1}} t_i$. Both sequences $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$ are 10-automatic: a deterministic finite automaton with output (DFAO) reading base-10 representations and outputting each sequence is given in Fig. 1. Note that the computations in this section are done using the *Mathematica* package IntegerSequences [32,33] written by Rowland. In fact, the b-kernel of a b-automatic sequence is in bijection with the set of states in the minimal b-automaton that generates it with the property that leading zeroes do not affect the output [3, Theorem 6.6.2]. In our case, the 10-kernel of $(s_n)_{n\geq 0}$ contains the four sequences

$$\begin{aligned}
(s_n)_{n\geq 0} &= 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, \ldots, \\
(s_{10n})_{n\geq 0} &= 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, \ldots, \\
(s_{10n+2})_{n\geq 0} &= 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, \ldots, \\
(s_{100n+90})_{n\geq 0} &= 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, \ldots,
\end{aligned}$$

and the zero-sequence $(s_{100n+12})_{n\geq 0} = 0, 0, 0, \ldots$. The 4×4 matrix whose columns are built on the terms corresponding to $n \in \{0, 1, 8, 12\}$ in each sequence has a non-zero determinant. Consequently, the $\mathbb{Q}$-vector space generated by the 10-kernel of $(s_n)_{n\geq 0}$ is finitely generated by the four linearly independent sequences $(s_n)_{n\geq 0}$, $(s_{10n})_{n\geq 0}$, $(s_{10n+2})_{n\geq 0}$, and $(s_{100n+90})_{n\geq 0}$. Using the 10-automaton in the top of Fig. 1, one builds up a linear representation $(L, M_0, \ldots, M_9, R)$ of $(s_n)_{n\geq 0}$ viewed as a 10-regular sequence. For instance, we

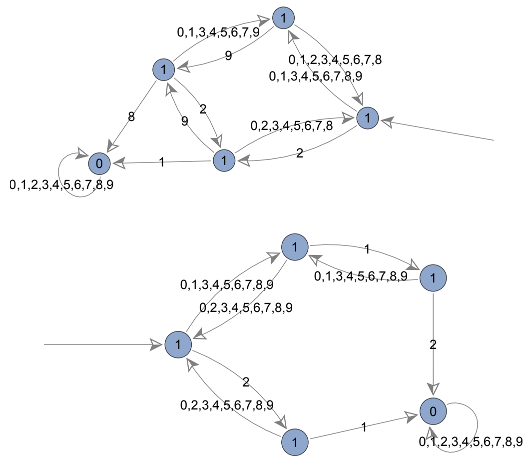

Fig. 1. The sequences $(s_n)_{n\geq 0}$ and $(t_n)_{n\geq 0}$ are generated by the base-10 DFAO's on top and on bottom respectively.

obtain the relations $s_{100n} = s_n$, $s_{100n+20} = s_n$, and $s_{1000n+900} = s_{10n}$, so

$$\begin{pmatrix} s_{10n} \\ s_{100n} \\ s_{100n+20} \\ s_{1000n+900} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}}_{=M_0} \begin{pmatrix} s_n \\ s_{10n} \\ s_{10n+2} \\ s_{100n+90} \end{pmatrix}.$$

The sum matrix $M_0 + \cdots + M_9$ has eigenvalues $\pm(5+2\sqrt{6})$ and $\pm(5-2\sqrt{6})$ (each with multiplicity 1) and does not fullfill the condition of Theorem 7. However, Heuberger and Krenn's more general result [16, Theorem A] shows that the growth of the main term of $\sum_{i=1}^{n} s_i$ is $n^{\log_{10}(5+2\sqrt{6})}$. (Note that we take the liberty not to recall [16, Theorem A] in full as its statement is quite lengthy.) Taking the logarithm and applying Lemma 1 thus gives an alternative proof of Theorem 4. A similar reasoning can be conducted for the sequence $(t_n)_{n\geq 0}$ to obtain an alternative proof of Theorem 5.

To end this section, let us consider yet a different example. We let L denote the language of all words $w_k w_{k-1} \cdots w_0$ over $[0, 9]$ such that $w_k \neq 0$ and, for all $i \in [0, k-1]$,

$$w_{i+1} w_i \neq \begin{cases} 12 \text{ or } 89, & \text{if } i \text{ is even;} \\ 89, & \text{if } i \text{ is odd.} \end{cases}$$

We define the sequence $(u_n)_{n\geq 0}$ by $u_n = 1$ if $\mathrm{rep}_{10}(n) \in L$, $u_n = 0$ otherwise. The first values of $(|L \cap [0, 9]^n|)_{n\geq 0}$ are $1, 9, 88, 872, 8534, 84566, 827622$, which does not seem to be in the OEIS. Computing further values of the latter sequence seems difficult with a classical computer set-up. The DFAO in Fig. 2

generates the 10-automatic sequence $(u_n)_{n\geq 0}$. Following the approach of [16,

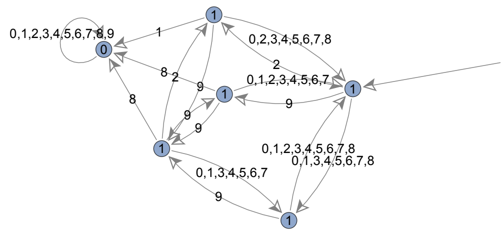

Fig. 2. The base-10 DFAO generates the sequence $(u_n)_{n\geq 0}$.

Theorem A] and from the DFAO in Fig. 2, we deduce a linear representation of $(u_n)_{n\geq 0}$ viewed as a 10-regular sequence, whose sum matrix has eigenvalues $\pm\sqrt{\frac{1}{2}\left(97+\sqrt{9401}\right)}$, $\pm\frac{2}{\sqrt{97+\sqrt{9401}}}$, and 0 (each having multiplicity 1). Combining Lemma 1 and [16, Theorem A] gives us the following result.

Theorem 8. *Let* $\lambda = \sqrt{\frac{1}{2}\left(97+\sqrt{9401}\right)}$. *Then the abscissa of convergence of the restricted Dirichlet series* $F_L(s)$ *is* $\frac{\log \lambda}{\log 10}$.

5 An unusual example

Let us consider an example. Let $\mathbf{t} = (t_n)_{n\geq 0} = 01101001\cdots$ be the Thue-Morse word. We say that an integer n is *evil* if $t_n = 0$. A non-evil number is called *odious*. The first few evil numbers are $0, 3, 5, 6, 9, 10, 12, 15, 17, 18, 20$.

Let L be the set of binary words of the form $w_k w_{k-1}\cdots w_0$ with the property that there is no factor 10 where the 0 appears in an evil position i for some $0 \leq i \leq k$. (Note that leading zeroes are allowed.) For example, $10111 \notin L$ because the word 10 appears as $w_4 w_3$ and 3 is evil. Define $a_n = 1$ if $\mathrm{rep}_2(n) \in L$.

Let us find a recurrence for the number $b_n = |L \cap \{0,1\}^n|$ of length-n words in L. Here the Goulden-Jackson cluster method does not seem to be applicable.

Lemma 3. *We have* $b_0 = 1$, $b_1 = 2$, $b_2 = 3$, *and for all* $n \geq 3$,

$$b_n = \begin{cases} 2b_{n-1}, & \text{if } t_{n-2} = 1; \\ b_{n-1} + b_{n-3}, & \text{if } t_{n-2} = t_{n-3} = 0; \\ b_{n-1} + b_{n-2}, & \text{if } t_{n-2} = 0 \text{ and } t_{n-3} = 1. \end{cases} \tag{10}$$

Proof. Let $L_n = L \cap \{0,1\}^n$ denote the set of all words $w_{n-1}\cdots w_0$ having no factor 10 where the 0 appears in an evil position, so that $b_n = |L_n|$. Suppose $n \geq 3$.

Case 1: Assume that $t_{n-2} = 1$. Then $L_n = 0L_{n-1} \cup 1L_{n-1}$, because appending either 0 or 1 to a word in L_{n-1} cannot introduce a forbidden word. This gives the first equality in (10).

Case 2: Assume that $t_{n-2} = t_{n-3} = 0$. Then $L_n = 0L_{n-1} \cup 111L_{n-3}$. The first term in (10) arises because appending a 0 to a word in L_{n-1} cannot introduce a forbidden word, and the second term arises because the first three symbols of a word in L_n cannot be 100, 101, or 110. This gives the second equality in (10).

Case 3: Assume that $t_{n-2} = 0$ and $t_{n-3} = 1$. Then $L_n = 0L_{n-1} \cup 11L_{n-2}$. The situation is like the previous case, except now the first two symbols of a word in L_n cannot be 10. This ends the proof. □

The first few values of the sequence $(b_n)_{n\geq 0}$ from Lemma 3 are given in Table 2.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
b_n	1	2	3	6	12	18	36	54	72	144	288	432	576	1152	1728	3456	6912	10368	20736	31104	41472

Table 2. The number b_n of length-n words in L, i.e., words with the property that there is no factor 10 where the 0 appears in an evil position.

We expect b_n to behave roughly like α^n, for some α with $1 < \alpha < 2$, because forbidding 10 with 0 in an evil position is more or less like forbidding 10 "half the time". This intuition is confirmed in the following result.

Theorem 9. *There are constants c_1, c_2 such that $n^{c_1}\alpha^n \leq b_n \leq n^{c_2}\alpha^n$, where $\alpha = 24^{1/6} \approx 1.69838$.*

We require the next lemma.

Lemma 4. *For $n \geq 3$, we have*

$$b_n = \begin{cases} 2b_{n-1}, & \text{if } t_{n-2} = 1; \\ \frac{4}{3}b_{n-1}, & \text{if } t_{n-2} = t_{n-3} = 0; \\ \frac{3}{2}b_{n-1}, & \text{if } t_{n-2} = 0 \text{ and } t_{n-3} = 1. \end{cases} \tag{11}$$

Proof. The proof is by induction on n. The base cases are left to the reader.

Case 1: If $t_{n-2} = 1$, then the first equality in (11) follows immediately from (10).

Case 2: Assume that $t_{n-2} = t_{n-3} = 0$. Since $\mathbf{t}$ avoids cubes, we must have $t_{n-4} = 1$. Then (10) gives $b_n = b_{n-1} + b_{n-3}$ and $b_{n-1} = b_{n-2} + b_{n-3}$, and the induction hypothesis gives and $b_{n-2} = 2b_{n-3}$. Adding these last two equations and cancelling b_{n-2} gives $b_{n-1} = 3b_{n-3}$. The second equality in (11) now follows.

Case 3: Assume that $t_{n-2} = 0$ and $t_{n-3} = 1$. Then (10) gives $b_n = b_{n-1} + b_{n-2}$ and the induction hypothesis gives $b_{n-1} = 2b_{n-2}$. The result now follows. □

Proof (of Theorem 9). For a block $u \in \{0,1\}^*$ define $e_u(n)$ to be the number of (possibly overlapping) occurrences of u in $\mathbf{t}[0..n-1]$. A simple induction, using (11), now gives

$$b_n = 2^{e_1(n-1)+2e_{00}(n-1)-e_{10}(n-1)} 3^{1+e_{10}(n-1)-e_{00}(n-1)} \tag{12}$$

for $n \geq 2$. For instance, if $t_{n-2} = t_{n-3} = 0$, then $\mathbf{t}[0..n-2] = \mathbf{t}[0..n-4]00$, so $e_1(n-1) = e_1(n-2)$, $e_{10}(n-1) = e_{10}(n-2)$, and $e_{00}(n-1) = e_{00}(n-2)+1$. This together with the induction hypothesis and the second equality in (11) yields Equation (12). The other two cases can be handled in a similar fashion.

Since it is known that

$$\begin{aligned} e_1(n) &= n/2 + O(1), \\ e_{00}(n) &= n/6 + O(\log n), \\ e_{10}(n) &= n/3 + O(\log n), \end{aligned}$$

the desired result now follows. □

Define L' to be the subset of L that contains binary words not starting with 0. For all $n \geq 0$, we let

$$a_n = \begin{cases} 1, & \text{if } \mathrm{rep}_2(n) \in L'; \\ 0, & \text{otherwise.} \end{cases}$$

The following observation particularly shows that one cannot use Theorem 7 nor [16, Theorem A] to study the behavior of the summatory function of the sequence $(a_n)_{n\geq 0}$.

Remark 2. Writing $b_n = 2^{x_n}3^{y_n}$, we see from Equation (12) that $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ are both 2-regular sequences, but $(b_n)_{n\geq 0}$ is not (as it grows too quickly).

On the other hand, the sequence $(a_n)_{n\geq 0}$ is not 2-automatic. Consider those n with base-2 representation of the form 101^i for some $i \geq 0$, i.e., $n = 3 \cdot 2^i - 1$. Then $a_n = 1$ if and only if i is odious, i.e., $t_i = 1$. By a well-known theorem ([3, Theorem 5.5.2], if $(a_n)_{n\geq 0}$ is 2-automatic, then $(a_{3\cdot 2^i-1})_{i\geq 0}$ is ultimately periodic. But $a_{3\cdot 2^i-1} = t_i$, and $\mathbf{t} = (t_i)_{i\geq 0}$ cannot be ultimately periodic because it is overlap-free.

Nevertheless, we can handle the behavior of the summatory function of $(a_n)_{n\geq 0}$, and thus the abscissa of convergence of the corresponding language L', as follows.

Corollary 4. *Let $\alpha = \sqrt[6]{24}$. The abscissa of convergence of the restricted Dirichlet series $F_{L'}(s)$ is $\frac{\log \alpha}{\log 2}$.*

Proof. For all $n \geq 0$, we let $A(n) = \sum_{0\leq i\leq n} a_i$. By Theorem 9, since leading zeroes are allowed in L, we have $n^{c_1}\alpha^n \leq A(2^n - 1) \leq n^{c_2}\alpha^n$. Now we may apply Lemma 2 first, then Lemma 1 to obtain the statement. □

6 Conclusion

Do the cases we have described call for a general theorem (and how general such a theorem can be)? We have chosen to give various examples, yielding methods to address other cases. The growth of the summatory function of a sequence gives the abscissa of convergence of the Dirichlet generating series of this sequence (Lemma 1). In the case of sequences related to missing digits, the point is to count words with some prescribed structure. Of course, other approaches than ours can be used (we think, e.g., of [14]), and we have already mentioned the "Goulden-Jackson cluster method" and the paper [31].

Note that the coefficients of the Dirichlet series that we have studied here are q-automatic or q-regular ([3]): one can ask about the abscissa of convergence of general q-automatic or q-regular Dirichlet series. It is worth briefly explaining why *automatic* Dirichlet series are interesting. A sequence $(u_n)_n$ is b-automatic for some $b \geq 2$ if the set of subsequences $\{(u(b^k n + j)_n \mid k \geq 0, j \in [0, b^k - 1]\}$ is finite. This implies equalities between these subsequences, which in turn yield properties (e.g., functional properties) of the corresponding Dirichlet series hence of the Dirichlet series we started from. This has been used in a different context (see, e.g., [1,8]). Another class of Dirichlet series related to combinatorics on words that has been studied consists of Dirichlet series with Sturmian coefficients (see, e.g., [27,35]). A class of Dirichlet series more general than the Sturmian series was studied in [25].

Acknowledgements

Manon Stipulanti is an FNRS Research Associate supported by the Research grant 1.C.104.24F. We thank France Gheeraert, Daniel Krenn, Eric Rowland, and Pierre Stas for useful discussions when this work was in its early stammering.

References

1. Allouche, J.-P., Mendès France, M., Peyrière, J.: Automatic Dirichlet series. J. Number Theory **81**(2), 359–373 (2000). https://doi.org/10.1006/jnth.1999.2487, https://doi.org/10.1006/jnth.1999.2487
2. Allouche, J.-P., Shallit, J., Skordev, G.: Self-generating sets, integers with missing blocks, and substitutions. Discrete Math. **292**(1-3), 1–15 (2005). https://doi.org/10.1016/j.disc.2004.12.004
3. Allouche, J.-P., Shallit, J.: Automatic sequences. Theory, Applications, Generalizations. Cambridge University Press, Cambridge (2003). https://doi.org/10.1017/CBO9780511546563
4. Berthé, V., Rigo, M. (eds.): Combinatorics, automata and number theory, Encyclopedia of Mathematics and its Applications, vol. 135. Cambridge University Press, Cambridge (2010). https://doi.org/10.1017/CBO9780511777653
5. Birmajer, D., Gil, J.B., Weiner, M.D.: On the enumeration of restricted words over a finite alphabet. J. Integer Seq. **19**(1), Article 16.1.3, 16 (2016), https://cs.uwaterloo.ca/journals/JIS/VOL19/Gil/gil6.html

6. Brauer, A.: On algebraic equations with all but one root in the interior of the unit circle. Math. Nachr. **4**, 250–257 (1951). https://doi.org/10.1002/mana.3210040123
7. de Bruijn, N.G., Knuth, D.E., Rice, S.O.: The average height of planted plane trees. In: Graph theory and computing, pp. 15–22. Academic Press, New York (1972)
8. Coons, M.: (Non)automaticity of number theoretic functions. J. Théor. Nombres Bordeaux **22**(2), 339–352 (2010), http://jtnb.cedram.org/item?id=JTNB_2010__22_2_339_0
9. Drmota, M., Grabner, P.J.: Analysis of digital functions and applications. In: Combinatorics, automata and number theory, Encyclopedia Math. Appl., vol. 135, pp. 452–504. Cambridge Univ. Press, Cambridge (2010)
10. Dumont, J.-M., Thomas, A.: Systèmes de numération et fonctions fractales relatifs aux substitutions. Theoret. Comput. Sci. **65**(2), 153–169 (1989). https://doi.org/10.1016/0304-3975(89)90041-8
11. Ekhad, S.B., Zeilberger, D.: Counting clean words according to the number of their clean neighbors. ACM Commun. Comput. Algebra **57**(1), 5–9 (2023). https://doi.org/10.1145/3610377.3610379
12. Flajolet, P., Sedgewick, R.: Analytic combinatorics. Cambridge University Press, Cambridge (2009). https://doi.org/10.1017/CBO9780511801655
13. Fogg, N.P.: Substitutions in dynamics, arithmetics and combinatorics, Lecture Notes in Mathematics, vol. 1794. Springer-Verlag, Berlin (2002). https://doi.org/10.1007/b13861, edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel
14. Guibas, L.J., Odlyzko, A.M.: String overlaps, pattern matching, and nontransitive games. J. Combin. Theory Ser. A **30**(2), 183–208 (1981). https://doi.org/10.1016/0097-3165(81)90005-4, https://doi.org/10.1016/0097-3165(81)90005-4
15. von Haeseler, F.: Automatic sequences, De Gruyter Expositions in Math., vol. 36. Walter de Gruyter & Co., Berlin (2003). https://doi.org/10.1515/9783110197969
16. Heuberger, C., Krenn, D.: Asymptotic analysis of regular sequences. Algorithmica **82**(3), 429–508 (2020). https://doi.org/10.1007/s00453-019-00631-3
17. Janjić, M.: On linear recurrence equations arising from compositions of positive integers. J. Integer Seq. **18**(4), Article 15.4.7, 14 (2015), https://cs.uwaterloo.ca/journals/JIS/VOL18/Janjic/janjic63.html
18. Janjić, M.: Binomial coefficients and enumeration of restricted words. J. Integer Seq. **19**(7), Article 16.7.3, 18 (2016), https://cs.uwaterloo.ca/journals/JIS/VOL19/Janjic/janjic73.html
19. Janjić, M.: Some formulas for numbers of restricted words. J. Integer Seq. **20**(6), Art. 17.6.5, 10 (2017), https://cs.uwaterloo.ca/journals/JIS/VOL20/Janjic/janjic83.html
20. Janjic, M.: Pascal matrices and restricted words. J. Integer Seq. **21**(5), Art. 18.5.2, 13 (2018), https://cs.uwaterloo.ca/journals/JIS/VOL21/Janjic2/janjic103.html
21. Janjić, M.: Words and linear recurrences. J. Integer Seq. **21**(1), Art. 18.1.4, 17 (2018), https://cs.uwaterloo.ca/journals/JIS/VOL21/Janjic/janjic93.html
22. Kärki, T., Lacroix, A., Rigo, M.: On the recognizability of self-generating sets. In: Mathematical foundations of computer science 2009, Lecture Notes in Comput. Sci., vol. 5734, pp. 525–536. Springer, Berlin (2009). https://doi.org/10.1007/978-3-642-03816-7_45
23. Kärki, T., Lacroix, A., Rigo, M.: On the recognizability of self-generating sets. J. Integer Seq. **13**(2), Article 10.2.2, 18 (2010), https://cs.uwaterloo.ca/journals/JIS/VOL13/Rigo/rigo6.html

24. Kempner, A.J.: A curious convergent series. Amer. Math. Monthly **21**(2), 48–50 (1914). https://doi.org/10.2307/2972074
25. Knill, O., Lesieutre, J.: Analytic continuation of Dirichlet series with almost periodic coefficients. Complex Anal. Oper. Theory **6**(1), 237–255 (2012). https://doi.org/10.1007/s11785-010-0064-7, https://doi.org/10.1007/s11785-010-0064-7
26. Köhler, G., Spilker, J.: Dirichlet-Reihen zu Kempners merkwürdiger konvergenter Reihe. Math. Semesterber. **56**(2), 187–199 (2009). https://doi.org/10.1007/s00591-009-0059-5
27. Kwon, D.: A one-parameter family of Dirichlet series whose coefficients are Sturmian words. J. Number Theory **147**, 824–835 (2015). https://doi.org/10.1016/j.jnt.2014.08.018, https://doi.org/10.1016/j.jnt.2014.08.018
28. Mandelbrojt, S.: Dirichlet Series. D. Reidel Publishing Co., Dordrecht (1972), Principles and Methods
29. Mukherjee, R., Sarkar, N.: A short note on a curious convergent series. Asian-Eur. J. Math. **14**(9), Paper No. 2150158 (2021). https://doi.org/10.1142/S1793557121501588
30. Nathanson, M.B.: Dirichlet series of integers with missing digits. J. Number Theory **222**, 30–37 (2021). https://doi.org/10.1016/j.jnt.2020.10.002
31. Noonan, J., Zeilberger, D.: The Goulden-Jackson cluster method: Extensions, applications and implementations. J. Difference Equ. Appl. **5**(4-5), 355–377 (1999). https://doi.org/10.1080/10236199908808197
32. Rowland, E.: IntegerSequences software package, https://github.com/ericrowland/IntegerSequences
33. Rowland, E.: IntegerSequences: a package for computing with k-regular sequences. In: Mathematical software – ICMS 2018. 6th international conference, South Bend, IN, USA, July 24–27, 2018. Proceedings, pp. 414–421. Cham: Springer (2018). https://doi.org/10.1007/978-3-319-96418-8_49
34. Sloane, N.J.A., et al.: The On-Line Encyclopedia of Integer Sequences, https://oeis.org
35. Sourmelidis, A.: On the meromorphic continuation of Beatty zeta-functions and Sturmian Dirichlet series. J. Number Theory **194**, 303–318 (2019). https://doi.org/10.1016/j.jnt.2018.07.009, https://doi.org/10.1016/j.jnt.2018.07.009
36. Titchmarsh, E.C.: The theory of functions. Oxford University Press, Oxford, second edn. (1939)
37. Wilf, H.S.: generatingfunctionology. A K Peters, Ltd., Wellesley, MA, third edn. (2006), https://web.archive.org/web/20230322192727/https://www2.math.upenn.edu/~wilf/DownldGF.html
38. Wintner, A.: On Riemann's reduction of Dirichlet series to power series. Amer. J. Math. **69**, 769–789 (1947). https://doi.org/10.2307/2371798