Output sum of transducers:
Limiting distribution and periodic fluctuation

Clemens Heuberger[*]

Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Klagenfurt, Austria
clemens.heuberger@aau.at

Sara Kropf[*]

Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Klagenfurt, Austria
sara.kropf@aau.at

Helmut Prodinger[†]
Department of Mathematical Sciences,
Stellenbosch University
Stellenbosch, South Africa
hproding@sun.ac.za

Abstract

As a generalization of the sum of digits function and other digital sequences, sequences defined as the sum of the output of a transducer are asymptotically analyzed. The input of the transducer is a random integer in $[0, N)$. Analogues in higher dimensions are also considered. Sequences defined by a certain class of recursions can be written in this framework.

Depending on properties of the transducer, the main term, the periodic fluctuation and an error term of the expected value and the variance of this sequence are established. The periodic fluctuation of the expected value is Hölder continuous and, in many cases, nowhere differentiable. A general formula for the Fourier coefficients of this periodic function is derived. Furthermore, it turns out that the sequence is asymptotically normally distributed for many transducers. As an example, the abelian complexity function of the paperfolding sequence is analyzed. This sequence has recently been studied by Madill and Rampersad.

Keywords: Central limit theorem, periodic fluctuation, Fourier coefficient, transducer, automatic sequence, non-differentiability.

1 Introduction

 Over the last decades, asymptotic properties of digital sequences have been studied by many authors. The simplest example is the q-ary sum of digits, see Delange [7]. This has been generalized to various other number systems (cf. [26], [27], [12], [34], [18], [3], [15], [16], [23], [21]). Similar results have been obtained for other digital sequences (cf. [6] and [4]). Frequently observed phenomena in the asymptotic analysis of these sequences include periodic fluctuations in the second order term and asymptotic normality (see also [9]).

The purpose of this article is to use finite state machines as a uniform framework to derive such asymptotic results. The results mentioned above will follow as corollaries from our main results, see the end of the introduction for more details. As an example of a new result fitting into this framework, we study the abelian complexity function of the paperfolding sequence (cf. [28]), see Example 2.8.

Our main focus lies on transducers: these finite state machines transform input words to output words using a finite memory (see Section 2 for a more precise definition). In our case, the input is the q-ary digit expansion of a random integer in the interval $[0, N)$. We then asymptotically study the sum of the output of the transducer for $N \to \infty$. This is also extended to higher dimensions.

While some of the examples can easily be formulated by transducers, other examples are more readily expressed in terms of recursions of the shape

$$a(q^{\kappa}n + \lambda) = a(q^{\kappa_\lambda}n + r_\lambda) + t_\lambda \quad \text{for} \quad 0 \leqslant \lambda < q^{\kappa} \tag{1}$$

with fixed κ, κ_λ, $r_\lambda \in \mathbb{Z}$, $t_\lambda \in \mathbb{R}$ and $\kappa_\lambda < \kappa$. We transform such a recursion into a transducer in Theorem 4 in Section 2.6.

Several notions abstracting the sum-of-digits and related problems have been studied. One of them is the notion of completely q-additive functions $a : \mathbb{N}_0 \to \mathbb{R}$ with

$$a(qn + \lambda) = a(n) + a(\lambda)$$

for $0 \leqslant \lambda < q$ (cf. [4]). These have been generalized to digital sequences as defined in [1, 6]: A sequence $a(n)$ is a digital sequence if it can be represented as a sum $\sum_w f(w)$ where f is a given function and w runs over all windows of a fixed length κ of the q-ary digit representation of n. These digital sequences can easily be formulated by a recursion as in (1).

For a transducer $\mathcal{T}$, let $\mathcal{T}(n)$ be the sum of the output labels of $\mathcal{T}$ when reading the q-ary expansion of n. For a positive integer N, we study the behavior of $\mathcal{T}(n)$ for a uniformly chosen random n in $\{0, \ldots, N-1\}$. Assuming suitable connectivity properties of the underlying graph of the transducer, we obtain the following results.

- The expected value is given by

$$\mathbb{E}(\mathcal{T}(n)) = e_{\mathcal{T}} \log_q N + \Psi_1(\log_q N) + o(1)$$

for a constant $e_{\mathcal{T}}$ and a periodic, continuous function Ψ_1 (Theorem 1).

An extended abstract with less general Theorems 1, 2 and 4 and without proofs appears as [22].

- The variance is

$$\mathbb{V}(\mathcal{T}(n)) = v_{\mathcal{T}} \log_q N - \Psi_1^2(\log_q N) + \Psi_2(\log_q N) + o(1)$$

with constant $v_{\mathcal{T}}$ and a periodic, continuous function $\Psi_2(x)$ (Theorem 1).

- After suitable renormalization, $\mathcal{T}(n)$ is asymptotically normally distributed (Theorem 1).

- The Fourier coefficients of Ψ_1 are given explicitly in Theorem 2 and the Fourier series converges absolutely and uniformly.

- The function Ψ_1 is nowhere differentiable provided that $e_{\mathcal{T}}$ is not an integer (Theorem 3).

The exact assumptions for the various results are given in detail in the respective theorems. Results for higher dimensional input are available for expectation, variance, normal distribution and Fourier coefficients.

Our theorems are generalizations of the following known results.

- For the sum of digits of the standard q-ary digit representations (cf. [7]), we obtain an asymptotic normal distribution, the Fourier coefficients and the non-differentiability (for even[1] q). The error term vanishes, as stated in Remark 3.4. Therefore, the formula is not only asymptotic but also exact. The formulas for the Fourier coefficients by Delange [7] also follow from our Theorem 2.

- The occurrence of subblocks in standard and non-standard digit representations is defined by a strongly connected, aperiodic transducer. Thus we obtain the expected value, the variance, the limit law and the Fourier coefficients (cf. [26, 27, 15] for the expected value). For one dimensional digit representations, we also obtain the non-differentiability (assuming $e_{\mathcal{T}} \neq 0, 1$) of the fluctuation in the expectation.

- The Hamming weight is a special case of the occurrence of subblocks. Thus, Theorem 1 is a generalization of the results about the width-w non-adjacent form [21], the simple joint sparse form [16] and the asymmetric joint sparse form [21].

- A transducer defining a completely q-additive function consists of only one state. Therefore, we obtain an asymptotic normal distribution (as in [4]), the Fourier coefficients and the non-differentiability (assuming $e_{\mathcal{T}} \notin \mathbb{Z}$ and integer output). Here, the error term vanishes, too.

- A digital sequence is defined by a strongly connected, aperiodic transducer. Thus, digital sequences are asymptotically normally distributed or degenerate. Assuming $e_{\mathcal{T}} \notin \mathbb{Z}$ and integer output, the periodic fluctuation $\Psi_1(x)$ is non-differentiable. The Fourier coefficients can be computed by Theorem 2. See also [6] for results on the expected value.

---

[1]Our approach in Theorem 3 requires that the constant $e_{\mathcal{T}}$ of the main term of the expected value is not an integer. In this case, $e_{\mathcal{T}} = \frac{q-1}{2}$, which is an integer if q is odd.

- Automatic sequences [1] are also defined by transducers: The output labels of all transitions are 0 and the final output labels are as in the definition of such sequences. Theorem 1 gives the expected value with $e_{\mathcal{T}} = 0$ (see also [30]) and, depending on the transducer, also the variance with $v_{\mathcal{T}} = 0$. The Fourier coefficients of the periodic fluctuation of the expected value are given explicitly in Theorem 2.

- In [18], Grabner and Thuswaldner investigate the sum of digits function for negative bases $s_{-q}(n)$. They give a transducer to compute the function $s_{-q}(n) - s_{-q}(-n)$. Their result about the limit law follows directly from our Theorem 1.

As an example of a new result obtained by Theorem 1, we give an asymptotic estimate of the abelian complexity function of the paperfolding sequence in Example 2.8. In [28], the authors prove that this sequence satisfies a recursion of type (1). As consequences of Theorem 1, the expected value is $\sim \frac{8}{13} \log_2 N$, the variance is $\sim \frac{432}{2197} \log_2 N$ and the sequence is asymptotically normally distributed.

In the sequel, we discuss the relation of our setting and our results with the notion of q-regular sequences introduced in [1].

A sequence is q-regular if it is the first coordinate of a vector $\boldsymbol{v}(n)$ and there exist matrices $V_0, \ldots, V_{q-1}$ such that

$$\boldsymbol{v}(qn + \varepsilon) = V_{\varepsilon} \boldsymbol{v}(n) \tag{2}$$

for $\varepsilon \in \{0, 1, \ldots, q-1\}$.

The concept of q-regular sequences is more general than our setting, but a broader variety of asymptotic behavior is observed which precludes any generalization of our results to general q-regular sequences.

While $\mathcal{T}(n)$ is a q-regular sequence for any transducer $\mathcal{T}$ (see Remark 3.10), the converse is not necessarily true: Obviously, the sum of the output of a transducer reading the input n is always bounded by $\mathcal{O}(\log n)$. However, the 2-regular sequence[2]

$$a(n) = \begin{cases} n & \text{if } n \text{ is a power of } 2, \\ 0 & \text{otherwise} \end{cases}$$

can clearly not be bounded by $\mathcal{O}(\log n)$.

Asymptotic estimates for q-regular sequences are given by Dumas [10, 11]. By restricting our attention to sequences defined by transducers, we obtain an asymptotic estimate of the variance, explicit expressions for the Fourier coefficients of the fluctuation in the second term of the expected value, non-differentiability of this fluctuation as well as a central limit theorem.

Section 2 contains all the theorems and the required notions. In Section 2.2, Theorem 1, formulas for the first and second moment of the output sum of a transducer and its limiting distribution are presented. In Theorem 2 in Section 2.4, the Fourier coefficients of the periodic fluctuation $\Psi_1(x)$ of the expected value are stated. We discuss the non-differentiability of $\Psi_1(x)$ in Theorem 3 in Section 2.5.

---

[2]Use $\boldsymbol{v}(0) = (0,1)^{\top}$ (where $^{\top}$ denotes transposition), $V_0 = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix} \right)$ and $V_1 = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix} \right)$.

Section 2.6 deals with sequences satisfying the recursion (1) and higher dimensional analogues. We construct a transducer computing this sequence in Theorem 4. Thus, from Theorem 1, the expected value, the variance and the limit distribution follow in many cases.

This construction and the computations for the constants $e_{\mathcal{T}}$, $v_{\mathcal{T}}$ and the Fourier coefficients can be done algorithmically by the mathematical software system Sage [32]: The general framework is included in Sage version 6.4.1 using its finite state machine package described in [20]. The code for the Fourier coefficients and the construction from a recursion is submitted for inclusion in future versions of Sage, see http://trac.sagemath.org/17222 and http://trac.sagemath.org/17221, respectively.

In Sections 3 to 6, we give the proofs of all the theorems from Section 2.

2 Results

This section starts with the definition of some notions about the connectivity of a transducer. Then we will state the theorems about the moments and the limiting distribution, the Fourier coefficients, the non-differentiability, and the construction of a transducer computing a sequence given by a recursion as in (1).

2.1 Notions

We consider complete, deterministic and subsequential transducers (cf. [5, Chapter 1]). In our case, the input alphabet is $\{0, \ldots, q-1\}^d$ for a positive integer d and the output alphabet $\mathbb{R}$. A transducer is said to be *deterministic* and *complete* if for every state and every digit of the input alphabet, there is exactly one transition starting in this state with this input label. A *subsequential* transducer $\mathcal{T}$ (cf. [31]) is defined to be a finite deterministic automaton with one initial state, an output label for every transition and a final output label for every state.
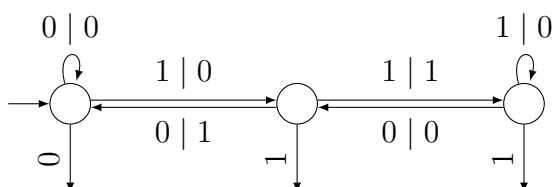

Figure 1: Transducer computing the Hamming weight of the non-adjacent form.

Figure 1 presents an example of a complete, deterministic, subsequential transducer. The label of a transition with input ε and output δ is written as $\varepsilon \mid \delta$.

The input of the transducer is the standard q-ary joint digit representation of an integer vector $\boldsymbol{n} \in \mathbb{N}_0^d$, i.e. the standard q-ary digit representation at each coordinate of the vector $\boldsymbol{n}$. The input is read from right (least significant digit) to left (most significant digit), without leading zeros. Then the output of the transducer is the sequence of the outputs of the transitions along the unique path starting in the initial state with the given

input and the final output of the last state of this path. The element $\mathcal{T}(\boldsymbol{n})$ of the sequence defined by the transducer $\mathcal{T}$ is the sum of this output sequence.

Using final output labels is convenient for our purposes. Clearly, it would also be possible to model the final output labels by using an "end-of-input" marker and additional transitions. In the context of digital expansions, the behavior can usually also be obtained by reading a sufficient number of leading zeros. But the approach using final outputs is more general as it is not required that the final outputs are compatible with the output generated by leading zeros.

For the various results, different properties of the complete, deterministic, subsequential transducer and its underlying digraph are needed. All states of the underlying digraph are assumed to be accessible from the initial state. Contracting each strongly connected component of the underlying digraph gives an acyclic digraph, the so-called condensation. A strongly connected component is said to be *final strongly connected* if it corresponds to a leaf (i.e., a vertex with outdegree 0) in the condensation. Let c be the number of final strongly connected components. We call a transducer or a digraph *finally connected* if $c = 1$.

For the asymptotic expressions, only the final strongly connected components are important. All other strongly connected components only influence the error term. Thus, we are not interested in the periodicity of the whole underlying digraph, but in the periodicity of the final strongly connected components. The *period* of a digraph is defined as the greatest common divisor of all lengths of directed cycles of the digraph. For $j = 1, \ldots, c$, let p_j be the period of the final strongly connected component C_j. Define the *final period* of the digraph as

$$p = \operatorname{lcm}\{p_j \mid j = 1, \ldots, c\}.$$

We call a digraph *finally aperiodic* if $p = 1$. If the underlying digraph is strongly connected, its final period is equal to its period.

For proving the non-differentiability of the fluctuation, we not only need a finally aperiodic, finally connected digraph ($p = c = 1$), but also a reset sequence. A *reset sequence* is an input sequence such that starting at any state and reading this sequence leads to a specific state s. If the transducer is not finally aperiodic and finally connected, then there cannot exist a reset sequence.

2.2 Moments and Limiting Distribution

This section contains the theorem about the moments of the output sum $\mathcal{T}(\boldsymbol{n})$ and the limiting distribution. Further results about the periodic fluctuation can be found in Theorems 2 and 3.

As probability space, we use $\Omega_N = \{0, 1, \ldots, N-1\}^d$ endowed with the equidistribution measure.

Denote by Φ_{μ,σ^2} the cumulative distribution function of the normal distribution with mean μ and variance $\sigma^2 \neq 0$. Thus,

$$\Phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right) dy.$$

Theorem 1. *Let $d \geqslant 1$, $\mathcal{T}$ be a complete, deterministic, subsequential transducer with input alphabet $\{0, 1, \ldots, q-1\}^d$, output alphabet $\mathbb{R}$, final period p, and c final components.*

Then $\mathcal{T}(\boldsymbol{n})$ has the expected value

$$\mathbb{E}(\mathcal{T}(\boldsymbol{n})) = e_{\mathcal{T}} \log_q N + \Psi_1(\log_q N) + \mathcal{O}(N^{-\xi} \log N) \tag{3}$$

where the constants $e_{\mathcal{T}}$ and $\xi > 0$ are given in (5) in Section 2.3 and $\Psi_1(x)$ is a p-periodic, Hölder continuous function.

If all b_j given in (5) are positive, the distribution function of $\mathcal{T}(\boldsymbol{n})$ can be approximated by a mixture of c Gaussian distributions with weights λ_j, means $a_j \log_q N$ and variances $b_j \log_q N$ for some constants a_j and $\lambda_j > 0$ with $\sum_{j=1}^{c} \lambda_j = 1$, given in (5). In particular,

$$\mathbb{P}\left(\frac{\mathcal{T}(\boldsymbol{n})}{\sqrt{\log_q N}} \leqslant x \right) = \sum_{j=1}^{c} \lambda_j \Phi_{a_j \sqrt{\log_q N}, b_j}(x) + \mathcal{O}\left(\log^{-\frac{1}{2}} N \right)$$

for all $x \in \mathbb{R}$.

If all a_j are equal, then $\mathcal{T}(\boldsymbol{n})$ has the variance

$$\mathbb{V}(\mathcal{T}(\boldsymbol{n})) = v_{\mathcal{T}} \log_q N - \Psi_1^2(\log_q N) + \Psi_2(\log_q N) + \mathcal{O}(N^{-\xi} \log^2 N) \tag{4}$$

with constant $v_{\mathcal{T}} \in \mathbb{R}$ (given in (5)) and a p-periodic, continuous function $\Psi_2(x)$. Otherwise, the variance is $\mathbb{V}(\mathcal{T}(\boldsymbol{n})) = \Theta(\log^2 N)$.

If all a_j are equal, $\mathcal{T}(\boldsymbol{n})$ converges in distribution to a mixture of Gaussian (or degenerate) distributions with means 0 and variances b_j, weighted by λ_j. In particular, if all $b_j > 0$,

$$\mathbb{P}\left(\frac{\mathcal{T}(\boldsymbol{n}) - \mathbb{E}(\mathcal{T}(\boldsymbol{n}))}{\sqrt{\log_q N}} \leqslant x \right) = \sum_{j=1}^{c} \lambda_j \Phi_{0, b_j}(x) + \mathcal{O}\left(\log^{-\frac{1}{2}} N \right)$$

holds for all $x \in \mathbb{R}$.

If furthermore $c = 1$ and $v_{\mathcal{T}} \neq 0$, then $\mathcal{T}(\boldsymbol{n})$ is asymptotically normally distributed.

We give the proof of this theorem in Section 3.

Remark 2.1. The assumption that $b_j > 0$ is essential for obtaining uniform convergence of the distribution function and the speed of convergence in particular. To see this, consider the transducer in Figure 2. It is easily seen that $\mathcal{T}(n) = (-1)^n$. For even N, the distribution function of $\mathcal{T}(n)/\sqrt{\log_2 N}$ is given by

$$\mathbb{P}\left(\frac{\mathcal{T}(n)}{\sqrt{\log_2 N}} \leqslant x \right) = \begin{cases} 0 & \text{if } x < -1/\sqrt{\log_2 N}, \\ 1/2 & \text{if } -1/\sqrt{\log_2 N} \leqslant x < 1/\sqrt{\log_2 N}, \\ 1 & \text{if } 1/\sqrt{\log_2 N} \leqslant x, \end{cases}$$

which does not converge uniformly.

$$0 \mid 0, 1 \mid 0$$
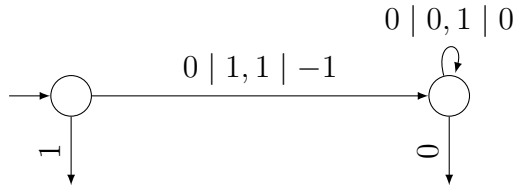
$$0 \mid 1, 1 \mid -1$$

Figure 2: Transducer for Remark 2.1.

2.3   Eigenvalues and Eigenvectors of the Transition Matrix

For the constants in Theorem 1 and the Fourier coefficients in Theorem 2, we need the notion of a transition matrix of the transducer and properties of its eigenvalues and eigenvectors.

We label the states of the transducer with contiguous positive integers starting with 1. We denote the indicator vector of the initial state by $\boldsymbol{e}_1$.

Definition 2.2. Let $t \in \mathbb{R}$ be in a neighborhood of 0.

The transition matrix $M_{\boldsymbol{\varepsilon}}$ for $\boldsymbol{\varepsilon} \in \{0, \ldots, q-1\}^d$ is the matrix whose (s_1, s_2)-th entry is $e^{it\delta}$ if there is a transition from state s_1 to state s_2 with input label $\boldsymbol{\varepsilon}$ and output label δ, and 0 otherwise.

Let M be the sum of all these transition matrices.

Lemma 2.3. *There are differentiable functions $\mu_j(t)$ in a neighborhood of $t = 0$ for $j = 1, \ldots, c$ such that the dominant eigenvalues of M are $\mu_j(t) \exp(\frac{2\pi i l}{p})$ in this neighborhood of $t = 0$ for some of the $l \in \mathcal{P} = \{k \in \mathbb{Z} \mid -p/2 < k \leqslant p/2\}$. For each of these dominant eigenvalues, the algebraic and geometric multiplicities coincide. For $t = 0$, $\mu_j(0) = q^d$.*

The proof of this lemma is given in Section 3.

Let $l \in \mathbb{Z}$. Consider the (not necessarily orthogonal) projection onto the direct sum of the left eigenspaces of M corresponding to the eigenvalues $\mu_j(t) \exp(\frac{2\pi i l}{p})$ for $j = 1, \ldots, c$ such that the kernel is the direct sum of the remaining generalized left eigenspaces. Let $\boldsymbol{w}_l^{\top}(t)$ be the image of $\boldsymbol{e}_1^{\top}$ under this projection, where $^{\top}$ denotes transposition. The definition of $\boldsymbol{w}_l^{\top}(t)$ only depends on l modulo p.

We write $\boldsymbol{w}_l^{\top}$ for $\boldsymbol{w}_l^{\top}(0)$ and $\boldsymbol{w}_l'^{\top}$ for the derivative of $\boldsymbol{w}_l^{\top}(t)$ at $t = 0$. Furthermore, $\boldsymbol{w}_l^{\top}$ is either the null vector or a left eigenvector of M corresponding to the eigenvalue $q^d \exp(\frac{2\pi i l}{p})$.

Let C_j be a final component with corresponding indicator vector $\boldsymbol{c}_j$. Define the constants

$$\lambda_j = \boldsymbol{w}_0^{\top} \boldsymbol{c}_j.$$

In Section 3.1, we will show that $\lambda_j > 0$ and $\sum_{j=1}^{c} \lambda_j = 1$.

With these definitions, the constants in Theorem 1 can be expressed as

$$a_j = -iq^{-d}\mu_j'(0),$$

$$e_{\mathcal{T}} = \sum_{j=1}^{c} \lambda_j a_j,$$

$$b_j = \frac{\mu_j'(0)^2 - q^d\mu_j''(0)}{q^{2d}}, \tag{5}$$

$$v_{\mathcal{T}} = \sum_{j=1}^{c} \lambda_j b_j.$$

Finally, $\xi > 0$ is chosen such that all non-dominant eigenvalues of M have modulus strictly less than $q^{d-\xi}$ at $t = 0$.

These constants can be interpreted as follows: $a_j \log_q N$ and $b_j \log_q N$ are the main terms of the mean and the variance, respectively, of the output sum of the final component C_j. These expressions including the derivatives of the eigenvalues correspond to the formulas for mean and variance given in [13, Theorem IX.9]. The constants $e_{\mathcal{T}}$ and $v_{\mathcal{T}}$ are convex combinations of the corresponding constants of the final components C_j.

The positive weight λ_j in these convex combinations turns out to be the asymptotic probability of reaching the final component C_j. This is connected to the following interpretation of the left eigenvector $\boldsymbol{w}_0^\top$: If the final period p is 1, the entries of $\boldsymbol{w}_0^\top$ will be shown to be the asymptotic probabilities of reaching the corresponding states. This corresponds to the left eigenvector used in a steady-state analysis. If $p > 1$, these probabilities depend on the length of the input modulo p. Then, we will prove that $\boldsymbol{w}_0^\top$ gives the average of these probabilities taken over all residues modulo p. These interpretations are justified in Section 3.1.

2.4 Fourier Coefficients

This section contains the formulas for the Fourier coefficients of the periodic fluctuation $\Psi_1(x)$. For this purpose, we need the following definitions.

Let $\chi_k = \frac{2\pi ik}{p\log q}$ for $k \in \mathbb{Z}$ and $\boldsymbol{1}$ be a vector whose entries are all one.

The s-th coordinate of the vector $\boldsymbol{b}(\boldsymbol{n})$ is the sum of the output of the transducer $\mathcal{T}$ (including the final output) if starting in state s with input the q-ary joint expansion of $\boldsymbol{n}$. In particular, the first coordinate of $\boldsymbol{b}(\boldsymbol{n})$ is $\mathcal{T}(\boldsymbol{n})$, and $\boldsymbol{b}(0)$ is the vector of final outputs. Furthermore, define the vector-valued function $\boldsymbol{H}(z)$ by the Dirichlet series

$$\boldsymbol{H}(z) = \sum_{\substack{\boldsymbol{n} \geqslant 0 \\ \boldsymbol{n} \neq 0}} \boldsymbol{b}(\boldsymbol{n}) \|\boldsymbol{n}\|_\infty^{-z}, \tag{6}$$

where the inequality in the summation index is considered coordinate-wise and $\|\cdot\|_\infty$ is the maximum norm.

Theorem 2. *Let $\mathcal{T}$ be a subsequential, complete, deterministic transducer. Then the Fourier coefficients of the p-periodic fluctuation $\Psi_1(x)$ are*

$$c_0 = -\frac{e_{\mathcal{T}}}{d \log q} - i\boldsymbol{w}_0'^{\top}\boldsymbol{1} + \frac{1}{d}\operatorname{Res}_{z=d}\boldsymbol{w}_0^{\top}\boldsymbol{H}(z),$$

$$c_k = \frac{1}{d+\chi_k}\operatorname{Res}_{z=d+\chi_k}\boldsymbol{w}_k^{\top}\boldsymbol{H}(z) \tag{7}$$

for $k \neq 0$.

The Fourier series $\sum_{k\in\mathbb{Z}} c_k \exp(\frac{2\pi ik}{p}x)$ converges absolutely and uniformly.

The function $\boldsymbol{w}_k^{\top}\boldsymbol{H}(z)$ is meromorphic in $\Re z > d-1$. It has a possible double pole at $z = d$ for $k = 0$ and possible simple poles at $z = d + \chi_k$ for $k \neq 0$.

The proof of this theorem is in Section 4.

The infinite recursion given in Lemma 4.5 can be used to numerically evaluate the Dirichlet series $\boldsymbol{H}(z)$ with arbitrary precision and to compute its residues at $z = d + \chi_l$ (see Lemma 4.7 and [17]). For $d = 1$, the computation of the Fourier coefficients can be done by the mathematical software system Sage [32] (using the code submitted at http://trac.sagemath.org/17222).

Example 2.4. The (artificial) transducer in Figure 3 has two final components with periods 2 and 3, respectively. Thus the final period is 6 and the function $\Psi_1(x)$ is 6-periodic. The constant $e_{\mathcal{T}}$ of the expected value is $\frac{11}{8}$. In Figure 4, the partial Fourier series with 2550 Fourier coefficients[3] is compared with the empirical values of the periodic fluctuation Ψ_1, i.e.,

$$\frac{1}{N}\sum_{n<N}\mathcal{T}(n) - \frac{11}{8}\log_2 N \tag{8}$$

with integers N and $4 \leqslant \log_2 N \leqslant 16$.

The computation of these 2550 Fourier coefficients took less than 6 minutes using a standard dual-core PC.
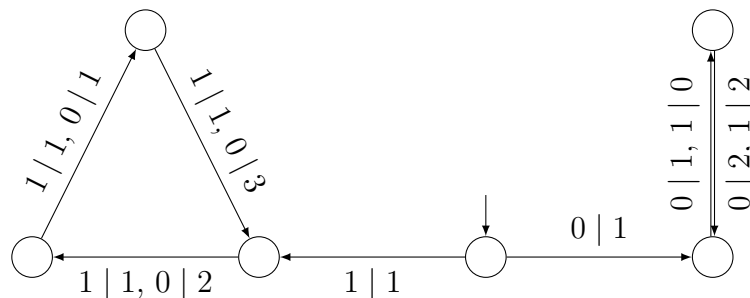

Figure 3: Transducer of Example 2.4: All states are final with final output 0.

---

[3]We use 2550 Fourier coefficients in this plot because the period length of the next summand of the Fourier series in Figure 4 is already less than the resolution of a standard printer.
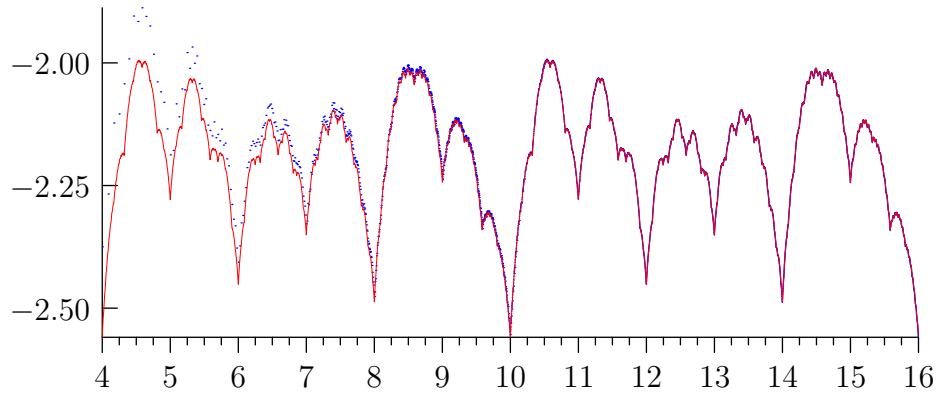
Figure 4: Partial Fourier series compared with the empirical values of the function Ψ_1 of Example 2.4.

In Example 2.8 we compute the first 2550 Fourier coefficients of the abelian complexity function of the paperfolding sequence.

As a corollary of Theorem 2, we obtain the following result which was already proved by Delange [7].

Corollary 2.5. *The Fourier coefficients of the periodic fluctuation*

$$\Psi_1(\log_q N) = \frac{1}{N} \sum_{n < N} s_q(n) - \frac{q-1}{2} \log_q N$$

for the q-ary sum-of-digits function $s_q(n)$ are

$$
\begin{aligned}
c_0 &= \frac{q-1}{2 \log q}(\log(2\pi) - 1) - \frac{q+1}{4}, \\
c_k &= -\frac{q-1}{\chi_k(1 + \chi_k) \log q} \zeta(\chi_k)
\end{aligned}
\tag{9}
$$

for $k \neq 0$ and $\chi_k = \frac{2\pi ik}{\log q}$ where ζ denotes the Riemann ζ-function.

We prove this corollary in Section 4.

2.5 Non-differentiability

In this section, we prove that for certain transducers, the periodic fluctuation $\Psi_1(x)$ of the expected value is nowhere differentiable.

Theorem 3. *Let $d = 1$. Assume that $e_{\mathcal{T}} \notin \mathbb{Z}$ and that the transducer $\mathcal{T}$ has a reset sequence and output alphabet $\mathbb{Z}$. Then the function $\Psi_1(x)$ is non-differentiable for any $x \in \mathbb{R}$.*

The proof can be found in Section 5. There, we follow the method presented by Tenenbaum [33], see also Grabner and Thuswaldner [18].

In [33, 18], the reset sequence consists only of 0's. If working with digit expansions, it is often possible to choose such a reset sequence. However, in the context of recursions, this is not always possible, see Example 2.8. There the reset sequence is (00001).

For a general finally aperiodic, finally connected transducer, the existence of a reset sequence cannot be guaranteed.

2.6 Recursions

In this section, we describe how to reduce a recursion to a transducer computing the given sequence. All inequalities in this section are considered coordinate-wise.

Let $q \geqslant 2$, κ, $\kappa_{\boldsymbol{\lambda}} \in \mathbb{Z}$, $\boldsymbol{r}_{\boldsymbol{\lambda}} \in \mathbb{Z}^d$, $t_{\boldsymbol{\lambda}} \in \mathbb{R}$ and $0 \leqslant \kappa_{\boldsymbol{\lambda}} < \kappa$ for $0 \leqslant \boldsymbol{\lambda} < q^{\kappa}\mathbf{1}$. If $d \geqslant 2$, then additionally let $\boldsymbol{r}_{\boldsymbol{\lambda}} \geqslant 0$ for all $\boldsymbol{\lambda}$.

Consider the sequence $a(\boldsymbol{n})$, $\boldsymbol{n} \in \mathbb{N}_0^d$, defined by the recursion

$$a(q^{\kappa}\boldsymbol{n} + \boldsymbol{\lambda}) = a(q^{\kappa_{\boldsymbol{\lambda}}}\boldsymbol{n} + \boldsymbol{r}_{\boldsymbol{\lambda}}) + t_{\boldsymbol{\lambda}} \quad \text{for} \quad 0 \leqslant \boldsymbol{\lambda} < q^{\kappa}\mathbf{1} \tag{10}$$

and for all integer vectors $\boldsymbol{n}$ such that the arguments on both sides are non-negative. Furthermore, initial values $a(\boldsymbol{n})$ for $\boldsymbol{n} \in \mathcal{I}$ have to be given for a suitable finite set $\mathcal{I} \subset \mathbb{N}_0^d$.

It must be ensured that the recursion (10) does not lead to conflicts and that the set of $\mathcal{I}$ is appropriate. Additionally, we require that $\mathcal{I}$ is minimal (with respect to inclusion). In that case, we say that the recursion is well-posed.

In Section 6, we construct a subsequential, complete, deterministic transducer $\mathcal{T}$ (also when the recursion is not well-posed) reading the q-ary joint expansion of integer vectors without leading zeros. We will define a distinguished subset of its states, called *simple states*. Furthermore, disjoint classes $F_1, \ldots, F_K$ of integer vectors will be defined.

Theorem 4. *The recursion* (10) *is well-posed if and only if*

1. *for each cycle consisting of simple states with transitions with zero input label, the sum of its output transitions vanishes and*

2. *the set $\mathcal{I}$ consists of one representative of each F_j, $1 \leqslant j \leqslant K$.*

In that case, the sum of the output of $\mathcal{T}$ is the sequence a, i.e., $\mathcal{T}(\boldsymbol{n}) = a(\boldsymbol{n})$ for all $\boldsymbol{n} \geqslant 0$.

The proof of this theorem is in Section 6. Combining this result with Theorem 1 yields an asymptotic analysis of the sequence $a(\boldsymbol{n})$, as in Example 2.8. Moreover, this asymptotic analysis can be performed algorithmically in Sage for $d = 1$ (using the code submitted at http://trac.sagemath.org/17221). A combinatorial description of the sets F_i involving an auxiliary transducer is given in Remark 6.1.

Remark 2.6. For $d \geqslant 2$, and $\boldsymbol{r}_{\boldsymbol{\lambda}} \not\geqslant 0$, the sequence cannot be computed by a finite transducer: For every $j \geqslant 0$, there are non-zero integer vectors $\boldsymbol{n} \geqslant 0$, $\boldsymbol{n}' \geqslant 0$ with $\boldsymbol{n} \equiv \boldsymbol{n}' \pmod{q^j}$—i.e., a finite deterministic transducer cannot distinguish between $\boldsymbol{n}$ and $\boldsymbol{n}'$—such that the recursion (10) can be applied for the argument $q^{\kappa}\boldsymbol{n} + \boldsymbol{\lambda}$ but cannot be applied for $q^{\kappa}\boldsymbol{n}' + \boldsymbol{\lambda}$.

This problem does not arise in the case of dimension $d = 1$: if the end of the input is not yet reached (this is something the transducer knows), there is a guaranteed forthcoming digit $\geqslant 1$ (instead of $\neq 0$ in the higher dimensional case). This information is enough to decide whether the recursion can be used.

Remark 2.7. Suppose that the given sequence is defined for $\boldsymbol{n} \geqslant \boldsymbol{n}_0$ for some constant $\boldsymbol{n}_0$. Then the sequence $b(\boldsymbol{n}) = a(\boldsymbol{n} + \boldsymbol{n}_0)$ fulfills (10) with $\kappa_{\boldsymbol{\lambda}}$, $\boldsymbol{r}_{\boldsymbol{\lambda}}$ and $t_{\boldsymbol{\lambda}}$ replaced by $\kappa_{\boldsymbol{\mu}}$, $q^{\kappa_{\boldsymbol{\mu}}} \boldsymbol{s} + \boldsymbol{r}_{\boldsymbol{\mu}} - \boldsymbol{n}_0$ and $t_{\boldsymbol{\mu}}$, respectively, where $\boldsymbol{n}_0 + \boldsymbol{\lambda} = q^{\kappa} \boldsymbol{s} + \boldsymbol{\mu}$ for $0 \leqslant \boldsymbol{\mu} < q^{\kappa} \mathbf{1}$. Then Theorem 4 can be applied.

Example 2.8. Consider the abelian complexity function $\rho(n)$ of the paperfolding sequence. The paperfolding sequence is obtained by repeatedly folding a strip of paper in half in the same direction. Then we open the strip and encode a right turn by 1 and a left turn by 0. The abelian complexity function $\rho(n)$ gives the number of abelian equivalence classes of subwords of length n of the paperfolding sequence. Two subwords of length n are equivalent if they are permutations of each other. In [28], the authors prove that this sequence satisfies the recursion

$$\rho(4n) = \rho(2n),$$
$$\rho(4n + 2) = \rho(2n + 1) + 1,$$
$$\rho(16n + 1) = \rho(8n + 1),$$
$$\rho(16n + 3) = \rho(2n + 1) + 2,$$
$$\rho(16n + 5) = \rho(4n + 1) + 2,$$
$$\rho(16n + 7) = \rho(2n + 1) + 2,$$
$$\rho(16n + 9) = \rho(2n + 1) + 2,$$
$$\rho(16n + 11) = \rho(4n + 3) + 2,$$
$$\rho(16n + 13) = \rho(2n + 1) + 2,$$
$$\rho(16n + 15) = \rho(2n + 2) + 1$$

with $\rho(1) = 2$ and $\rho(0) = 0$. The constructed transducer is shown in Figure 5.

For simplicity, we do not state the final output labels in this figure. The expected value and the variance are

$$\mathbb{E}(\rho(n)) = \frac{8}{13} \log_2 N + \Psi_1(\log_2 N) + \mathcal{O}(N^{-\xi} \log N),$$
$$\mathbb{V}(\rho(n)) = \frac{432}{2197} \log_2 N - \Psi_1^2(\log_2 N) + \Psi_2(\log_2 N) + \mathcal{O}(N^{-\xi} \log^2 N)$$

with $0 < \xi < 0.5604267891$, as the second largest eigenvalues of the transition matrix are $-0.7718445063 \pm 1.1151425080\, i$. The sequence $\rho(n)$ is asymptotically normally distributed. The functions $\Psi_1(x)$ and $\Psi_2(x)$ are 1-periodic and continuous. The reset sequence of the transducer is (00001) (reading from right to left). The function $\Psi_1(x)$ is nowhere differentiable and its Fourier series converges absolutely and uniformly. The first 24 Fourier coefficients of $\Psi_1(x)$ are listed in Table 1. In Figure 6, the trigonometric polynomial formed with the first 2550 Fourier coefficients is compared with the empirical values of the function $\Psi_1(x)$ (see (8)).
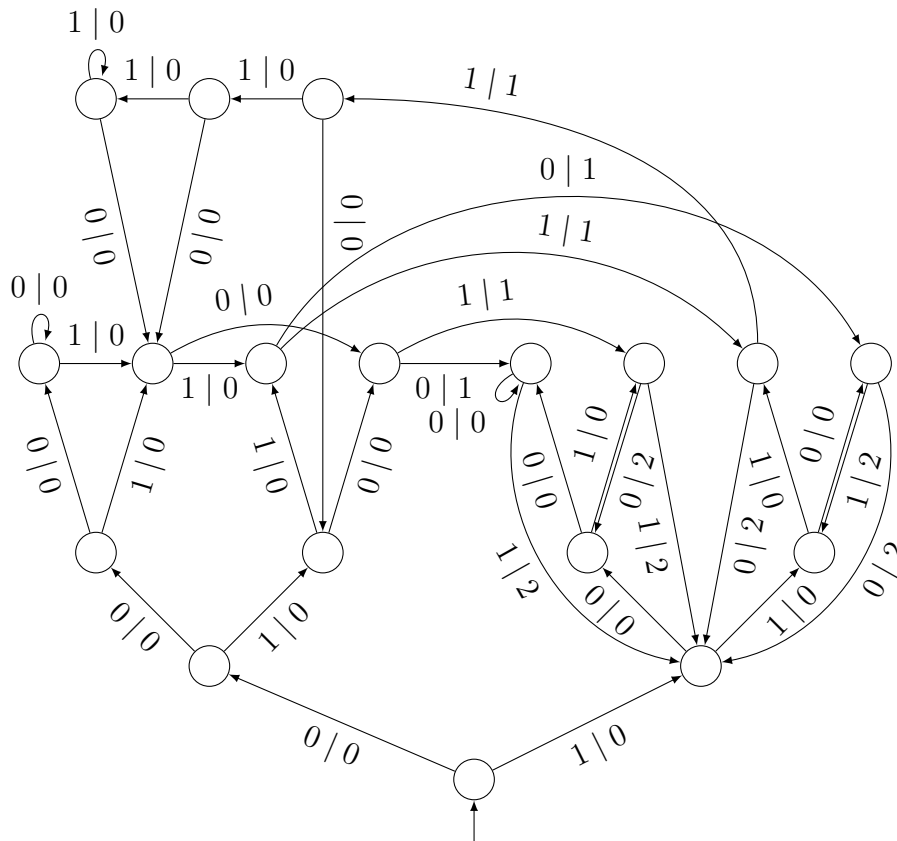
Figure 5: Transducer computing the abelian complexity function $\rho(n)$ of the paperfolding sequence. For simplicity, the final output labels are omitted.
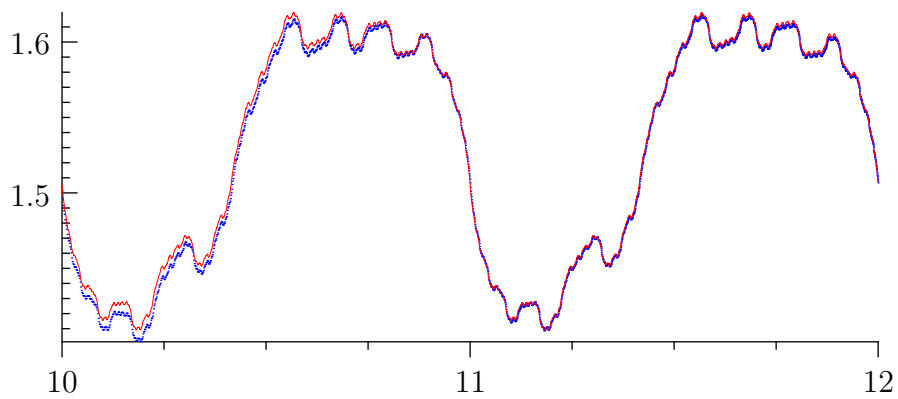

Figure 6: Partial Fourier series compared with the empirical values of $\Psi_1(x)$ of the abelian complexity function of the paperfolding sequence.

l	c_l	l	c_l
0	1.5308151288	12	$-0.0002297481 + 0.0009687657\,i$
1	$-0.0162585750 + 0.0478637218\,i$	13	$0.0006425378 + 0.0006516706\,i$
2	$0.0054521982 + 0.0075023586\,i$	14	$0.0000413217 - 0.0003867709\,i$
3	$-0.0028294724 + 0.0086495903\,i$	15	$-0.0005632948 - 0.0001843541\,i$
4	$0.0036818110 + 0.0021908312\,i$	16	$0.0009051717 - 0.0000476354\,i$
5	$-0.0028244495 + 0.0014519078\,i$	17	$-0.0004621780 - 0.0000594551\,i$
6	$-0.0008962222 + 0.0030512180\,i$	18	$-0.0000127264 - 0.0003100798\,i$
7	$0.0015033904 + 0.0013217107\,i$	19	$0.0004112716 + 0.0001954204\,i$
8	$-0.0006766166 - 0.0015392566\,i$	20	$-0.0000011706 + 0.0004183253\,i$
9	$0.0016074870 - 0.0000503663\,i$	21	$-0.0001027596 + 0.0004091624\,i$
10	$-0.0006908394 + 0.0018753575\,i$	22	$-0.0004725451 + 0.0004237489\,i$
11	$-0.0008974336 + 0.0007658455\,i$	23	$-0.0000596181 + 0.0002323317\,i$

Table 1: First 24 Fourier coefficients of the abelian complexity function $\rho(n)$ of the paperfolding sequence.

3   Asymptotic Distribution — Proof of Theorem 1

This section contains some lemmas which will together imply Theorem 1. Our plan is as follows: First, we give auxiliary lemmas about the eigenvalues and eigenvectors of the transition matrix M in Section 3.1. Section 3.2 contains an asymptotic formula for the characteristic function of the random variable $\mathcal{T}(\boldsymbol{n})$. We use this characteristic function to give formulas for the expected value and the variance in Section 3.3, and prove the continuity of the periodic fluctuations in Section 3.4. Finally, we prove the central limit theorem in Section 3.5.

We use the notation $(\boldsymbol{\varepsilon}_L \dots \boldsymbol{\varepsilon}_0)_q$ for the standard q-ary joint digit representation of an integer vector with $\boldsymbol{\varepsilon}_L \neq 0$. For a real number in the interval $[0, q)$, we write $(\varepsilon_0 \boldsymbol{.}\, \varepsilon_1 \dots)_q$ for the q-ary digit representation choosing the representation ending on 0^ω in the case of ambiguity. Furthermore, we use Iverson's notation [19]: $[expression]$ is 1 if $expression$ is true and 0 otherwise. All $\mathcal{O}$-constants depend only on q, d and the number of states.

3.1   Transition Matrix and its Eigenvectors

This section contains the proofs of some results on the eigenvalues, eigenvectors and eigenprojections of the transition matrix M.

For the proof of Theorem 1, we use the following lemma which describes the eigenvalues of a matrix in a similar way as the Perron–Frobenius theorem (cf. [14]).

Lemma 3.1. *Let M be a matrix with complex entries whose underlying directed graph is p-periodic and strongly connected. Then the set of non-zero eigenvalues of M can be partitioned into disjoint sets of cardinality p where each set is invariant under multiplication by $e^{2\pi i/p}$ and all eigenvalues in one set have the same algebraic multiplicities.*

Proof. Since the underlying directed graph of $M \in \mathbb{C}^{n \times n}$ is a strongly connected, p-periodic graph, we can write M as

$$M = \begin{pmatrix} 0 & A_2 & 0 & \cdots & 0 \\ \vdots & \ddots & A_3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & A_p \\ A_1 & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

with block matrices A_i by reordering the vertices. Then $M - xI$ is the product of the matrices

$$\begin{pmatrix} -xI & 0 & \cdots & \cdots & & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & -xI & & 0 \\ A_1 & \frac{1}{x}\prod_{j=1}^{2} A_j & \cdots & \frac{1}{x^{p-2}}\prod_{j=1}^{p-1} A_j & \frac{1}{x^{p-1}}\prod_{j=1}^{p} A_j - xI \end{pmatrix}$$

and

$$\begin{pmatrix} I & -\frac{1}{x}A_2 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & I & -\frac{1}{x}A_p \\ 0 & \cdots & \cdots & 0 & I \end{pmatrix} .$$

Let $h(x)$ be the characteristic polynomial of $\prod_{j=1}^{p} A_j \in \mathbb{C}^{m \times m}$. Thus the characteristic polynomial of M is $x^{n-m-(p-1)m}h(x^p)$. Therefore, the eigenvalues of M are either 0 or any p-th root of a non-zero eigenvalue of $\prod_{j=1}^{p} A_j$. $\qquad \square$

With this lemma, we can prove Lemma 2.3 about the eigenvalues of the matrix M:

Proof of Lemma 2.3. First, consider the case $t = 0$. By construction, q^d is an eigenvalue with right eigenvector $\mathbf{1}$ of M. As $\|M\|_\infty \leqslant q^d$, where $\|\cdot\|_\infty$ denotes the row sum norm, q^d is a dominant eigenvalue.

Consider the strongly connected components of the underlying graph of $\mathcal{T}$. Each final strongly connected component C_j induces a final transducer $\mathcal{T}_j$ which is strongly connected, complete, deterministic and p_j-periodic. Thus, the adjacency matrix at $t = 0$ of this final transducer has a dominant eigenvalue q^d with right eigenvector $\mathbf{1}$. By the Perron–Frobenius theorem (cf. [14, Theorem 8.8.1]), all dominant eigenvalues of this final transducer are $\{q^d e^{2\pi i l/p} \mid l \in \mathcal{P} \text{ with } p \mid lp_j\}$, each with algebraic and geometric multiplicity one.

A non-final strongly connected component induces a transducer $\mathcal{S}$ with the adjacency matrix S. This transducer is not complete. Let $\mathcal{S}^+$ be the complete transducer where loops are added to states of $\mathcal{S}$ where necessary. The adjacency matrix of $\mathcal{S}^+$ is S^+. Since

$\mathcal{S}^+$ is complete, deterministic and strongly connected, $\rho(\mathcal{S}^+) = q^d$. As $S \leqslant S^+$ but $S \neq S^+$, Theorem 8.8.1 in [14] implies $\rho(S) < \rho(S^+) = q^d$.

Thus, the dominant eigenvalues are $q^d e^{2\pi i l/p}$ with an $l \in \mathcal{P}$ such that there exists a $j \in \{1, \ldots, c\}$ with $p \mid lp_j$. We determine the geometric multiplicities of these dominant eigenvalues of M in Lemma 3.2.

Now, fix a final strongly connected component C_j and some $l \in \mathcal{P}$ with $p \mid lp_j$. In a small neighborhood of $t = 0$, let $\mu_{lj}(t)$ be the eigenvalue of the submatrix of M corresponding to the complete transducer $\mathcal{T}_j$ with $\mu_{lj}(0) = q^d e^{2\pi i l/p}$. Because of Lemma 3.1 applied to the final component C_j separately, we have $\mu_{lj}(t) = e^{2\pi i l/p} \mu_j(t)$ where $\mu_j(t)$ is defined to be $\mu_{0j}(t)$.

All other moduli of eigenvalues of M are less than $\min_{l,j} |\mu_{lj}(t)|$ because of the continuity of eigenvalues.

We prove the differentiability of the eigenvalues in Lemma 3.2. $\qquad\square$

Lemma 3.2. *Let $\mu_j(t) \exp(\frac{2\pi i l}{p})$ be a dominant eigenvalue of the matrix M. There exists a corresponding left eigenvector of M with zero entries except in coordinates corresponding to the final component C_j.*

At $t = 0$, the algebraic and geometric multiplicities of $q^d \exp(\frac{2\pi i l}{p})$ coincide.

Furthermore the eigenvalues and the eigenprojection corresponding to the eigenvalues $\mu_j \exp(\frac{2\pi i l}{p})$ are analytic at $t = 0$.

Proof. Let $q^d \exp(\frac{2\pi i l}{p})$ be a dominant eigenvalue of M. Its algebraic multiplicity at $t = 0$ is $|\{j : p \mid lp_j\}|$. We construct exactly one left eigenvector in the neighborhood of $t = 0$ for each final component C_j with $p \mid lp_j$: Let $\mathcal{T}_j$ be the induced transducer of the final component C_j. Let $\tilde{\boldsymbol{v}}^\top(t)$ be a left eigenvector of the adjacency matrix of $\mathcal{T}_j$ corresponding to the eigenvalue $\mu_j(t) \exp(\frac{2\pi i l}{p})$. As the algebraic multiplicity is 1 in this final component, the choice of $\tilde{\boldsymbol{v}}^\top(t)$ is unique up to multiplication with a scalar function in t. Then, we construct the left eigenvector $\boldsymbol{v}^\top(t)$ by padding $\tilde{\boldsymbol{v}}^\top(t)$ with zeros.

These left eigenvectors are linearly independent because of the block structure induced by the final components. Thus the geometric and the algebraic multiplicities of $q^d \exp(\frac{2\pi i l}{p})$ coincide.

Furthermore, $\mu_j(t) \exp(\frac{2\pi i l}{p})$ is a simple eigenvalue of the adjacency matrix of $\mathcal{T}_j$. Therefore, [25, Chapter II] implies the differentiability of the eigenvalues and eigenprojections. $\qquad\square$

From now on we use the convention that the eigenspace corresponding to $\mu_j(t) \exp(\frac{2\pi i l}{p})$ is the null space if $\mu_j(t) \exp(\frac{2\pi i l}{p})$ is not an eigenvalue. Then its eigenprojection is the constant null function.

Definition 3.3. Let $\boldsymbol{w}_{lj}^\top(t)$ be the eigenprojection of $\boldsymbol{e}_1^\top$ onto the left eigenspace corresponding to the possible eigenvalue $\mu_j(t) \exp(\frac{2\pi i l}{p})$. The vector $\boldsymbol{w}_{lj}^\top(t)$ is thus a null vector or a left eigenvector of M corresponding to the eigenvalue $\mu_j(t) \exp(\frac{2\pi i l}{p})$.

Define
$$\boldsymbol{w}^\top(t) = \boldsymbol{e}_1^\top - \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} \boldsymbol{w}_{lj}^\top(t).$$

As an abbreviation, we write $\boldsymbol{w}_{lj}^\top$, $\boldsymbol{w}^\top$, $\boldsymbol{w}_{lj}'^\top$ and $\boldsymbol{w}'^\top$ for these projections and their derivatives at $t = 0$.

Remark 3.4. If there are only dominant eigenvalues, then $\boldsymbol{w}^\top(t) = 0$. This will imply that there is no error term in the asymptotic expansion of the expected value and the variance. This occurs in the case of the sum of digits of the standard q-ary digit representation and other completely q-additive functions because the transducer has only one state.

Lemma 3.5. *In a fixed neighborhood of $t = 0$, let $\xi > 0$ be as defined in* (5), *i.e., all non-dominant eigenvalues have modulus less than $q^{d-\xi}$. Then*

$$\left\| \frac{d^k}{dt^k} \boldsymbol{w}^\top(t) M^m \right\| = \mathcal{O}(c_k^{(1)} q^{(d-\xi)(m-k)} m^k)$$

for $m, k \geqslant 0$ and a constant $c_k^{(1)}$.

Proof. Let P be the matrix such that $x^\top \mapsto x^\top P$ is the sum of the eigenprojections onto the left eigenspaces corresponding to $\mu_j \exp(\frac{2\pi i l}{p})$ for $j = 1, \ldots, c$ and $l \in \mathcal{P}$. Then $\boldsymbol{w}^\top = \boldsymbol{e}_1^\top(I - P)$ and

$$\boldsymbol{w}^\top M^m = \boldsymbol{e}_1^\top((I - P)M)^m.$$

As the spectral radius of $(I - P)M$ is less than $q^{d-\xi}$, we obtain the stated estimates. □

With $\boldsymbol{w}_l^\top$ defined in Section 2.3, we have

$$\boldsymbol{w}_l^\top(t) = \sum_{j=1}^{c} \boldsymbol{w}_{lj}^\top(t). \tag{11}$$

Note that left and right eigenvectors corresponding to different eigenvalues annihilate each other. Because of the block structure of the eigenvectors in Lemma 3.2 and because $\mathbf{1}$ is a right eigenvector to q^d, we have

$$[l = 0]\lambda_j = \boldsymbol{w}_{lj}^\top \mathbf{1} \tag{12}$$

where λ_j is defined in Section 2.3. Furthermore, $\boldsymbol{w}^\top \mathbf{1} = 0$ and

$$\sum_{j=1}^{c} \lambda_j = \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} \boldsymbol{w}_{lj}^\top \mathbf{1} + \boldsymbol{w}^\top \mathbf{1} = \boldsymbol{e}_1^\top \mathbf{1} = 1.$$

Denote by $\boldsymbol{\delta}$ the vector whose s-th component is the sum of the outputs of all transitions leaving the state s. By the definition of the transition matrix $M(t)$, $\boldsymbol{\delta}$ can be expressed as

$$i\boldsymbol{\delta} = \frac{d}{dt} M(t)\mathbf{1} \Big|_{t=0}. \tag{13}$$

We now establish a relation between $\boldsymbol{\delta}$, the left eigenvector $\boldsymbol{w}_l^\top$ and its derivative at $t = 0$. By definition of the left eigenvectors $\boldsymbol{w}_{lj}^\top(t)$ and (11),

$$\boldsymbol{w}_l^\top(t)M\mathbf{1} = \sum_{j=1}^c \mu_j(t)\exp\left(\frac{2\pi il}{p}\right)\boldsymbol{w}_{lj}^\top(t)\mathbf{1}.$$

Differentiation, (12), (5) and (11) yield

$$\boldsymbol{w}_l^\top\boldsymbol{\delta} = [l = 0]e_{\mathcal{T}}q^d - q^d\left(e^{\frac{2\pi il}{p}} - 1\right)i\boldsymbol{w}_l'^\top\mathbf{1}. \tag{14}$$

To establish the interpretation of $\boldsymbol{w}_0^\top$ given at the end of Section 2.3, we consider

$$\hat{\boldsymbol{w}}_k^\top := \lim_{m\to\infty} \boldsymbol{e}_1^\top M^{mp+k}q^{-d(mp+k)},$$

the stationary distribution on the state space of all states of the transducer under the assumption that the input length is congruent to k modulo p. Using (11) and Lemma 3.5 yields

$$\begin{aligned}
\hat{\boldsymbol{w}}_k^\top &= \lim_{m\to\infty}\left(\sum_{l\in\mathcal{P}}\boldsymbol{w}_l^\top + \boldsymbol{w}^\top\right)M^{mp+k}q^{-d(mp+k)} \\
&= \lim_{m\to\infty}\sum_{l\in\mathcal{P}}\exp\left(\frac{2\pi ilk}{p}\right)\boldsymbol{w}_l^\top + \mathcal{O}(q^{-\xi(mp+k)}) \\
&= \sum_{l\in\mathcal{P}}\exp\left(\frac{2\pi ilk}{p}\right)\boldsymbol{w}_l^\top.
\end{aligned}$$

Summation leads to $\frac{1}{p}\sum_{k=0}^{p-1}\hat{\boldsymbol{w}}_k^\top = \boldsymbol{w}_0^\top$. Thus, λ_j is the hitting probability of the final component C_j when starting in the initial state. As every state is accessible from the initial state, λ_j is positive.

Finally, for $l = 0$, (14) reads $q^{-d}\boldsymbol{w}_0^\top\boldsymbol{\delta} = e_{\mathcal{T}}$, which can be interpreted as the steady state analysis of the expectation: the probability distribution $\boldsymbol{w}_0^\top$ is multiplied with the expected output $q^{-d}\boldsymbol{\delta}$.

3.2 Characteristic function

To obtain a central limit law in Section 3.5, we compute an asymptotic formula for the characteristic function in this section.

The next lemma can be proved by induction on L. It is a generalization of Lemma 3 in [21].

Lemma 3.6. *Let A_ε, $\varepsilon = 0,\ldots,q-1$ be matrices in $\mathbb{C}^{n\times n}$, $H_\varepsilon\colon \mathbb{N}_0 \to \mathbb{C}^{n\times n}$ be known functions with $H_0(0) = 0$. Let $G\colon \mathbb{N}_0 \to \mathbb{C}^{n\times n}$ be a function which satisfies the recurrence relation*

$$G(qN + \varepsilon) = A_\varepsilon G(N) + H_\varepsilon(N)$$

for $N \geqslant 0$, $\varepsilon \in \{0, \ldots, q-1\}$ and $G(0) = 0$. Then

$$G\left((\varepsilon_L \ldots \varepsilon_0)_q\right) = \sum_{l=0}^{L} \left(\prod_{i=0}^{l-1} A_{\varepsilon_i}\right) H_{\varepsilon_l}\left((\varepsilon_L \ldots \varepsilon_{l+1})_q\right).$$

The solution of this recursion finally leads to an asymptotic formula for the characteristic function.

We choose the branch $-\pi + \frac{\pi}{p} < \arg z \leqslant \pi + \frac{\pi}{p}$ of the complex logarithm. After setting $t = 0$, we use only the logarithm of complex numbers for which our branch coincides the principal branch $-\pi < \arg z \leqslant \pi$.

Lemma 3.7. *The characteristic function of the random variable $\mathcal{T}(\boldsymbol{n})$ is*

$$\mathbb{E}(\exp(it\mathcal{T}(\boldsymbol{n}))) =$$

$$\frac{1}{N^d} \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} \mu_j(t)^{\log_q N} \exp\left(\frac{2\pi i l \log_q N}{p}\right) \Psi_{lj}(\log_q N, t) + R(N, t)$$

with functions $\Psi_{lj}(x, t)$ (defined in (24)), which are arbitrarily often differentiable in t and 1-periodic in x, and an error term $R(N, t)$. This error term $R(N, t)$ is arbitrarily often differentiable, too, and satisfies $\frac{d^k}{dt^k} R(N, t) = \mathcal{O}(c_k^{(2)} N^{-\xi} \log^k N)$, for $k \geqslant 0$, a constant $c_k^{(2)}$ and the constant $\xi > 0$ defined in Section 2.3, in a neighborhood of $t = 0$. At $t = 0$, we have $R(N, 0) = 0$.

Proof. For a transducer $\mathcal{T}$, consider the characteristic function

$$F(N) = \mathbb{E}(\exp(it\mathcal{T}(n))) = \frac{1}{N^d} \sum_{\boldsymbol{n} \in \Omega_N} e^{it\mathcal{T}(\boldsymbol{n})} \tag{15}$$

of the discrete random variable $\mathcal{T}(\boldsymbol{n})$.

Then the summands in (15) can be expressed as a matrix product

$$e^{it\mathcal{T}(\boldsymbol{n})} = \boldsymbol{e}_1^\top \prod_{l=0}^{L} M_{\varepsilon_l} \boldsymbol{u}$$

where $(\boldsymbol{\varepsilon}_L \ldots \boldsymbol{\varepsilon}_0)_q$ is the standard q-ary joint digit representation of $\boldsymbol{n}$ with $\boldsymbol{\varepsilon}_L \neq 0$ and the vector $\boldsymbol{u}$ has entries $e^{itb(s)}$ where $b(s)$ is the final output of the state s. Again, the vector $\boldsymbol{e}_1$ is the indicator vector of the initial state.

Let

$$g(\boldsymbol{n}) = \prod_{l=0}^{L} M_{\varepsilon_l}$$

and

$$G(N) = \sum_{\boldsymbol{n} \in \Omega_N} g(\boldsymbol{n}),$$

hence

$$F(N) = \frac{1}{N^d} \boldsymbol{e}_1^\top G(N) \boldsymbol{u}. \tag{16}$$

The function $g(\boldsymbol{n})$ satisfies the recursion

$$g(q\boldsymbol{n} + \boldsymbol{\varepsilon}) = M_{\boldsymbol{\varepsilon}} g(\boldsymbol{n}) \tag{17}$$

for $\boldsymbol{\varepsilon} \in \{0, 1, \ldots, q-1\}^d$, $\boldsymbol{n} \geqslant 0$ with $q\boldsymbol{n} + \boldsymbol{\varepsilon} \neq 0$.

We define further functions

$$G_C(N) = \sum_{\substack{0 \leqslant n_i < N \\ i \notin C}} \sum_{\substack{n_i = N \\ i \in C}} g(\boldsymbol{n}) \tag{18}$$

where the coordinates $n_1, \ldots, n_d$ of $\boldsymbol{n}$ with indices in the set $C \subseteq \{1, \ldots, d\}$ are fixed to N. This yields $G(N) = G_\emptyset(N)$. Furthermore, we define the matrices

$$M_{C,D}^{\varepsilon} = \sum_{\substack{\beta_i = 0 \\ i \notin C \cup D}}^{q-1} \sum_{\substack{\beta_i = 0 \\ i \in D}}^{\varepsilon-1} \sum_{\substack{\beta_i = \varepsilon \\ i \in C}} M_{\boldsymbol{\beta}} \tag{19}$$

for disjoint sets C, $D \subseteq \{1, \ldots, d\}$ and $\varepsilon \in \{0, 1, \ldots, q-1\}$. In this definition, we restrict the i-th coordinate β_i of $\boldsymbol{\beta}$ to be ε or less than ε if $i \in C$ or $i \in D$, respectively. Otherwise, the i-th coordinate can be arbitrary. Then, $M = M_{\emptyset,\emptyset}^{\varepsilon}$ holds independently of ε.

Then, (17) yields the following recursions for $G_C(N)$, $\varepsilon = 0, \ldots, q-1$, $N \geqslant 0$ and $C \neq \{1, \ldots, d\}$:

$$\begin{aligned}
G_C(qN + \varepsilon) &= \sum_{\substack{\beta_i = 0 \\ i \notin C}}^{q-1} \sum_{\substack{\beta_i = \varepsilon \\ i \in C}} \sum_{\substack{0 \leqslant qm_i + \beta_i < qN + \varepsilon \\ i \notin C}} \sum_{\substack{qm_i + \beta_i = qN + \varepsilon \\ i \in C}} g(q\boldsymbol{m} + \boldsymbol{\beta}) \\
&= [C = \emptyset \wedge qN + \varepsilon \neq 0](I - M_0) \\
&\quad + \sum_{\substack{\beta_i = 0 \\ i \notin C}}^{q-1} \sum_{\substack{\beta_i = \varepsilon \\ i \in C}} M_{\boldsymbol{\beta}} \sum_{\substack{0 \leqslant m_i < N + \frac{\varepsilon - \beta_i}{q} \\ i \notin C}} \sum_{\substack{m_i = N \\ i \in C}} g(\boldsymbol{m}) \\
&= [C = \emptyset \wedge qN + \varepsilon \neq 0](I - M_0) + \sum_{D \subseteq C^c} M_{C,D}^{\varepsilon} G_{C \cup D}(N).
\end{aligned} \tag{20}$$

This recursion for G_C only depends on $G_{C'}$ for $C' \supsetneq C$. As

$$G_{\{1,\ldots,d\}}(N) = g(N\boldsymbol{1}),$$

we can recursively determine G_C using Lemma 3.6. In particular, for $G(N)$, this yields the recursion formula

$$G(qN + \varepsilon) = MG(N) + H_\varepsilon(N) \tag{21}$$

for $N \geqslant 0$, $\varepsilon \in \{0, \ldots q - 1\}$ where H_ε are known functions with

$$H_\varepsilon(N) = [qN + \varepsilon \neq 0](I - M_0) + \sum_{\emptyset \neq D \subseteq \{1, \ldots, d\}} M_{\emptyset, D}^\varepsilon G_D(N). \tag{22}$$

Thus by Lemma 3.6, we get

$$G((\varepsilon_L \ldots \varepsilon_0)_q) = \sum_{m=0}^{L} M^m H_{\varepsilon_m}((\varepsilon_L \ldots \varepsilon_{m+1})_q). \tag{23}$$

By construction, $\|M_{\boldsymbol{\varepsilon}}\|_\infty = 1$ for every $\boldsymbol{\varepsilon} \in \{0, \ldots, q-1\}^d$. We conclude that $\|M_{C,D}^\varepsilon\|_\infty \leqslant q^{d-|C|-|D|} \varepsilon^{|D|}$. By the definition of $G_C(N)$, the growth rates of the functions $G_C(N)$ and $H_\varepsilon(N)$ are $\|G_C(N)\|_\infty = \mathcal{O}(N^{d-|C|})$ and $\|H_\varepsilon(N)\|_\infty = \mathcal{O}(N^{d-1})$, respectively. For $k \geqslant 0$, the k-th derivative of $H_\varepsilon(N)$ at $t = 0$ can be bounded by $\mathcal{O}(c_k^{(3)} N^{d-1} \log^k N)$ for a constant $c_k^{(3)}$.

We define

$$R(N, t) = \frac{1}{N^d} \sum_{m=0}^{L} \boldsymbol{w}^\top M^m H_{\varepsilon_m}((\varepsilon_L \ldots \varepsilon_{m+1})_q) \boldsymbol{u},$$

which constitutes an explicit expression for the error term contributed by the non-dominant eigenvalues. By Lemma 3.5, its derivatives satisfy

$$\frac{d^k}{dt^k} R(N, t) = \mathcal{O}(c_k^{(2)} N^{-\xi} \log^k N)$$

for $k \geqslant 0$. Because $\boldsymbol{u}(0) = \boldsymbol{1}$ and left and right eigenvectors corresponding to different eigenvalues annihilate each other, we have $R(N, 0) = 0$.

By (16), (23) and $\boldsymbol{e}_1^\top = \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} \boldsymbol{w}_{lj}^\top + \boldsymbol{w}^\top$,

$$\begin{aligned}
F(N) = {} & \frac{1}{N^d} \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} \mu_j^L \exp\left(\frac{2\pi i l L}{p}\right) \\
& \cdot \sum_{m=0}^{L} \mu_j^{m-L} \exp\left(\frac{2\pi i l(m-L)}{p}\right) \boldsymbol{w}_{lj}^\top H_{\varepsilon_m}((\varepsilon_L \ldots \varepsilon_{m+1})_q) \boldsymbol{u} \\
& + R(N, t) \\
= {} & \frac{1}{N^d} \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} \mu_j^{\log_q N} \exp\left(\frac{2\pi i l \log_q N}{p}\right) \Psi_{lj}(\log_q N, t) + R(N, t)
\end{aligned}$$

with

$$\begin{aligned}
\Psi_{lj}(x, t) = {} & \mu_j(t)^{-\{x\}} \exp\left(-\frac{2\pi i l \{x\}}{p}\right) \\
& \cdot \sum_{m=0}^{\infty} \mu_j(t)^{-m} \exp\left(-\frac{2\pi i l m}{p}\right) \boldsymbol{w}_{lj}^\top H_{x_m}((x_0 \ldots x_{m-1})_q) \boldsymbol{u} \quad (24)
\end{aligned}$$

and $q^{\{x\}} = (x_0 \bullet x_1 \ldots)_q$, choosing the representation ending on 0^ω in the case of ambiguity.

The functions $\Psi_{lj}(x,t)$ are periodic in x with period 1 and well defined for all $x \in \mathbb{R}$ since they are dominated by geometric series. Furthermore, they are arbitrarily often differentiable in t. $\qquad\square$

3.3 Moments

In this section we give the moments of the output sum $\mathcal{T}(\boldsymbol{n})$.

Lemma 3.8. *The expected value and the variance of $\mathcal{T}(n)$ are as stated in Theorem 1 with constants given in (5) and periodic functions given in Lemma 3.9 and (28).*

Proof. The derivative of $\mathbb{E}(\exp(it\mathcal{T}(\boldsymbol{n})))$ with respect to t at $t = 0$ gives the expected value of the sum of the output of the transducer

$$\mathbb{E}(\mathcal{T}(n)) = \frac{1}{N^d} \sum_{\boldsymbol{n} \in \Omega_N} \mathcal{T}(n) = \Psi_0(\log_q N)\log_q N + \Psi_1(\log_q N) + \mathcal{O}(N^{-\xi}\log N)$$

with p-periodic functions

$$\Psi_0(x) = \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} a_j e^{\frac{2\pi i l x}{p}} \Psi_{lj}(x,0),$$

$$\Psi_1(x) = -i \sum_{l \in \mathcal{P}} \sum_{j=1}^{c} e^{\frac{2\pi i l x}{p}} \Psi_{lj}'(x,0) \tag{25}$$

and constants a_j defined in (5). Here, Ψ_{lj}' denotes the derivative with respect to t.

We now compute $\Psi_0(x)$ for some x with $q^{\{x\}} = (x_0 \bullet x_1 \ldots)_q$. To compute $H_\varepsilon(N)$, we use (21) and the definition of $G(N)$ to obtain

$$H_\varepsilon(N)\mathbf{1} = ((qN + \varepsilon)^d - (qN)^d)\mathbf{1} \tag{26}$$

for $t = 0$, because $\mathbf{1}$ is a right eigenvector of $M_{\boldsymbol{\varepsilon}}$ for every $\boldsymbol{\varepsilon}$. Together with (24), this results in

$$\Psi_{lj}(x,0) = q^{-d\{x\}} \exp\left(-\frac{2\pi i l\{x\}}{p}\right) \boldsymbol{w}_{lj}^\top \mathbf{1} D\left(q^d e^{\frac{2\pi i l}{p}}\right)$$

with

$$D(z) = \sum_{m=0}^{\infty} z^{-m}((x_0 \ldots x_m)_q^d - (x_0 \ldots x_{m-1}0)_q^d).$$

By (12), we have $\Psi_{lj}(x,0) = 0$ for $l \neq 0$.

To compute $D(q^d)$, observe that

$$D(q^d) = \sum_{m=0}^{\infty} \left((x_0 \bullet x_1 \ldots x_m)_q^d - (x_0 \bullet x_1 \ldots x_{m-1})_q^d\right) = \lim_{m \to \infty} (x_0 \bullet x_1 \ldots x_m)_q^d = q^{d\{x\}}$$

because $D(q^d)$ is a telescoping sum.

We conclude that

$$\Psi_{lj}(x,0) = \lambda_j[l=0] \tag{27}$$

and therefore

$$\Psi_0(x) = \sum_{j=1}^c a_j \lambda_j = e_{\mathcal{T}}$$

by (5). This completes the proof of the expectation as given in (3).

Using Lemma 3.7 and (27), the second derivative of $\mathbb{E}(\exp(it\mathcal{T}(n)))$ gives

$$\frac{1}{N^d} \sum_{\boldsymbol{n}\in\Omega_N} \mathcal{T}(n)^2 = \log_q^2 N \sum_{j=1}^c a_j^2 \lambda_j + v_{\mathcal{T}} \log_q N$$

$$- 2i \log_q N \sum_{l\in\mathcal{P}} \sum_{j=1}^c a_j \exp\left(\frac{2\pi i l \log_q N}{p}\right) \Psi_{lj}'(\log_q N, 0)$$

$$+ \Psi_2(\log_q N) + \mathcal{O}(N^{-\xi} \log^2 N)$$

with $v_{\mathcal{T}}$ given in (5) and

$$\Psi_2(x) = -\sum_{l\in\mathcal{P}} \sum_{j=1}^c e^{\frac{2\pi i l x}{p}} \Psi_{lj}''(x,0). \tag{28}$$

Here, Ψ_{lj}'' denotes the second derivative with respect to t. Thus, by (3), the variance is

$$\mathbb{V}(\mathcal{T}(n)) = \frac{1}{N^d} \sum_{\boldsymbol{n}\in\Omega_N} \mathcal{T}(n)^2 - \left(\frac{1}{N^d} \sum_{\boldsymbol{n}\in\Omega_N} \mathcal{T}(n)\right)^2$$

$$= \left(\sum_{j=1}^c a_j^2 \lambda_j - e_{\mathcal{T}}^2\right) \log_q^2 N$$

$$+ \left(v_{\mathcal{T}} - 2i \sum_{l\in\mathcal{P}} \sum_{j=1}^c a_j \exp\left(\frac{2\pi i l \log_q N}{p}\right) \Psi_{lj}'(\log_q N, 0)\right. \tag{29}$$

$$\left. - 2e_{\mathcal{T}} \Psi_1(\log_q N)\right) \log_q N$$

$$+ \Psi_2(\log_q N) - \Psi_1^2(\log_q N) + \mathcal{O}(N^{-\xi} \log^2 N).$$

By Jensen's inequality, the coefficient of $\log_q^2 N$ is zero if and only if all a_j are equal. If all a_j are equal, then the coefficient of $\log_q N$ in (29) simplifies by (25), too, and we obtain (4). $\qquad\square$

For the computation of the Fourier coefficients and the proof of the Hölder condition, we need an explicit expression for Ψ_1.

In analogy to the definition of G_C in (18), define

$$\boldsymbol{B}_C(N) = \sum_{\substack{0 \leqslant n_i < N \\ i \notin C}} \sum_{\substack{n_i = N \\ i \in C}} \boldsymbol{b}(\boldsymbol{n}) \tag{30}$$

for $C \subseteq \{1, \ldots, d\}$.

Lemma 3.9. *For $q^{\{x\}} = (x_0 \, . \, x_1 \ldots)_q$, the fluctuation $\Psi_1(x)$ can be expressed as*

$$\Psi_1(x) = -e_{\mathcal{T}}\{x\} - q^{-d\{x\}} \sum_{l \in \mathcal{P}} \sum_{m=0}^{\infty} q^{-dm} e^{\frac{2\pi i l}{p}(\lfloor x \rfloor - m)} f_l((x_0 \ldots x_m)_q) \tag{31}$$

with

$$\begin{aligned}
f_l(r) = {}& [l = 0] e_{\mathcal{T}} \Big(\lfloor \log_q r \rfloor (r^d - (q\lfloor rq^{-1} \rfloor)^d) + (q\lfloor rq^{-1} \rfloor)^d \Big) \\
& + i \boldsymbol{w}_l'^{\top} \mathbf{1} \Big(r^d - \exp\Big(\frac{2\pi i l}{p}\Big)(q\lfloor rq^{-1} \rfloor)^d \Big) \\
& - \boldsymbol{w}_l^{\top} \boldsymbol{B}_{\emptyset}(r) + q^d \exp\Big(\frac{2\pi i l}{p}\Big) \boldsymbol{w}_l^{\top} \boldsymbol{B}_{\emptyset}(\lfloor rq^{-1} \rfloor).
\end{aligned} \tag{32}$$

The estimate $f_l(r) = \mathcal{O}(r^{d-1} \log r)$ holds.

Proof. From (25), (24), (5), (27) and (12) and the absolute convergence of Ψ_{lj}, we obtain (31) with

$$f_l(r) = [l = 0] e_{\mathcal{T}} \lfloor \log_q r \rfloor (r^d - (q\lfloor rq^{-1} \rfloor)^d) + i \left. \frac{d}{dt} \boldsymbol{w}_l^{\top}(t) H_{r \bmod q}(\lfloor rq^{-1} \rfloor) \boldsymbol{u}(t) \right|_{t=0}.$$

From the combinatorial interpretation of $\boldsymbol{b}(\boldsymbol{n})$ and $g(\boldsymbol{n})\boldsymbol{u}(t)$, we obtain

$$i\boldsymbol{b}(\boldsymbol{n}) = \left. \frac{d}{dt} g(\boldsymbol{n})\boldsymbol{u}(t) \right|_{t=0}, \tag{33}$$

in analogy to (13). As the range of summation of G_C and $\boldsymbol{B}_C$ coincides, we immediately get

$$i\boldsymbol{B}_C(N) = \left. \frac{d}{dt} G_C(N)\boldsymbol{u}(t) \right|_{t=0}. \tag{34}$$

By (26) and by differentiating $H_\varepsilon(N)\boldsymbol{u}(t)$ using (21), (34) and (13),

$$\begin{aligned}
f_l(r) = {}& [l = 0] e_{\mathcal{T}} \lfloor \log_q r \rfloor (r^d - (q\lfloor rq^{-1} \rfloor)^d) + i \boldsymbol{w}_l'^{\top} \mathbf{1}\big(r^d - (q\lfloor rq^{-1} \rfloor)^d \big) \\
& - \boldsymbol{w}_l^{\top} \big(\boldsymbol{B}_{\emptyset}(r) - M\boldsymbol{B}_{\emptyset}(\lfloor rq^{-1} \rfloor) - \lfloor rq^{-1} \rfloor^d \boldsymbol{\delta} \big).
\end{aligned}$$

The fact that $\boldsymbol{w}_l^{\top}$ is a left eigenvector of M and (14) establish (32).

For the growth estimate of $f_l(r)$, we use the explicit definition of H_ε in (22), (34) and the trivial estimate $\|\boldsymbol{b}(\boldsymbol{n})\| = \mathcal{O}(\log \|\boldsymbol{n}\|)$. $\qquad \square$

To formulate $\mathcal{T}(\boldsymbol{n})$ as a q-regular sequence, we first define output vectors. The s-th entry of the vector $\boldsymbol{\delta}_\varepsilon$ is the output label of the transition from state s with input label ε. By (17), (33), and

$$\frac{d}{dt} M_\varepsilon \mathbf{1} \Big|_{t=0} = i \boldsymbol{\delta}_\varepsilon, \tag{35}$$

we have

$$\boldsymbol{b}(q\boldsymbol{n} + \boldsymbol{\varepsilon}) = M_\varepsilon \boldsymbol{b}(\boldsymbol{n}) + \boldsymbol{\delta}_\varepsilon. \tag{36}$$

Remark 3.10. We can use the matrices

$$V_\varepsilon = \begin{pmatrix} M_\varepsilon & \boldsymbol{\delta}_\varepsilon & [\varepsilon = 0]I \\ 0 & 1 & 0 \\ 0 & 0 & [\varepsilon = 0]I \end{pmatrix}$$

and $\boldsymbol{v}(\boldsymbol{n}) = (\boldsymbol{b}(\boldsymbol{n}), 1, [\boldsymbol{n} = 0](\boldsymbol{b}(0) - M_0\boldsymbol{b}(0) - \boldsymbol{\delta}_0))^\top$ in the definition of a q-regular sequence (2) to realize that the output sum of a transducer is q-regular. If $d > 1$, then this is a multidimensional q-regular sequence (cf. [1]).

3.4 Hölder Continuity

In this section, we prove the continuity of the fluctuations Ψ_1 and Ψ_2 as well as the Hölder continuity of Ψ_1. This will be used to establish the convergence of the Fourier series.

Lemma 3.11. *The functions $\Psi_1(x)$ and, if all a_j are equal, $\Psi_2(x)$ are continuous for $x \in \mathbb{R}$.*

Proof. First note that continuity of Ψ_1 for $x \in \mathbb{R}$ with $x = \log_q y$ where y has no finite q-ary expansion follows from the definitions (24) and (25). To prove it for $x = \log_q y$ with $0 \leqslant x < p$ where y has a finite q-ary expansion, observe that the two one-sided limits exist due to the definition. Next, we prove that they are the same. Consider the two integer sequences $N_k = yq^{pk}$ and $\tilde{N}_k = N_k - 1$ for k large enough such that N_k is an integer. For a real number z, we write $\{z\}_p = p\{z/p\}$ for the unique real number in the interval $[0, p)$ such that $z - \{z\}_p$ is an integer multiple of p.

This yields

$$\lim_{k \to \infty} \{\log_q N_k\}_p = \lim_{k \to \infty} \{\log_q y + pk\}_p = \{x\}_p = \lim_{z \to x^+} \{z\}_p,$$

$$\lim_{k \to \infty} \{\log_q \tilde{N}_k\}_p = \lim_{k \to \infty} \{\log_q N_k + \log_q(1 - N_k^{-1})\}_p$$
$$= \lim_{k \to \infty} \{x + \log_q(1 - N_k^{-1})\}_p = \lim_{z \to x^-} \{z\}_p.$$

If we insert the two sequences N_k and $\tilde{N}_k$ in

$$\sum_{\boldsymbol{n} \in \Omega_N} \mathcal{T}(\boldsymbol{n}) = e_{\mathcal{T}} N^d \log_q N + N^d \Psi_1(\log_q N) + \mathcal{O}(N^{d-\xi} \log N)$$

(cf. (3)) and take the difference, we get

$$\mathcal{O}(N_k^{d-1} \log N_k) = N_k^d \Psi_1(\log_q N_k) - \tilde{N}_k^d \Psi_1(\log_q \tilde{N}_k) + \mathcal{O}(N_k^{d-\xi} \log N_k).$$

Because $\Psi_1(x)$ is bounded by a geometric series by definition, we have

$$\Psi_1(\log_q N_k) - \Psi_1(\log_q \tilde{N}_k) = O(N_k^{-\xi} \log N_k)$$

and in particular

$$\lim_{k \to \infty} \Psi_1(\{\log_q N_k\}_p) = \lim_{k \to \infty} \Psi_1(\{\log_q \tilde{N}_k\}_p).$$

Therefore, Ψ_1 is continuous in x.

The continuity of $\Psi_2(x)$ at $x = \log_q(y)$ for y with infinite q-ary expansion again follows from the definition of Ψ_2. If all a_j are equal, the continuity of the fluctuation $-\Psi_1^2 + \Psi_2$ of the variance (4) follows as above, where $\log N_k$ has to be replaced by $\log^2 N_k$ in the error terms. Thus Ψ_2 is also continuous in this case. $\qquad\square$

Lemma 3.12. *The function Ψ_1 satisfies a Hölder condition of order α for all $\alpha \in (0, 1)$.*

Proof. Let $0 < \alpha < 1$ be any constant. We want to prove that there exists a positive constant C such that

$$|\Psi_1(y) - \Psi_1(x)| \leqslant C|y - x|^\alpha \tag{37}$$

holds for all $x, y \in \mathbb{R}$.

For $x = y$, the left-hand side of (37) is 0 and the inequality is obviously satisfied. From now on, assume that $x < y$. By the periodicity of Ψ_1, it is sufficient to prove (37) for $0 \leqslant x < p$.

First, we prove (37) for the case $0 \leqslant x < y$ and sufficiently small $y - x < 1$.

Fix such x and y and choose the integer k such that

$$q^{-k-1} \leqslant |q^y - q^x| < q^{-k}.$$

Note that the continuous differentiability of $z \mapsto q^z$ on the compact interval $[0, p+1]$ implies that $q^y - q^x = \mathcal{O}(|y - x|)$ and therefore

$$q^{-k} = \mathcal{O}(|y - x|). \tag{38}$$

We prove (37) in three steps.

Statement 3.13. *Let $a, b \in \mathbb{R}$ with $x \leqslant a < b \leqslant y$ and $\lfloor a \rfloor = \lfloor b \rfloor$ such that the first $k+1$ digits of the expansions*

$$q^{\{a\}} = (a_0 . a_1 \ldots)_q, \quad q^{\{b\}} = (b_0 . b_1 \ldots)_q$$

coincide, i.e., $a_i = b_i$ for $0 \leqslant i \leqslant k$. Then

$$|\Psi_1(b) - \Psi_1(a)| = \mathcal{O}(|y - x|^\alpha).$$

Proof. Lemma 3.9 yields

$$
\begin{aligned}
|\Psi_1(b) - \Psi_1(a)| &\leqslant |e_{\mathcal{T}}||\{b\} - \{a\}| \\
&\quad + q^{-d\{b\}} \sum_{l \in \mathcal{P}} \sum_{m \geqslant 0} q^{-dm} |f_l((b_0 \ldots b_m)_q) - f_l((a_0 \ldots a_m)_q)| \\
&\quad + |q^{-d\{b\}} - q^{-d\{a\}}| \sum_{l \in \mathcal{P}} \sum_{m \geqslant 0} q^{-dm} |f_l((a_0 \ldots a_m)_q)| \\
&\leqslant |e_{\mathcal{T}}||\{b\} - \{a\}| \\
&\quad + \sum_{l \in \mathcal{P}} \sum_{m > k} q^{-dm} (|f_l((b_0 \ldots b_m)_q)| + |f_l((a_0 \ldots a_m)_q)|) \\
&\quad + |q^{-d\{b\}} - q^{-d\{a\}}| \sum_{l \in \mathcal{P}} \sum_{m \geqslant 0} q^{-dm} |f_l((a_0 \ldots a_m)_q)|
\end{aligned}
$$

because the summands for $m \leqslant k$ cancel in the first sum as the first $k+1$ digits coincide. By using the estimates

$$
\begin{aligned}
|\{b\} - \{a\}| &\leqslant |\{b\} - \{a\}|^\alpha = |b - a|^\alpha, \\
|q^{-d\{b\}} - q^{-d\{a\}}| &= \mathcal{O}(|b - a|^\alpha), \\
|f_l((b_0 \ldots b_m)_q)| &= \mathcal{O}(q^{(d-1)m} m)
\end{aligned}
$$

(see Lemma 3.9 for the last estimate), we obtain

$$
\begin{aligned}
|\Psi_1(b) - \Psi_1(a)| &= \mathcal{O}\left(|b - a|^\alpha + \sum_{m > k} m q^{-m} + |b - a|^\alpha \right) \\
&= \mathcal{O}(|b - a|^\alpha + k q^{-k}) = \mathcal{O}(|b - a|^\alpha + q^{-\alpha k}) \\
&= \mathcal{O}(|b - a|^\alpha + |y - x|^\alpha) = \mathcal{O}(|y - x|^\alpha).
\end{aligned}
$$

Here, (38) has been used in the penultimate step. $\qquad \square$

We now use the continuity of Ψ_1 and Statement 3.13 to remove the condition on coinciding digits from Statement 3.13.

Statement 3.14. *Let $a, b \in \mathbb{R}$ with $x \leqslant a < b \leqslant y$ and $\lfloor a \rfloor = \lfloor b \rfloor$. Then*

$$
|\Psi_1(b) - \Psi_1(a)| = \mathcal{O}(|y - x|^\alpha).
$$

Proof. We write the expansions of $q^{\{a\}}$ and $q^{\{b\}}$ as

$$
q^{\{a\}} = (a_0 \boldsymbol{.} a_1 \ldots)_q, \quad q^{\{b\}} = (b_0 \boldsymbol{.} b_1 \ldots)_q.
$$

This yields

$$
0 < q^{\{b\}} - q^{\{a\}} = \frac{1}{q^{\lfloor a \rfloor}}(q^b - q^a) \leqslant q^b - q^a \leqslant q^y - q^x < q^{-k}.
$$

Thus

$$
0 \leqslant (b_0 \ldots b_k)_q - (a_0 \ldots a_k)_q \leqslant 1.
$$

If $(b_0 \ldots b_k)_q = (a_0 \ldots a_k)_q$, the result follows immediately from Statement 3.13. Otherwise, we have

$$(b_0 \ldots b_k)_q = (a_0 \ldots a_k)_q + 1. \tag{39}$$

For $m \geqslant 0$, define z and z_m by $\lfloor z \rfloor = \lfloor z_m \rfloor = \lfloor a \rfloor = \lfloor b \rfloor$ and

$$q^{\{z\}} = (b_0 \boldsymbol{.} b_1 \ldots b_k)_q,$$
$$q^{\{z_m\}} = (a_0 \boldsymbol{.} a_1 \ldots a_k (q-1)^m)_q.$$

Then $\lim_{m \to \infty} z_m = z$ because of (39).

By construction of z and z_m, we have $a < z_m < z \leqslant b$ for sufficiently large m.

By continuity of Ψ_1,

$$|\Psi_1(z) - \Psi_1(z_m)| \leqslant |y - x|^\alpha \tag{40}$$

holds for sufficiently large m.

This yields

$$\begin{aligned}
|\Psi_1(b) - \Psi_1(a)| &\leqslant |\Psi_1(b) - \Psi_1(z)| + |\Psi_1(z) - \Psi_1(z_m)| \\
&\quad + |\Psi_1(z_m) - \Psi_1(a)|.
\end{aligned}$$

The third summand can be bounded by Statement 3.13 (for a and z_m) and the second by (40). The first summand is either 0 or can be bounded by Statement 3.13 (for z and b). $\qquad\square$

To finally prove (37) for sufficiently small $y - x < 1$, we only have to remove the assumption $\lfloor a \rfloor = \lfloor b \rfloor$ from Statement 3.14. We use the idea of the proof of Statement 3.14 once more.

Assume that $\lfloor y \rfloor > \lfloor x \rfloor$. By our assumption $y < x + 1$, this amounts to $\lfloor y \rfloor = \lfloor x \rfloor + 1$. For $m \geqslant 0$, define z and z_m by $z = \lfloor y \rfloor$, $\lfloor z_m \rfloor = \lfloor x \rfloor$ and $q^{\{z_m\}} = ((q-1) \boldsymbol{.} (q-1)^m)_q$. Then $\lim_{m \to \infty} z_m = z$. By continuity of Ψ_1, we have

$$|\Psi_1(z) - \Psi_1(z_m)| \leqslant |y - x|^\alpha \tag{41}$$

and $x < z_m < z \leqslant y$ for sufficiently large m.

Then, this yields

$$\begin{aligned}
|\Psi_1(y) - \Psi_1(x)| &\leqslant |\Psi_1(y) - \Psi_1(z)| + |\Psi_1(z) - \Psi_1(z_m)| \\
&\quad + |\Psi_1(z_m) - \Psi_1(x)|.
\end{aligned}$$

The third summand can be bounded by Statement 3.14 for x and z_m and the second by (41). The first vanishes or can be bounded by Statement 3.14 for z and y.

This yields

$$|\Psi_1(y) - \Psi_1(x)| = \mathcal{O}(|y - x|^\alpha).$$

Therefore, (37) is satisfied with a suitable positive constant C for $y - x < \varepsilon$ for some $\varepsilon > 0$.

Assume $y - x \geqslant \varepsilon$. As Ψ_1 is continuous and periodic, $|\Psi_1(y) - \Psi_1(x)|$ is bounded. Thus, (37) holds for a suitable positive constant C for $|y - x| \geqslant \varepsilon$.

Therefore, the function Ψ_1 is Hölder continuous of order $\alpha < 1$. $\qquad\square$

3.5 Limiting distribution

Finally, we can prove the parts of Theorem 1 concerning the approximation of the distribution function and the central limit theorem.

Proof. To prove that the distribution function can be approximated by a Gaussian mixture, we use the Berry-Esseen inequality (cf., for instance, [13, Theorems IX.5]) to estimate the difference between distribution functions. The proof follows the proof of Hwang's Quasi-Power Theorem [24]. First, we describe the two corresponding characteristic functions.

Let $\hat{g}_N(t)$ be the characteristic function of a mixture of Gaussian or degenerate distributions with weights λ_j, means $a_j \sqrt{\log_q N}$ and variances b_j for $j = 1, \ldots, c$, that is

$$\hat{g}_N(t) = \sum_{j=1}^{c} \lambda_j \exp\left(i a_j \sqrt{\log_q N} t - \frac{b_j}{2} t^2 \right)$$

with a_j, b_j and λ_j defined in (5).

By Lemma 3.7, the characteristic function $\hat{f}_N(t)$ of $\mathcal{T}(\boldsymbol{n}) / \sqrt{\log_q N}$ is

$$\hat{f}_N(t) = \sum_{j=1}^{c} \exp\left(i a_j \sqrt{\log_q N} t - \frac{b_j}{2} t^2 + \mathcal{O}\left(\frac{t^3}{\sqrt{\log N}} \right) \right)$$

$$\cdot \sum_{l \in \mathcal{P}} e^{\frac{2\pi i l}{p} \log_q N} \Psi_{lj}\left(\log_q N, \frac{t}{\sqrt{\log_q N}} \right) + R\left(N, \frac{t}{\sqrt{\log N}} \right)$$

for $t \log_q^{-\frac{1}{2}} N$ in a fixed neighborhood of 0.

Because of (27) and $R(N, 0) = 0$ (see Lemma 3.7), we have

$$\hat{f}_N(t) = \sum_{j=1}^{c} \exp\left(i a_j \sqrt{\log_q N} t - \frac{b_j}{2} t^2 \right) \exp\left(\mathcal{O}\left(\frac{t^3}{\sqrt{\log N}} \right) \right)$$

$$\cdot \left(\lambda_j + \mathcal{O}\left(\frac{t}{\sqrt{\log N}} \right) \right) + \mathcal{O}\left(N^{-\xi} t \sqrt{\log N} \right).$$

Now we use the inequality $|e^w - 1| \leqslant |w| e^{|w|}$, valid for all complex numbers w, to obtain

$$\left| \frac{1}{t} (\hat{f}_N(t) - \hat{g}_N(t)) \right| = \sum_{j=1}^{c} \mathcal{O}\left(\left(\frac{t^2 + 1}{\sqrt{\log N}} \right) \exp\left(-\frac{b_j}{2} t^2 + \mathcal{O}\left(\frac{t^3}{\sqrt{\log N}} \right) \right) \right)$$

$$+ \mathcal{O}(N^{-\xi} \log^{-\frac{1}{2}} N) \quad (42)$$

for $t \log_q^{-\frac{1}{2}} N$ in a small neighborhood of 0.

From now on, we assume that $b_j \neq 0$. There is a small neighborhood of 0 for $t \log_q^{-\frac{1}{2}} N$ such that

$$\mathcal{O}\left(\exp\left(-\frac{b_j}{2} t^2 + \mathcal{O}\left(\frac{t^3}{\sqrt{\log N}} \right) \right) \right) = \mathcal{O}\left(\exp\left(-\frac{b_j}{4} t^2 \right) \right)$$

holds.

This yields

$$\left| \frac{1}{t}(\hat{f}_N(t) - \hat{g}_N(t)) \right| = \sum_{j=1}^{c} \mathcal{O}\left(\exp\left(-\frac{b_j}{4}t^2 \right) \frac{t^2+1}{\sqrt{\log N}} \right) + \mathcal{O}(N^{-\xi} \log^{\frac{1}{2}} N).$$

Now, the Berry-Esseen inequality with $T = c\sqrt{\log_q N}$ for a small constant $c > 0$ (cf., for instance, [13, Theorem IX.5]) implies that

$$\sup_{x \in \mathbb{R}} |F_N(x) - G_N(x)| = \mathcal{O}\left(\frac{1}{\sqrt{\log N}} \right)$$

where F_N is the cumulative distribution function of $\mathcal{T}(\boldsymbol{n})$ and G_N is the cumulative distribution function of the mixture of Gaussian distributions.

If all a_j are equal and $b_j \geqslant 0$, G_N is the distribution function of a mixture of normal (or degenerate) distributions with mean $e_{\mathcal{T}}\sqrt{\log_q N}$ and variances $b_j \geqslant 0$. After subtracting the mean, (42) converges to 0. Thus,

$$\frac{\mathcal{T}(\boldsymbol{n}) - \mathbb{E}(\mathcal{T}(\boldsymbol{n}))}{\sqrt{\log_q N}}$$

converges in distribution. If all $b_j > 0$, then the same estimates as above yield the speed of convergence. $\qquad \square$

This completes the proof of Theorem 1.

4 Fourier Coefficients — Proof of Theorem 2

This section contains the proof of the theorem about the Fourier coefficients. First, we investigate some Dirichlet series which we will use later. Then, we prove the formulas given in Theorem 2. We use the Hölder condition for Ψ_1 to prove that its Fourier series converges.

Lemma 4.1. *The Dirichlet series*

$$L(z) = \sum_{r \geqslant 1} \lfloor \log_q r \rfloor (r^d - (r-1)^d) r^{-z}$$

is meromorphic in $\Re z > d - 1$ with poles in $z = d + \frac{2\pi i l}{\log q}$ for $l \in \mathbb{Z}$. The main part at $z = d$ is

$$\frac{d}{(z-d)^2 \log q} - \frac{d}{2(z-d)}$$

and, for $l \neq 0$, the residue at $z = d + \frac{2\pi i l}{\log q}$ is $\frac{d}{2\pi i l}$.

Proof. First, we use the binomial theorem to obtain

$$L(z) = dL_1(z - d + 1) - \sum_{j=0}^{d-2} \binom{d}{j} (-1)^{d-j} L_1(z - j) \tag{43}$$

with $L_1 = \sum_{r \geqslant 1} \lfloor \log_q r \rfloor r^{-z}$. The Dirichlet series $L_1(z)$ is holomorphic for $\Re z > 1$. Thus, the second summand in (43) is holomorphic for $\Re z > d - 1$. To obtain the expansion of $L(z)$ at z with $\Re z = d$, we investigate the Dirichlet series $L_1(z)$ at $\Re z = 1$.

Let $k \geqslant 0$ be an integer. We use Euler-Maclaurin summation with $f(x) = kx^{-z}$ to obtain

$$\sum_{q^k \leqslant r < q^{k+1}} \frac{\lfloor \log_q r \rfloor}{r^z} = \int_{q^k}^{q^{k+1}} kx^{-z}\, dx - \frac{k}{2}(q^{-(k+1)z} - q^{-kz})$$

$$- kz \int_{q^k}^{q^{k+1}} B_1(\{x\}) x^{-z-1}\, dx$$

$$= \frac{1}{1-z}(kq^{(k+1)(1-z)} - kq^{k(1-z)})$$

$$- \frac{1}{2}(kq^{-(k+1)z} - kq^{-kz})$$

$$- z \int_{q^k}^{q^{k+1}} B_1(\{x\}) x^{-z-1} \lfloor \log_q(x) \rfloor\, dx$$

where $B_1(x)$ is the first Bernoulli polynomial. For $\Re z > 1$, summation over $k \geqslant 0$ yields

$$L_1(z) = \frac{1}{1-z} \sum_{k \geqslant 1} q^{k(1-z)}((k-1) - k) - \frac{1}{2} \sum_{k \geqslant 1} q^{-zk}((k-1) - k)$$

$$- z \int_1^\infty B_1(\{x\}) x^{-z-1} \lfloor \log_q(x) \rfloor\, dx$$

$$= \frac{1}{z-1} \frac{1}{q^{z-1} - 1} + \frac{1}{2} \frac{1}{q^z - 1} - z \int_1^\infty B_1(\{x\}) x^{-z-1} \lfloor \log_q(x) \rfloor\, dx.$$

The second summand and the integral are clearly holomorphic for $\Re z > 0$. Thus, $L_1(z)$ can be continued meromorphically to $\Re z > 0$ with poles coming from the first summand.

The expansion around $z = 1$ is

$$\frac{1}{z-1} \frac{1}{q^{z-1} - 1} + O(1) = \frac{1}{(z-1)^2 \log q} - \frac{1}{2(z-1)} + O(1).$$

Thus, by (43), we obtain the main part and the residues of $L(z)$ at $z = d + \frac{2\pi i l}{\log q}$ for $l \in \mathbb{Z}$ as stated in the lemma. $\qquad\square$

Lemma 4.2. *The Dirichlet series*

$$Z(z) = \sum_{r \geqslant 1} (r^d - (r-1)^d) r^{-z}$$

is meromorphic in $\mathbb{C}$ with simple poles in $z = j$, $j \in \{1, \dots, d\}$ with residues $\binom{d}{j-1}(-1)^{d-j}$.

Proof. The binomial theorem yields

$$Z(z) = \sum_{j=0}^{d-1} \binom{d}{j} (-1)^{d-j+1} \zeta(z-j),$$

where ζ is the Riemann ζ-function. The result follows from the unique pole of $\zeta(z)$ at $z = 1$ with residue 1. $\qquad\square$

Denote by $\zeta(z, \alpha)$ the Hurwitz ζ-function

$$\zeta(z, \alpha) = \sum_{r > -\alpha} (r + \alpha)^{-z}.$$

Furthermore ψ is the digamma function.

Lemma 4.3. *For $0 \leqslant \alpha < 1$ and and an integer $0 \leqslant j \leqslant d-1$, the Dirichlet series*

$$J(z, \alpha, j) = \sum_{r \geqslant 1} r^j (r + \alpha)^{-z}$$

is analytic for $\Re z > j+1$. For $j = d-1$, it is meromorphic for $\Re z > d-1$ with a simple pole at $z = d$ with expansion

$$J(z, \alpha, d-1) = \frac{1}{z - d} - \psi(\alpha + [\alpha = 0]) - [\alpha > 0 \wedge d = 1]\alpha^{-1}$$

$$+ \sum_{k=0}^{d-2} \binom{d-1}{k} (-\alpha)^{d-1-k} \zeta(d-k, \alpha) + \mathcal{O}(z-d). \tag{44}$$

Proof. As $r^j (r + \alpha)^{-z} = \mathcal{O}(r^{j - \Re z})$, J is analytic for $\Re z > j+1$. Now, let $j = d-1$.
    The binomial theorem yields

$$J(z, \alpha, d-1) = \sum_{r \geqslant 1} (r + \alpha - \alpha)^{d-1} (r + \alpha)^{-z}$$

$$= \sum_{k=0}^{d-1} \binom{d-1}{k} (-\alpha)^{d-1-k} \sum_{r \geqslant 1} (r + \alpha)^{-(z-k)}$$

$$= \sum_{k=0}^{d-1} \binom{d-1}{k} (-\alpha)^{d-1-k} \left(\zeta(z-k, \alpha) - [\alpha > 0]\alpha^{-z+k} \right)$$

$$= \zeta(z-d+1, \alpha) + \sum_{k=0}^{d-2} \binom{d-1}{k} (-\alpha)^{d-1-k} \zeta(z-k, \alpha)$$

$$- [\alpha > 0 \wedge d = 1]\alpha^{-z}.$$

Using the expansion (cf. [35, p. 271])

$$\zeta(z, \alpha) = \frac{1}{z - 1} - \psi(\alpha + [\alpha = 0]) + \mathcal{O}(z-1)$$

yields (44). $\qquad\square$

Lemma 4.4. *Let $k \in \mathbb{Z}$. The Dirichlet series*

$$B(z) = \boldsymbol{w}_k^\top \sum_{r=1}^\infty \left(\boldsymbol{B}_\emptyset(r+1) - 2\boldsymbol{B}_\emptyset(r) + \boldsymbol{B}_\emptyset(r-1)\right) r^{-z}$$

is analytic for $\Re z > d - 1$.

Proof. By the definition (30), we have

$$\boldsymbol{B}_\emptyset(r+1) - \boldsymbol{B}_\emptyset(r) = \sum_{\emptyset \neq C \subseteq \{1,\ldots,d\}} \boldsymbol{B}_C(r), \tag{45}$$

which can be bounded by $\|\boldsymbol{B}_C(r)\| = \mathcal{O}(r^{d-1} \log r)$. Thus,

$$B(z) = \boldsymbol{w}_k^\top \sum_{\emptyset \neq C \subseteq \{1,\ldots,d\}} \sum_{r \geqslant 1} (\boldsymbol{B}_C(r) - \boldsymbol{B}_C(r-1)) r^{-z}$$

which converges for $\Re z > d - 1$ by [2, Theorem 8.1]. $\qquad\square$

The vector-valued functions $\boldsymbol{H}_C(z)$ are defined by the Dirichlet series

$$\boldsymbol{H}_C(z) = \sum_{r \geqslant 1} \boldsymbol{B}_C(r) r^{-z}. \tag{46}$$

By (6) and (45), this yields

$$\boldsymbol{H}(z) = \sum_{\emptyset \neq C \subseteq \{1,\ldots,d\}} \boldsymbol{H}_C(z) = \sum_{r \geqslant 1} (\boldsymbol{B}_\emptyset(r+1) - \boldsymbol{B}_\emptyset(r)) r^{-z}. \tag{47}$$

Next, we investigate the Dirichlet series $\boldsymbol{H}_C$. In particular, we determine its behavior at $z = d + \chi_k$ and provide an infinite functional equation to compute its residues at these points. This will finally give us the residues of $\boldsymbol{H}$ in (7). We use a similar method as Grabner and Hwang in [17].

For this infinite recursion, define

$$\boldsymbol{\delta}_{C,D}^\varepsilon = \sum_{\substack{\beta_i=0 \\ i \notin C \cup D}}^{q-1} \sum_{\substack{\beta_i=0 \\ i \in D}}^{\varepsilon-1} \sum_{\substack{\beta_i=\varepsilon \\ i \in C}} \boldsymbol{\delta}_\beta, \tag{48}$$

in analogy to the definition of $M_{C,D}^\varepsilon$. As before, the s-th entry of $\boldsymbol{\delta}_\varepsilon$ is the output label of the transition starting in s with input label ε. Then, $\boldsymbol{\delta} = \boldsymbol{\delta}_{\emptyset,\emptyset}^\varepsilon$ holds independently of ε. Furthermore, $\boldsymbol{\delta}_{C,D}^\varepsilon = \frac{d}{dt} M_{C,D}^\varepsilon \boldsymbol{1}\big|_{t=0}$ by (35).

Lemma 4.5. *Let $C \neq \emptyset$. For $\Re z > d$ and $C \neq \emptyset$, the Dirichlet series $\boldsymbol{H}_C(z)$ satisfies the following infinite recursion*

$$
\left(1 - q^{-z} \sum_{\varepsilon=0}^{q-1} M_{C,\emptyset}^\varepsilon\right) \boldsymbol{H}_C(z) =
$$
$$
\sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon)\varepsilon^{-z} + q^{-z} \sum_{\emptyset \neq D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \boldsymbol{H}_{C \cup D}(z)
$$
$$
+ q^{-z} \sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,D}^\varepsilon J\left(z, \frac{\varepsilon}{q}, d - |D| - |C|\right) \tag{49}
$$
$$
+ \sum_{D \subseteq C^c} \sum_{m \geqslant 1} \binom{-z}{m} q^{-z-m} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \varepsilon^m \boldsymbol{H}_{C \cup D}(z+m).
$$

It is analytic for $\Re z > d - |C| + 1$. For $|C| = 1$ and $k \neq 0$, $\boldsymbol{w}_k^\top \boldsymbol{H}_C$ has a possible simple pole in $z = d + \chi_k$ with residue the right-hand side of (49) evaluated at $z = d + \chi_k$ and divided by $\log q$. For $|C| = 1$, $\boldsymbol{w}_0^\top \boldsymbol{H}_C$ has a possible double pole with main part

$$
\frac{e_\mathcal{T}}{\log q} \frac{1}{(z-d)^2} + \left(\frac{e_\mathcal{T}}{2} + \frac{h_C}{\log q}\right) \frac{1}{z-d}
$$

where h_C is given in (52).

Remark 4.6. The infinite recursion (49) can be used to numerically compute the values of $\boldsymbol{H}_C$ and its residues at $z = d + \chi_k$ with arbitrary precision. It numerically converges fast if the first terms of the Dirichlet series $\boldsymbol{H}_C$ are computed explicitly.

Proof. As $\boldsymbol{B}_C(r) = \mathcal{O}(r^{d-|C|} \log r)$, the Dirichlet series $\boldsymbol{H}_C$ is analytic for $\Re z > d - |C| + 1$.

By multiplying (20) with $\boldsymbol{u}(t)$, differentiating with respect to t at $t = 0$ and using (34), (19) and (48), we obtain the recursion

$$
\boldsymbol{B}_C(qr + \varepsilon) = \sum_{D \subseteq C^c} M_{C,D}^\varepsilon \boldsymbol{B}_{C \cup D}(r) + \boldsymbol{\delta}_{C,D}^\varepsilon r^{d-|D|-|C|} \tag{50}
$$

for $C \neq \emptyset, \{1, \ldots, d\}$ and $qr + \varepsilon \geqslant 0$. By (17), this recursion is also valid for $C = \{1, \ldots, d\}$ and $qr + \varepsilon > 0$.

By (50), we have

$$
\boldsymbol{H}_C(z) = \sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon)\varepsilon^{-z} + \sum_{\varepsilon=0}^{q-1} \sum_{r \geqslant 1} \boldsymbol{B}_C(qr + \varepsilon)(qr + \varepsilon)^{-z}
$$
$$
= \sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon)\varepsilon^{-z} +
$$

$$\sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \sum_{r \geqslant 1} (M_{C,D}^\varepsilon \boldsymbol{B}_{C \cup D}(r) + \boldsymbol{\delta}_{C,D}^\varepsilon r^{d-|D|-|C|}) q^{-z} r^{-z} \left(1 + \frac{\varepsilon}{qr}\right)^{-z}$$

for $C \neq \emptyset$. Expanding $(1 + \varepsilon/(qr))^{-z}$ as a binomial series yields

$$
\begin{aligned}
\boldsymbol{H}_C(z) &= \sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon) \varepsilon^{-z} \\
&\quad + \sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \sum_{r \geqslant 1} \sum_{m \geqslant 0} \binom{-z}{m} M_{C,D}^\varepsilon \varepsilon^m q^{-z-m} \boldsymbol{B}_{C \cup D}(r) r^{-z-m} \\
&\quad + q^{-z} \sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,D}^\varepsilon J\left(z, \frac{\varepsilon}{q}, d - |D| - |C|\right) \\
&= \sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon) \varepsilon^{-z} + q^{-z} \sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \boldsymbol{H}_{C \cup D}(z) \\
&\quad + q^{-z} \sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,D}^\varepsilon J\left(z, \frac{\varepsilon}{q}, d - |D| - |C|\right) \\
&\quad + \sum_{D \subseteq C^c} \sum_{m \geqslant 1} \binom{-z}{m} q^{-z-m} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \varepsilon^m \boldsymbol{H}_{C \cup D}(z+m)
\end{aligned}
$$

for $\Re z > d$ and $C \neq \emptyset$. Collecting $\boldsymbol{H}_C(z)$ on the left-hand side results in (49).

To compute the residues of $\boldsymbol{w}_k^\top \boldsymbol{H}_C$ for $|C| = 1$ at $z = d + \chi_k$, note that $\sum_{\varepsilon=0}^{q-1} M_{C,\emptyset}^\varepsilon = M$ holds independently of C.

We multiply (49) with the left eigenvector $\boldsymbol{w}_k^\top$ which results in

$$
\begin{aligned}
\left(1 - q^{d-z} \exp\left(\frac{2\pi i k}{p}\right)\right) \boldsymbol{w}_k^\top \boldsymbol{H}_C(z) &= \\
\boldsymbol{w}_k^\top \sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon) \varepsilon^{-z} & \\
+ q^{-z} \boldsymbol{w}_k^\top \sum_{\emptyset \neq D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \boldsymbol{H}_{C \cup D}(z) & \\
+ q^{-z} \boldsymbol{w}_k^\top \sum_{D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,D}^\varepsilon J\left(z, \frac{\varepsilon}{q}, d - |D| - 1\right) & \\
+ \boldsymbol{w}_k^\top \sum_{D \subseteq C^c} \sum_{m \geqslant 1} \binom{-z}{m} q^{-z-m} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \varepsilon^m \boldsymbol{H}_{C \cup D}(z+m). &
\end{aligned}
\tag{51}
$$

As $|C \cup D| \geqslant 2$ or $\Re z + m > d$, all $\boldsymbol{H}_{C \cup D}$ used on right-hand side of (51) are well defined for $\Re z > d - 1$. The Dirichlet series J have simple poles at $z = d$ for $|C| = 1$ and

$D = \emptyset$ (Lemma 4.3). Thus the right-hand side of (51) is meromorphic for $\Re z > d - 1$ with a simple pole at $z = d$.

The factor $1 - q^{d-z} \exp(\frac{2\pi i k}{p})$ has a zero exactly for $z = d + \chi_k$, $k \in \mathbb{Z}$. Thus for $k \neq 0$, $\boldsymbol{w}_k^\top \boldsymbol{H}_C$ has a possible simple pole at $z = d + \chi_k$. Its residue is the right-hand side of (51) evaluated at $z = d + \chi_k$ divided by $\log q$.

If $k = 0$, we have $z = d$. In this case the expansion of the right-hand side of (51) is

$$\frac{e_\mathcal{T}}{z - d} + h_C + \mathcal{O}(z - d)$$

with

$$h_C = -e_\mathcal{T} \log q - q^{-d} \boldsymbol{w}_0^\top \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,\emptyset}^\varepsilon \psi\left(\frac{\varepsilon}{q} + [\varepsilon = 0]\right) \tag{52}$$

$$- [d = 1] \boldsymbol{w}_0^\top \sum_{\varepsilon=1}^{q-1} \boldsymbol{\delta}_{C,\emptyset}^\varepsilon \varepsilon^{-1}$$

$$+ q^{-d} \boldsymbol{w}_0^\top \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,\emptyset}^\varepsilon \sum_{k=0}^{d-2} \binom{d-1}{k}\left(-\frac{\varepsilon}{q}\right)^{d-1-k} \zeta\left(d - k, \frac{\varepsilon}{q}\right)$$

$$+ \boldsymbol{w}_0^\top \sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_C(\varepsilon) \varepsilon^{-d} + q^{-d} \boldsymbol{w}_0^\top \sum_{\emptyset \neq D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon H_{C \cup D}(d)$$

$$+ q^{-d} \boldsymbol{w}_0^\top \sum_{\emptyset \neq D \subseteq C^c} \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,D}^\varepsilon J\left(d, \frac{\varepsilon}{q}, d - |D| - 1\right)$$

$$+ \boldsymbol{w}_0^\top \sum_{D \subseteq C^c} \sum_{m \geqslant 1} \binom{-d}{m} q^{-d-m} \sum_{\varepsilon=0}^{q-1} M_{C,D}^\varepsilon \varepsilon^m \boldsymbol{H}_{C \cup D}(d + m)$$

where we used the expansion of J in Lemma 4.3, $\boldsymbol{\delta} = \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}_{C,\emptyset}^\varepsilon$ and (14). $\qquad \square$

From the previous lemma and (47), the residues of the Dirichlet function $\boldsymbol{H}$ follow. Only $\boldsymbol{H}_C$ with $|C| = 1$ contribute as all other summands are holomorphic.

Lemma 4.7. *The Dirichlet function $\boldsymbol{H}$ is meromorphic in $\Re z > d - 1$ with possible simple poles at $z = d + \chi_k$, $k \neq 0$ and a possible double pole at $z = d$.*

The residue at $z = d + \chi_k$, $k \neq 0$ is

$$\frac{1}{\log q} \sum_{j=1}^{d}\left(\sum_{\varepsilon=1}^{q-1} \boldsymbol{B}_{\{j\}}(\varepsilon) \varepsilon^{-d-\chi_k}\right.$$

$$\left. + q^{-d-\chi_k} \sum_{\emptyset \neq D \subseteq \{j\}^c} \sum_{\varepsilon=0}^{q-1} M_{\{j\},D}^\varepsilon \boldsymbol{H}_{\{j\} \cup D}(d + \chi_k)\right.$$

$$+ q^{-d-\chi_k} \sum_{D \subseteq \{j\}^c} \sum_{\varepsilon=0}^{q-1} \boldsymbol{\delta}^\varepsilon_{\{j\},D} J\left(d + \chi_k, \frac{\varepsilon}{q}, d - |D| - 1\right)$$

$$+ \sum_{D \subseteq \{j\}^c} \sum_{m \geqslant 1} \binom{-d-\chi_k}{m} q^{-d-m-\chi_k} \sum_{\varepsilon=0}^{q-1} M^\varepsilon_{\{j\},D} \varepsilon^m \boldsymbol{H}_{\{j\} \cup D}(d + m + \chi_k)\Bigg).$$

The main part at $z = d$ is

$$\frac{e_\mathcal{T} d}{\log q} \frac{1}{(z-d)^2} + \left(\frac{e_\mathcal{T} d}{2} + \sum_{j=1}^{d} \frac{h_{\{j\}}}{\log q}\right) \frac{1}{z-d}$$

where $h_{\{j\}}$ is defined in (52).

Now we can prove the formulas for the Fourier coefficients.

Proof of Theorem 2. The periodic fluctuation Ψ_1 of the expected value is a p-periodic function. We use the explicit expression of Ψ_1 given in Lemma 3.9.

Due to absolute convergence, the k-th Fourier coefficient of $\Psi_1(x)$ is

$$c_k = \frac{1}{p} \int_0^p \Psi_1(x) e^{-\frac{2\pi i k}{p} x} \, dx$$

$$= -\frac{e_\mathcal{T}}{p} \int_0^p \{x\} e^{-\frac{2\pi i k}{p} x} \, dx - \sum_{l \in \mathcal{P}} \sum_{m=0}^{\infty} q^{-dm} e^{-\frac{2\pi i l m}{p}} I_{l,m}$$

with

$$I_{l,m} = \frac{1}{p} \int_0^p q^{-d\{x\}} \exp\left(\frac{2\pi i l}{p} \lfloor x \rfloor - \frac{2\pi i k}{p} x\right) f_l((x_0 \ldots x_m)_q) \, dx$$

and $q^{\{x\}} = (x_0 \centerdot x_1 \ldots)_q$. The value of the first integral is given by $-\frac{e_\mathcal{T}}{2}$ for $k = 0$, and $[k \equiv 0 \bmod p] \frac{e_\mathcal{T}}{\chi_k \log q}$ otherwise. Thus, we focus on the second integral $I_{l,m}$.

First, we partition the interval $[0, p)$ into intervals $[r, r+1)$ for $r = 0, \ldots, p-1$. After simplifying the sum of p-th roots of unity, we obtain

$$I_{l,m} = [k \equiv l \bmod p] \int_0^1 q^{-dx} f_l((x_0 \ldots x_m)_q) e^{-\frac{2\pi i k}{p} x} \, dx.$$

After partitioning the interval $[0, 1)$ into the intervals $[\log_q r - m, \log_q(r+1) - m)$ for $r = q^m, \ldots, q^{m+1} - 1$, the function $f_l((x_0 \ldots x_m)_q)$ is constant on the interval of integration. Therefore, we obtain

$$\sum_{l \in \mathcal{P}} \sum_{m=0}^{\infty} q^{-md} e^{-\frac{2\pi i l m}{p}} I_{l,m} = \frac{1}{(d + \chi_k) \log q} \sum_{r=1}^{\infty} f_{k \bmod p}(r) \left(r^{-d-\chi_k} - (r+1)^{-d-\chi_k}\right).$$

Next, consider the function

$$A(z) = \sum_{r=1}^{\infty} f_{k \bmod p}(r) \left(r^{-z} - (r+1)^{-z} \right).$$

We know that $f_l(r) = \mathcal{O}(r^{d-1} \log r)$. Thus, $A(z)$ is analytic for $\Re z > d - 1$.

By summation by parts, we can rearrange the series for $\Re z > d$ and obtain a sum of Dirichlet series

$$A(z) = [p \mid k] e_{\mathcal{T}} S_1(z) + i \boldsymbol{w}_k'^{\top} \boldsymbol{1} S_2(z) - S_3(z) + q^d \exp\left(\frac{2\pi i k}{p}\right) S_4(z) \qquad (53)$$

with coefficients $s_1(r)$, $s_2(r)$, $s_3(r)$ and $s_4(r)$ respectively. These coefficients are differences of the four summands in $f_{k \bmod p}(r)$ and $f_{k \bmod p}(r-1)$ in (32), respectively, e.g.,

$$s_1(r) = \lfloor \log_q(r) \rfloor (r^d - (q\lfloor r/q \rfloor)^d) + (q\lfloor r/q \rfloor)^d$$
$$- [r > 1]\big(\lfloor \log_q(r-1) \rfloor ((r-1)^d - (q\lfloor (r-1)/q \rfloor)^d)$$
$$- (q\lfloor (r-1)/q \rfloor)^d \big).$$

After some simplifications using $\lfloor \frac{r-1}{q} \rfloor = \lfloor \frac{r}{q} \rfloor - [q \mid r]$ and $\lfloor \log_q(r-1) \rfloor = \lfloor \log_q r \rfloor - [r \text{ is a power of } q]$ (for $r \geqslant 2$), we obtain

$$
\begin{aligned}
s_1(r) &= \lfloor \log_q r \rfloor (r^d - (r-1)^d) \\
&\quad - [q \mid r] q^d \lfloor \log_q rq^{-1} \rfloor ((rq^{-1})^d - (rq^{-1} - 1)^d) \\
&\quad + [r \neq 1 \text{ is a power of } q]((r-1)^d - (r-q)^d), \\
s_2(r) &= r^d - (r-1)^d - [q \mid r] q^d \exp\left(\frac{2\pi i k}{p}\right)((rq^{-1})^d - (rq^{-1} - 1)^d), \\
s_3(r) &= \boldsymbol{w}_k^{\top}(\boldsymbol{B}_\emptyset(r) - \boldsymbol{B}_\emptyset(r-1)), \\
s_4(r) &= [q \mid r] \boldsymbol{w}_k^{\top}(\boldsymbol{B}_\emptyset(rq^{-1}) - \boldsymbol{B}_\emptyset(rq^{-1} - 1)).
\end{aligned}
\qquad (54)
$$

For $\Re z > d$, we can split up the summation into the different cases in (54). This yields

$$
\begin{aligned}
S_1(z) &= (1 - q^{d-z}) L(z) + \sum_{j=0}^{d-1} \binom{d}{j} (-1)^{d-j} \frac{1 - q^{d-j}}{q^{z-j} - 1}, \\
S_2(z) &= \left(1 - q^{d-z} \exp\left(\frac{2\pi i k}{p}\right) \right) Z(z), \\
S_3(z) &= \boldsymbol{w}_k^{\top} \boldsymbol{H}(z) - B(z), \\
S_4(z) &= q^{-z} \boldsymbol{w}_k^{\top} \boldsymbol{H}(z) - q^{-z} B(z)
\end{aligned}
$$

where we used (45), (47) and the Dirichlet series defined in Lemmas 4.1, 4.2 and 4.4.

Thus, in (53), we obtain

$$
\begin{aligned}
A(z) = {} & [p \mid k] e_{\mathcal{T}} \sum_{j=0}^{d-1} \binom{d}{j} (-1)^{d-j} \frac{1 - q^{d-j}}{q^{z-j} - 1} \\
& + i \boldsymbol{w}_k'^{\top} \boldsymbol{1} \left(1 - q^{d-z} e^{\frac{2\pi i k}{p}} \right) Z(z) \\
& - \left(1 - q^{d-z} e^{\frac{2\pi i k}{p}} \right) \boldsymbol{w}_k^{\top} \boldsymbol{H}(z) \\
& + [p \mid k] e_{\mathcal{T}} (1 - q^{d-z}) L(z) \\
& + \left(1 - q^{d-z} e^{\frac{2\pi i k}{p}} \right) B(z).
\end{aligned}
\tag{55}
$$

We want to evaluate A at $z = d + \chi_k$. The factors $1 - q^{d-z} e^{\frac{2\pi i k}{p}}$ are zero if and only if $z = d + \chi_k$. Thus, the following Dirichlet series contribute to (55):

- The Dirichlet series Z only contributes if $k = 0$ (Lemma 4.2).

- The Dirichlet series $\boldsymbol{w}_k^{\top} \boldsymbol{H}$ has poles at $z = d + \chi_k$ for $k \in \mathbb{Z}$. The possible double pole at $z = d$ cancels with the one of L (Lemma 4.7).

- The residues of the Dirichlet series L contribute to the Fourier coefficients. The possible double pole at $z = d$ cancels with that of $\boldsymbol{w}_0^{\top} \boldsymbol{H}$ (Lemma 4.1).

- As the Dirichlet series B converges for $\Re z > d - 1$ (Lemma 4.4), it does not contribute to the Fourier coefficients.

As the second order poles of $\boldsymbol{w}_0^{\top} \boldsymbol{H}$ and L cancel, the right-hand side of (55) is well defined for the limit $z \to d + \chi_k$. After computing the limit and simplifying the summation, we obtain (7).

Then Lemma 3.12 and Bernstein's theorem (cf. [36, p. 240]) imply the absolute and uniform convergence of the Fourier series. $\qquad \square$

Now we use Theorem 2 to prove Corollary 2.5.

Proof of Corollary 2.5. The transducer in Figure 7 computes the q-ary sum-of-digits function $s_q(n)$ and we can use Theorem 2.
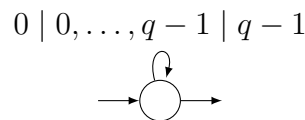
$$
0 \mid 0, \ldots, q - 1 \mid q - 1
$$

Figure 7: Transducer for the q-ary sum-of-digits function.

We transform the Dirichlet series

$$
D(z) = \sum_{m \geqslant 1} (s_q(m) - s_q(m-1)) m^{-z}
$$

in two different ways. This series is absolutely convergent for $\Re z > 1$.

First, we can rearrange the summation of the Dirichlet series $D(z)$ such that the Dirichlet series $H(z) = \sum_{m \geqslant 1} s_q(m) m^{-z}$, defined in (46), appears. We have

$$
\begin{aligned}
|H(z) - 1| &= \mathcal{O}\Big(2^{-\Re z} + \sum_{m \geqslant 3} m^{-\Re z} \log m\Big) \\
&= \mathcal{O}\Big(2^{-\Re z} + \int_2^\infty x^{-\Re z} \log x \, dx\Big) \\
&= \mathcal{O}(2^{-\Re z})
\end{aligned}
\tag{56}
$$

for $\Re z > 1$. By partial summation, we obtain

$$
\begin{aligned}
D(z) &= 1 - 2^{-z} + \sum_{m \geqslant 2} s_q(m)(m^{-z} - (m+1)^{-z}) \\
&= 1 - 2^{-z} + \sum_{m \geqslant 2} s_q(m) m^{-z}\big(1 - \big(1 + m^{-1}\big)^{-z}\big).
\end{aligned}
$$

Expanding the binomial series yields

$$
\begin{aligned}
D(z) &= 1 - 2^{-z} - \sum_{m \geqslant 2} s_q(m) m^{-z} \sum_{l \geqslant 1} \binom{-z}{l} m^{-l} \\
&= 1 - 2^{-z} - \sum_{l \geqslant 1} \binom{-z}{l}(H(z + l) - 1).
\end{aligned}
\tag{57}
$$

By (57), we have

$$
D(z) = 1 - 2^{-z} + zH(z+1) - z - \sum_{l \geqslant 2} \binom{-z}{l}(H(z+l) - 1)
$$

which is equivalent to

$$
H(z+1) = \frac{1}{z}D(z) + \frac{1}{z}(2^{-z} - 1) + 1 - \sum_{l \geqslant 2} \frac{1}{l}\binom{-z-1}{l-1}(H(z+l) - 1)
$$

for $\Re z > 1$. The sum on the right-hand side is holomorphic at $\Re z = 0$ because of (56). By meromorphic continuation, this equation also holds for $\Re z = 0$. This yields

$$
\operatorname{Res}_{z = 1 + \chi_k} H(z) = \operatorname{Res}_{z = \chi_k} H(z+1) = \operatorname{Res}_{z = \chi_k} \frac{1}{z} D(z).
\tag{58}
$$

On the other hand, we split up the summation in the definition of $D(z)$ into the q equivalence classes modulo q and we use the recursions[4]

$$
s_q(qm + \varepsilon) = s_q(m) + \varepsilon
$$

---

[4]Actually, these recursions are (36).

for $0 \leqslant \varepsilon < q$. This results in

$$s_q(m) - s_q(m-1) = 1 + [q \mid m] \left(s_q\left(q^{-1}m\right) - s_q\left(q^{-1}m - 1\right) - q \right)$$

for $m \geqslant 1$. Thus we obtain

$$D(z) = \sum_{m \geqslant 1} \left(1 + [q \mid m] \left(s_q\left(q^{-1}m\right) - s_q\left(q^{-1}m - 1\right) - q \right) \right) m^{-z}$$
$$= \zeta(z) + q^{-z}D(z) - q^{1-z}\zeta(z).$$

Thus, we obtain[5]

$$D(z) = \frac{1 - q^{1-z}}{1 - q^{-z}}\zeta(z). \tag{59}$$

This formula yields

$$\operatorname{Res}_{z=\chi_k} D(z) = -\frac{q-1}{\log q}\zeta(\chi_k). \tag{60}$$

For $k = 0$, we further use the expansion

$$\zeta(z) = -\frac{1}{2} - \frac{1}{2}\log(2\pi)z + \mathcal{O}(z^2)$$

(cf. [8, 25.6.1 and 25.6.11]) and (59) to obtain

$$D(z) = \frac{q-1}{2z \log q} + \frac{(q-1)\log(2\pi)}{2 \log q} - \frac{q+1}{4} + \mathcal{O}(z). \tag{61}$$

Thus, by (56) and (60), we obtain

$$\operatorname{Res}_{z=1+\chi_k} H(z) = \frac{1}{\chi_k}\operatorname{Res}_{z=\chi_k} D(z) = -\frac{q-1}{\chi_k \log q}\zeta(\chi_k)$$

for $k \neq 0$. For $k = 0$, (61) and (58) yield

$$\operatorname{Res}_{z=1} H(z) = \frac{(q-1)\log(2\pi)}{2 \log q} - \frac{q+1}{4}.$$

Now, (7) with $e_{\mathcal{T}} = \frac{q-1}{2}$ and $\boldsymbol{w}_0'^{\top} = 0$ yields (9). $\qquad \square$

5 Non-Differentiability — Proof of Theorem 3

In this section, we give the proof of the non-differentiability of $\Psi_1(x)$. We follow the method presented by Tenenbaum [33], see also Grabner and Thuswaldner [18].

---

[5]Note that this well-known identity can also be derived from $s_q(m) - s_q(m-1) = 1 - (q-1)v_q(m)$, where $v_q(m)$ is the q-adic valuation of m.

Proof of Theorem 3. Let $r = (r_{m-1} \dots r_0)_q$ be the value of the reset sequence $(r_{m-1} \dots r_0)$ leading to state ν.

Assume that Ψ_1 is differentiable at $x \in [0, 1)$. Let $q^x = (\varepsilon_0 \, . \, \varepsilon_1 \dots)_q$ be the standard q-ary digit expansion choosing the representation ending on 0^ω in the case of ambiguity. Further, let x_k be such that $q^{x_k} = (\varepsilon_0 \, . \, \varepsilon_1 \dots \varepsilon_k)_q$. Thus, we have $\lim_{k \to \infty} x_k = x$. For $f \in \{0, 1\}$, the function $L_f \colon \mathbb{Z} \to \mathbb{Z}$ is defined as $L_f(k) = ck + f$ with c a positive integer such that $c > \frac{1}{\xi} - 1$. Define $N_k = q^{x_k + k + L_f(k)}$ and $h(k) = \lfloor q^{ck + \frac{c}{c+1} x_k - m - 2} \rfloor$. Let y_k and z_k be such that $N_k + q^{ck - m - 1} r = q^{y_k + k + L_f(k)}$ and $N_k + q^{ck - m - 1} r + h(k) = q^{z_k + k + L_f(k)}$.

From these definitions, we know that

$$\frac{h(k)}{N_k} = \Theta(q^{-k}),$$
$$N_k^{1-\xi} \log N_k = o(h(k))$$

for $k \to \infty$. Apart from x_k, also, y_k and z_k converge to x and satisfy the following bounds:

$$z_k - y_k = \frac{1}{\log q} \frac{h(k)}{N_k} + \mathcal{O}\left(\frac{h(k)^2}{N_k^2}\right),$$
$$|y_k - x_k| = \mathcal{O}(q^{-k}),$$
$$x - x_k = \mathcal{O}(q^{-k}).$$

Now, we compute

$$\frac{1}{h(k)} \sum_{n \in \mathcal{N}_k} \mathcal{T}(n) \tag{62}$$

in two different ways where $\mathcal{N}_k = \{n \in \mathbb{Z} \mid N_k + q^{ck - m - 1} r \leqslant n < N_k + q^{ck - m - 1} r + h(k)\}$.

First, observe that $q^{ck - 1} \mid N_k$ and $h(k) < q^{ck - m - 1}$. Thus, the digit representations of the three summands in $N_k + q^{ck - m - 1} r + n$ are not overlapping at non-zero digits for $n < h(k)$. Since the digit expansion of r is a reset sequence, we have

$$\mathcal{T}(N_k + q^{ck - m - 1} r + n) = \boldsymbol{e}_\nu^\top \boldsymbol{b}(N_k q^{-ck + 1}) + \mathcal{T}(q^{ck - m - 1} r + n) - b(\nu)$$

where $\boldsymbol{e}_\nu^\top \boldsymbol{b}(N)$ is the output of the transducer when starting in state ν with input N and $b(\nu)$ is the final output at state ν.

Thus, we have

$$\frac{1}{h(k)} \sum_{n \in \mathcal{N}_k} \mathcal{T}(n) = \frac{1}{h(k)} \sum_{0 \leqslant n < h(k)} \mathcal{T}(N_k + q^{ck - m - 1} r + n)$$
$$= \boldsymbol{e}_\nu^\top \boldsymbol{b}(N_k q^{-ck + 1}) - b(\nu) + \frac{1}{h(k)} \sum_{n < h(k)} \mathcal{T}(q^{ck - m - 1} r + n)$$

where only the first summand depends on $L_f(k)$ and hence on f.

Taking the difference in (3), there is a second way of computing the sum in (62). Using the periodicity and continuity of $\Psi_1(x)$ yields

$$
\begin{aligned}
\sum_{n \in \mathcal{N}_k} \mathcal{T}(n) = {} & (N_k + q^{ck-m-1}r)e_{\mathcal{T}}(z_k - y_k) + h(k)e_{\mathcal{T}}(x + k + L_f(k)) \\
& + (N_k + q^{ck-m-1}r)(\Psi_1(z_k) - \Psi_1(y_k)) \\
& + h(k)\Psi(x) + o(h(k)).
\end{aligned}
\tag{63}
$$

Next, we use our assumption that Ψ_1 is differentiable at x to replace the difference by the derivative

$$
\Psi_1(z_k) - \Psi_1(y_k) = \Psi_1'(x)(z_k - y_k) + o(|z_k - x|) + o(|x - y_k|).
$$

Now, we insert this into (63), divide by $h(k)$ and obtain

$$
\frac{1}{h(k)} \sum_{n \in \mathcal{N}_k} \mathcal{T}(n) = \frac{e_{\mathcal{T}}}{\log q} + e_{\mathcal{T}}(x + k + L_f(k)) + \frac{1}{\log q}\Psi_1'(x) + \Psi_1(x) + o(1).
$$

Thus, we have the following equality

$$
\begin{aligned}
\boldsymbol{e}_\nu^\top \boldsymbol{b}(N_k q^{-ck+1}) - b(\nu) + \frac{1}{h(k)} \sum_{n < h(k)} \mathcal{T}(q^{ck-m-1}r + n) & \\
= \frac{e_{\mathcal{T}}}{\log q} + e_{\mathcal{T}}(x + k + L_f(k)) & + \frac{1}{\log q}\Psi_1'(x) + \Psi_1(x) + o(1)
\end{aligned}
$$

twice, for $f \in \{0, 1\}$. Subtracting these two from each other yields

$$
\boldsymbol{e}_\nu^\top \boldsymbol{b}(q^{x_k+k+2}) - \boldsymbol{e}_\nu^\top \boldsymbol{b}(q^{x_k+k+1}) = e_{\mathcal{T}} + o(1).
$$

Since the left-hand side is an integer, but the right-hand side is not for k large enough, this contradicts our assumption that Ψ_1 is differentiable at x. $\qquad \square$

6 Recursions — Proof of Theorem 4

In this section, we construct a transducer associated to the sequence defined by the recursion in (10). All inequalities, maxima and minima in this section are considered coordinate-wise.

Define the function $A \colon \mathbb{N}_0^d \to \mathbb{N}_0^d \cup \{\infty\}$ by

$$
A(q^\kappa \boldsymbol{n} + \boldsymbol{\lambda}) = \begin{cases} q^{\kappa\lambda}\boldsymbol{n} + \boldsymbol{r}_{\boldsymbol{\lambda}} & \text{if } q^{\kappa\lambda}\boldsymbol{n} + \boldsymbol{r}_{\boldsymbol{\lambda}} \geqslant 0, \\ \infty & \text{else} \end{cases}
$$

for $0 \leqslant \boldsymbol{\lambda} < q^\kappa \boldsymbol{1}$ and $\boldsymbol{n} \geqslant 0$. So, if $A(\boldsymbol{n}) < \infty$, then the recursion (10) can be used for this argument because the argument on the right-hand side is non-negative, i.e., $a(\boldsymbol{n}) = a(A(\boldsymbol{n})) + t_{\boldsymbol{n} \bmod q^\kappa}$.

First, we construct a non-deterministic transducer $\widetilde{\mathcal{T}}$. A priori, it has an infinite number of states; later, we will prove that only finitely many of them are accessible. We then simplify it to obtain a finite, deterministic, subsequential, complete transducer $\mathcal{T}$.

The set of states of $\widetilde{\mathcal{T}}$ is

$$\{(\boldsymbol{l}, j)_F \mid \boldsymbol{l} \in \mathbb{Z}^d, j \in \mathbb{N}_0\} \cup \{(\boldsymbol{l}, j)_N \mid \boldsymbol{l} \in \mathbb{Z}^d, j \in \mathbb{N}_0\}.$$

The initial state is $(0, 0)_F$; all states $(\boldsymbol{l}, j)_F$ are final states with final output $a(\boldsymbol{l})$ if $\boldsymbol{l} \geqslant 0$ and final output 0 otherwise[6]. As an abbreviation, we will frequently speak about "a state $(\boldsymbol{l}, j)$" if we do not want to distinguish between $(\boldsymbol{l}, j)_F$ and $(\boldsymbol{l}, j)_N$. We call $\boldsymbol{l}$ the *carry* and j the *level* of the state $(\boldsymbol{l}, j)$. A state $(\boldsymbol{l}, j)_F$ is called *simple*, if it is final, $\boldsymbol{l} \geqslant 0$ and $j \leqslant \kappa$.

There are two types of transitions in $\widetilde{\mathcal{T}}$, *recursion transitions* and *storing transitions*. Each state is either the origin of one recursion transition or of q^d storing transitions.

There is a recursion transition leaving $(\boldsymbol{l}, j)$ if

- $j \geqslant \kappa$ and

- $A(q^j \boldsymbol{n} + \boldsymbol{l}) < \infty$ for all $\boldsymbol{n} \geqslant 0$ with $\boldsymbol{n} \neq 0$.

In that case, we write $\boldsymbol{l} = q^\kappa \boldsymbol{s} + \boldsymbol{\lambda}$ for a $0 \leqslant \boldsymbol{\lambda} < q^\kappa \boldsymbol{1}$ and the transition leads to the state $(\boldsymbol{l}', j')_N$ with $j' = \kappa_{\boldsymbol{\lambda}} + j - \kappa$ and $\boldsymbol{l}' = q^{\kappa_{\boldsymbol{\lambda}}} \boldsymbol{s} + \boldsymbol{r}_{\boldsymbol{\lambda}}$. The input label is empty, the output label is $t_{\boldsymbol{\lambda}}$. Thus

$$A(q^j \boldsymbol{n} + \boldsymbol{l}) = q^{j'} \boldsymbol{n} + \boldsymbol{l}' \tag{64}$$

for $\boldsymbol{n} \geqslant 0$ with $\boldsymbol{n} \neq 0$. Note that (64) holds for $\boldsymbol{n} = 0$ if and only if $\boldsymbol{l} \geqslant 0$ and $\boldsymbol{l}' \geqslant 0$.

Otherwise, there are storing transitions from $(\boldsymbol{l}, j)$ to $(q^j \boldsymbol{\varepsilon} + \boldsymbol{l}, j + 1)_F$ with input $\boldsymbol{\varepsilon}$ and output 0 for all $0 \leqslant \boldsymbol{\varepsilon} < q\boldsymbol{1}$.

We now define the classes $F_1, \dots, F_K$ announced in Section 2.6. For each accessible cycle in $\widetilde{\mathcal{T}}$ with simple states and input 0, the carries of its states form one of these classes. The other classes are the singletons of those carries $\boldsymbol{l} \geqslant 0$ in the accessible part of $\widetilde{\mathcal{T}}$ with $A(\boldsymbol{l}) = \infty$. These sets will turn out to be disjoint by Lemma 6.6 and the finiteness of K will follow from the finiteness of the accessible part of $\widetilde{\mathcal{T}}$ (Lemma 6.4).

Remark 6.1. We also give a combinatorial description of those classes $F_1, \dots, F_K$ which do not come from cycles in $\widetilde{\mathcal{T}}$: Let $\boldsymbol{l} \geqslant 0$ be a carry of an accessible state of $\widetilde{\mathcal{T}}$. Then $A(\boldsymbol{l}) = \infty$ if and only if there is a recursion transition from some $(\boldsymbol{l}, j)$ to some $(\boldsymbol{l}', j')$ with $\boldsymbol{l}' \not\geqslant 0$.

Proof. Let $(\boldsymbol{l}, j_0)$ be any accessible state with carry $\boldsymbol{l}$. We use the longest path with input 0 using storing transitions only to arrive in some state $(\boldsymbol{l}, j)$—again, finiteness of this process will follow from the finiteness of the accessible part and the fact that the levels increase along storing transitions. As there is no storing transition leaving $(\boldsymbol{l}, j)$ by construction, there is a recursion transition from $(\boldsymbol{l}, j)$ to some $(\boldsymbol{l}', j')$. By the remark following (64), $\boldsymbol{l}' = A(\boldsymbol{l})$ or $\boldsymbol{l}' \not\geqslant 0$. $\qquad\square$

---

[6]In fact, we will prove that a path with valid input will never end in a state $(\boldsymbol{l}, j)_F$ with $\boldsymbol{l} \not\geqslant 0$, but the framework of subsequential transducers requires us to specify a final output even in that case. The non-final states $(\boldsymbol{l}, j)_N$ will disappear in the reduction to $\mathcal{T}$ anyway.

As usual, if reaching a state which is the origin of a transition with empty input, the process may stay in that state or may continue to the destination state writing the output of the transition without reading an input. This is the reason why the transducer is non-deterministic.

Note that in our case, transitions with empty input (i.e., recursion transitions) lead to non-final states and transitions with non-empty input (i.e., storing transitions) lead to final states. Combined with the fact that each state is either the origin of one recursion transition or of q^d storing transitions, processing an input is in fact deterministic: For every admissible input—we do not allow leading zeros—, there exists exactly one path leading from the initial state to a final state with the given input. This will enable us to simplify the transducer $\widetilde{\mathcal{T}}$ to a deterministic transducer $\mathcal{T}$ later on.

We need the property that the carries of accessible states are not "too negative":

Lemma 6.2. *1. If $(\boldsymbol{l}, j)$ is an accessible state, then*

$$q^j \boldsymbol{n} + \boldsymbol{l} \geqslant 0 \tag{65}$$

holds for all $\boldsymbol{n} \geqslant 0$ with $\boldsymbol{n} \neq 0$.

2. If $d \geqslant 2$ and $(\boldsymbol{l}, j)$ is an accessible state, then

$$\boldsymbol{l} \geqslant 0.$$

3. Any accessible transition with input $\boldsymbol{\varepsilon} \neq 0$ leads to a state $(\boldsymbol{l}, j)$ with $\boldsymbol{l} \geqslant 0$.

4. If $d = 1$ and (l, j) is an accessible state, then

$$l \geqslant l_{\min} = \min_{\lambda} \left\{ 0, \frac{-1 + \frac{r_\lambda}{q^{\kappa_\lambda}}}{\frac{1}{q^{\kappa_\lambda}} - \frac{1}{q^{\kappa}}} \right\}.$$

Proof. The first assertion is easily shown by induction and (64). The second assertion follows by induction and from the assumption that $r_\lambda \geqslant 0$ holds for all $\boldsymbol{\lambda}$. To prove the third assertion, we use (65) on the originating state of the transition.

The last assertion is shown by induction. It is clearly valid in the initial state. For storing transitions, the value of l is non-decreasing. If there is a recursion transition from some (l, j) to some $(l', j')_N$, we have

$$l' = q^{\kappa_\lambda} \left\lfloor \frac{l}{q^\kappa} \right\rfloor + r_\lambda \geqslant q^{\kappa_\lambda} \left(\frac{l}{q^\kappa} - 1 + \frac{r_\lambda}{q^{\kappa_\lambda}} \right)$$
$$\geqslant q^{\kappa_\lambda} \left(\frac{l_{\min}}{q^\kappa} + l_{\min} \left(\frac{1}{q^{\kappa_\lambda}} - \frac{1}{q^\kappa} \right) \right) = l_{\min}.$$

$\square$

As leading zeros are not allowed, the last transition in the computation path of any valid input has input $\boldsymbol{\varepsilon} \neq 0$ and thus leads to a state with a non-negative carry.

For our further investigations and finally the correctness proof, we need a suitable invariant:

Lemma 6.3. *Consider a path from $(\boldsymbol{l}, j)$ to $(\boldsymbol{l}', j')$ with input label $\boldsymbol{\varepsilon}_{m-1} \ldots \boldsymbol{\varepsilon}_0$, output label $\delta_{m'-1} \ldots \delta_0$ using L recursion transitions and $\boldsymbol{n} \geqslant 0$. Thus m' is the number of transitions and $m = m' - L$ is the number of storing transitions.*

If $\boldsymbol{n} \neq 0$ or if the last transition is a storing transition with non-zero input $\boldsymbol{\varepsilon}_{m-1}$, then

$$A^L(q^j(q^m\boldsymbol{n} + (\boldsymbol{\varepsilon}_{m-1} \ldots \boldsymbol{\varepsilon}_0)_q) + \boldsymbol{l}) = q^{j'}\boldsymbol{n} + \boldsymbol{l}', \tag{66}$$

and, if the recursion (10) *is well-posed,*

$$a(q^j(q^m\boldsymbol{n} + (\boldsymbol{\varepsilon}_{m-1} \ldots \boldsymbol{\varepsilon}_0)_q) + \boldsymbol{l}) = a(q^{j'}\boldsymbol{n} + \boldsymbol{l}') + \sum_{k=0}^{m'-1} \delta_k. \tag{67}$$

Proof. First consider the case that the path consists of a single transition. If it is a storing transition, then $L = 0$, $m = 1$, and all assertions follow from the definition and Lemma 6.2. On the other hand, if the transition is a recursion transition, we have $L = 1$, $m = 0$, and all assertions again follow from the definition, Lemma 6.2 and (64).

By induction on the length of the path, we obtain (66) and (67). $\qquad\square$

We are now able to prove the finiteness of the accessible part.

Lemma 6.4. *The transducer has a finite number of accessible states.*

Proof. For a recursion transition from $(\boldsymbol{l}, j)$ to $(\boldsymbol{l}', j')_N$, we have $j > j'$. Thus, there are no infinite paths consisting only of recursion transitions. In particular, there exist no cycles of recursion transitions.

For $d = 1$, let $J \geqslant \kappa$ be minimal such that $q^{J-\kappa} \geqslant -\lfloor \frac{l_{\min}}{q^\kappa} \rfloor - \min_\lambda q^{-\kappa_\lambda} r_\lambda$. Then $A(q^j + l) < \infty$ holds for all accessible states (l, j) with $j \geqslant J$. This implies $j \leqslant J$ for all accessible states (l, j). For $d \geqslant 2$, we have $j \leqslant \kappa =: J$ for all accessible states $(\boldsymbol{l}, j)$. Thus there are at most J consecutive recursion transitions.

To prove that only finitely many states are accessible, we introduce the notion of heights of states: The height of a state $(\boldsymbol{l}, j)$ is defined to be $\boldsymbol{h} = \boldsymbol{l}q^{-j}$. If there exists a storing transition from $(\boldsymbol{l}, j)$ of height $\boldsymbol{h}$ to $(\boldsymbol{l}', j')_F$ of height $\boldsymbol{h}'$, we have $\frac{1}{q}\boldsymbol{h} \leqslant \boldsymbol{h}' \leqslant \frac{1}{q}\boldsymbol{h}+1$. If there exists a recursion transition from $(\boldsymbol{l}, j)$ of height $\boldsymbol{h}$ to $(\boldsymbol{l}', j')_N$ of height $\boldsymbol{h}'$, we have $\boldsymbol{h}+\boldsymbol{s}^- - \mathbf{1} \leqslant \boldsymbol{h}' \leqslant \boldsymbol{h}+\boldsymbol{s}^+$ where $\boldsymbol{s}^+ = \max_\lambda\{r_\lambda q^{-\kappa_\lambda}, 0\}$ and $\boldsymbol{s}^- = \min_\lambda\{r_\lambda q^{-\kappa_\lambda}, 0\}$.

Assume that there is a path from $(\boldsymbol{l}, j)$ of height $\boldsymbol{h}$ to $(\boldsymbol{l}', j')$ of height $\boldsymbol{h}'$ with $L \leqslant J$ recursion transitions and one storing transition (in this order). Then we have

$$\frac{1}{q}\boldsymbol{h} + \frac{J}{q}(\boldsymbol{s}^- - \mathbf{1}) \leqslant \boldsymbol{h}' < \frac{1}{q}\boldsymbol{h} + \frac{J}{q}\boldsymbol{s}^+ + \mathbf{1}.$$

We can subdivide every path in the transducer starting with the initial state into a sequence of such paths and a final path consisting of only recursion transitions. Let $\boldsymbol{h}_m$ be the sequence of heights of the states where the subpaths starts. Then, we have

$$\frac{1}{q}\boldsymbol{h}_m + \frac{J}{q}(\boldsymbol{s}^- - \mathbf{1}) \leqslant \boldsymbol{h}_{m+1} < \frac{1}{q}\boldsymbol{h}_m + \frac{J}{q}\boldsymbol{s}^+ + \mathbf{1}.$$

Iteration leads to

$$\frac{J(\boldsymbol{s}^- - \mathbf{1})}{q-1} \leqslant \boldsymbol{h}_m \leqslant \frac{J\boldsymbol{s}^+ + q\mathbf{1}}{q-1}$$

for all m. Therefore, the height $\boldsymbol{h}$ of an accessible state is bounded. Since $0 \leqslant j \leqslant J$ is also bounded, the integer carry $\boldsymbol{l} = q^j \boldsymbol{h}$ of an accessible state $(\boldsymbol{l}, j)$ can only take finitely many different values. The accessible part of the transducer is thus finite. $\qquad\square$

Lemma 6.5. *Let $\mathcal{P}$ be an infinite path with input zero starting at some state of level j such that all of its states have non-negative carries. Then, after at most j transitions, it reaches a state $(\boldsymbol{l}_0, \kappa)$. From that point on, it only passes through simple states, namely*

$$(\boldsymbol{l}_0, \kappa), (\boldsymbol{l}_1, j_1)_N, (\boldsymbol{l}_1, j_1 + 1)_F, \dots, (\boldsymbol{l}_1, \kappa)_F,$$
$$(\boldsymbol{l}_2, j_2)_N, (\boldsymbol{l}_2, j_2 + 1)_F, \dots, (\boldsymbol{l}_2, \kappa)_F,$$
$$(\boldsymbol{l}_3, j_3)_N, (\boldsymbol{l}_3, j_3 + 1)_F, \dots, (\boldsymbol{l}_3, \kappa)_F,$$
$$\dots$$

where $\boldsymbol{l}_i = A(\boldsymbol{l}_{i-1})$ and $j_i = \kappa_{\boldsymbol{l}_{i-1} \bmod q^\kappa}$ for $i \geqslant 1$.

Proof. Denote the first state of $\mathcal{P}$ by $(\boldsymbol{l}, j)$.

First, assume that $j \geqslant \kappa$. As storing transitions always increase the level and the levels are bounded by Lemma 6.4, the path has to contain at least one recursion transition. Thus the path starts with $k \geqslant 0$ storing transitions leading from $(\boldsymbol{l}, j)$ to $(\boldsymbol{l}, j + k)$, followed by a recursion transition from $(\boldsymbol{l}, j + k)$ to $(\boldsymbol{l}', j')$. By assumption, we have $\boldsymbol{l} \geqslant 0$ and $\boldsymbol{l}' \geqslant 0$. Thus $A(\boldsymbol{l}) = \boldsymbol{l}' \neq \infty$ by (64). Therefore, there is a recursion transition leaving $(\boldsymbol{l}, j)$, i.e., there were no leading storing transitions. Recall that $j' < j$ holds for any recursion transition. We repeat the argument at most $j - \kappa$ times until we reach a simple state.

If we are in a simple state $(\boldsymbol{l}', j')$ with $j' < \kappa$, the next $\kappa - j'$ steps will be storing transitions, leading to $(\boldsymbol{l}', \kappa)$. This means that after at most j steps, we reach a state $(\boldsymbol{l}_0, \kappa)$.

We now apply the argument of the second paragraph again. Thus a recursion transition leads to $(\boldsymbol{l}_1, j_1)$ with $\boldsymbol{l}_1 = A(\boldsymbol{l}_0)$ and $j_1 = \kappa_{\boldsymbol{l}_0 \bmod q^\kappa}$.

The remainder of the lemma follows by induction. $\qquad\square$

As an auxiliary structure for deciding the well-posedness of the recursion, we introduce the *recursion digraph* $\mathcal{R}$. It has set of vertices $\mathbb{N}_0^d$ and arcs $(\boldsymbol{n}, A(\boldsymbol{n}))$ with label $t_{\boldsymbol{n} \bmod q^\kappa}$ for all $\boldsymbol{n} \in \mathbb{N}_0^d$ with $A(\boldsymbol{n}) < \infty$. Thus $a(\boldsymbol{n})$ can be computed from the successor of $\boldsymbol{n}$ in $\mathcal{R}$ using the recursion (10). By definition, each vertex of $\mathcal{R}$ has out-degree 1 or 0. Each component of $\mathcal{R}$ is a functional digraph or a rooted tree (oriented towards the root).

If

$$\|\boldsymbol{n}\|_\infty > \max_{\boldsymbol{\lambda}} \frac{\|\boldsymbol{\lambda}\|_\infty + \|\boldsymbol{r_\lambda}\|_\infty}{q^\kappa - q^{\kappa\lambda}},$$

we have

$$q^\kappa \|\boldsymbol{n}\|_\infty - \|\boldsymbol{\lambda}\|_\infty > q^{\kappa\lambda} \|\boldsymbol{n}\|_\infty + \|\boldsymbol{r_\lambda}\|_\infty$$

and therefore
$$\|q^\kappa \boldsymbol{n} + \boldsymbol{\lambda}\|_\infty > \|q^\kappa \boldsymbol{\lambda} + \boldsymbol{r_\lambda}\|_\infty$$
for all $0 \leqslant \boldsymbol{\lambda} < q^\kappa \mathbf{1}$. Thus we have $\|\boldsymbol{n}'\|_\infty < \|\boldsymbol{n}\|_\infty$ for all but finitely many arcs $(\boldsymbol{n}, \boldsymbol{n}')$ of $\mathcal{R}$.

Thus for every vertex of $\mathcal{R}$, there is a unique path starting in this vertex and leading to a vertex with out-degree 0 or a finite cycle.

From this description, it is clear that the recursion is well-posed if and only if

- the sum of the labels of each cycle in $\mathcal{R}$ is 0 and

- the set $\mathcal{I}$ consists of one element for every cycle in $\mathcal{R}$ as well as of the vertices with out-degree 0 in $\mathcal{R}$.

We now prove the essential connection between the recursive digraph and the transducer $\widetilde{\mathcal{T}}$. This also implies that the classes $F_1, \ldots, F_K$ are disjoint.

Lemma 6.6. *There exists a bijection between cycles in the recursive digraph $\mathcal{R}$ and accessible cycles in the transducer $\widetilde{\mathcal{T}}$ with input 0 and simple states. Corresponding cycles under this bijection have the same output sum and sum of labels.*

Proof. Let $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_L = \boldsymbol{n}_0$ be a cycle in the recursive digraph with $\boldsymbol{n}_R \geqslant 0$ for all $0 \leqslant R < L$.

Let k_0 be the length of the path $\mathcal{P}_0$ in $\widetilde{\mathcal{T}}$ starting in the initial state and reading the q-ary expansion of $\boldsymbol{n}_0$.

We determine the destinations of certain paths in the transducer associated with the cycle in the recursive digraph.

Statement 6.7. *Let $k \geqslant k_0$ and $\mathcal{P}$ be the path from the initial state $(0,0)$ to $(\boldsymbol{l}, j)$ of length k whose input label is the q-ary expansion of $\boldsymbol{n}_0$, padded with leading zeros. Assume that the number of recursion transitions in this path is $LQ + R$ for some $Q \geqslant 0$ and $0 \leqslant R < L$. Then $\boldsymbol{l} = \boldsymbol{n}_R \geqslant 0$.*

Proof of Statement 6.7. Let $k' = k - (LQ + R)$ be the number of storing transitions of $\mathcal{P}$. By (66), we have
$$A^{LQ+R}(q^{k'} \boldsymbol{n} + \boldsymbol{n}_0) = q^j \boldsymbol{n} + \boldsymbol{l} \tag{68}$$
for $\boldsymbol{n} \geqslant 0$, $\boldsymbol{n} \neq 0$.

Note that for $M \geqslant \kappa$ and $\boldsymbol{n} \equiv \boldsymbol{n}' \pmod{q^M}$ with $A(\boldsymbol{n}) < \infty$ and $A(\boldsymbol{n}') < \infty$, the definition of A implies $A(\boldsymbol{n}) \equiv A(\boldsymbol{n}') \pmod{q^{M-\kappa}}$.

Together with the definitions of $\boldsymbol{n}_R$ and the recursive digraph $\mathcal{R}$ as well as (68), this implies
$$\begin{aligned}
\boldsymbol{n}_R = A^{LQ+R}(\boldsymbol{n}_0) &\equiv A^{LQ+R}(q^{k'+M}\mathbf{1} + \boldsymbol{n}_0) \\
&= q^{j+M}\mathbf{1} + \boldsymbol{l} \pmod{q^{k'+M-(LQ+R)\kappa}}
\end{aligned}$$
for sufficiently large M. Coarsening yields
$$\boldsymbol{n}_R \equiv \boldsymbol{l} \pmod{q^{M-(LQ+R)\kappa}},$$
still valid for sufficiently large M. As $\boldsymbol{l}$ is bounded by Lemma 6.4, this implies $\boldsymbol{n}_R = \boldsymbol{l}$. $\square$

Now, we conclude the proof of Lemma 6.6.

Let $\mathcal{P}$ be the infinite path in $\widetilde{\mathcal{T}}$ starting at the destination of $\mathcal{P}_0$ and reading zeros. By Lemma 6.5 applied to $\mathcal{P}$ together with Statement 6.7 applied to $\mathcal{P}_0$ concatenated with prefixes of $\mathcal{P}$, $\mathcal{P}$ leads to a cycle in $\widetilde{\mathcal{T}}$. Its states are simple and have carries $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{L-1}$ and levels determined by $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{L-1}$ as in Lemma 6.5.

This construction defines a map from the cycles of the recursive digraph $\mathcal{R}$ to the accessible cycles with input 0 in the transducer with simple states. This map is injective by construction. Under this map, the sum of the labels of the cycle in $\mathcal{R}$ equals the sum of output labels of the cycle in $\widetilde{\mathcal{T}}$ by construction.

On the other hand, let

$$
\begin{aligned}
&(\boldsymbol{n}_0, j_0), (\boldsymbol{n}_0, j_0 + 1), \ldots, (\boldsymbol{n}_0, \kappa), \\
&(\boldsymbol{n}_1, j_1), (\boldsymbol{n}_1, j_1 + 1), \ldots, (\boldsymbol{n}_1, \kappa), \ldots \\
&(\boldsymbol{n}_{L-1}, j_{L-1}), (\boldsymbol{n}_{L-1}, j_{L-1} + 1), \ldots, (\boldsymbol{n}_{L-1}, \kappa), \\
&(\boldsymbol{n}_0, j_0)
\end{aligned}
$$

be an accessible cycle of simple states in the transducer with input 0. Lemma 6.5 yields $A(\boldsymbol{n}_R) = \boldsymbol{n}_{R+1 \bmod L} \geqslant 0$ for $0 \leqslant R < L$. Thus, this cycle in the transducer is the image of the cycle $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_L = \boldsymbol{n}_0$ in the recursive digraph. Thus the map is surjective. $\qquad\square$

To use Theorem 1, we simplify $\widetilde{\mathcal{T}}$ to obtain the deterministic transducer $\mathcal{T}$, that is one without transitions with empty input. As a first step, we remove all non-accessible states. By Lemma 6.4, this leaves us with finitely many states.

By Lemma 6.4 and the fact that recursion transitions decrease the level, the length of paths consisting of recursion transitions only is bounded. As a recursion transition always leads to a non-final state, processing an input never ends with a recursion transition.

Consider a recursion transition from $(\boldsymbol{l}, j)$ to $(\boldsymbol{l}', j')_N$ with output t such that no recursion transition originates in $(\boldsymbol{l}', j')_N$. For each transition originating in $(\boldsymbol{l}', j')_N$, say to some $(\boldsymbol{l}'', j'')_F$ with input $\boldsymbol{\varepsilon}$ and output t', we insert a storing transition from $(\boldsymbol{l}, j)$ to $(\boldsymbol{l}'', j'')_F$ with input $\boldsymbol{\varepsilon}$ and output $t + t'$. Then, the recursion transition from $(\boldsymbol{l}, j)$ to $(\boldsymbol{l}', j')_N$ is removed. The number of recursion transitions decreased by one and the new transducer generates the same output as the old transducer. We repeat this process until there are no more recursion transitions. Then, all non-final states are inaccessible and are removed.

Proof of Theorem 4. By Lemma 6.6 and the characterization of well-posedness via the recursive digraph, the recursion (10) is well-posed if and only if $\mathcal{I}$ consists of exactly one representative of each of the sets F_j, $1 \leqslant j \leqslant K$, and if $\widetilde{\mathcal{T}}$ has no cycle with simple states, input 0 and non-vanishing output sum.

We now show that the cycles of simple states with input 0 in $\mathcal{T}$ are exactly the reductions of the cycles of simple states with input 0 in $\widetilde{\mathcal{T}}$. As a cycle with simple states and input 0 in $\widetilde{\mathcal{T}}$ does not have consecutive recursion transitions (cf. Lemma 6.5), it is reduced to a cycle with simple states in $\mathcal{T}$. On the other hand, consider a cycle of $\widetilde{\mathcal{T}}$ with input 0 containing a non-simple state. If there is a state of level $> \kappa$, the state with

largest level is final and is not removed. If all states have level $\leqslant \kappa$, then there are no two consecutive recursion transitions, so no negative carry is completely removed from the cycle in the reduction to $\mathcal{T}$. Therefore, such a cycle is not reduced to a cycle with simple states and input 0 in $\mathcal{T}$.

Therefore, the assertion on well-posedness is proved.

To prove correctness of the transducer, we use (66) with $(\boldsymbol{l}, j) = (0, 0)$, the joint q-ary expansion of $\boldsymbol{n}$ as input leading to some state $(\boldsymbol{l}', j')_F$ with output $\delta_{m'-1} \ldots \delta_0$. By Lemma 6.2, we have $\boldsymbol{l}' \geqslant 0$ because the last transition is a storing transition with non-zero input. Thus by (67), $a(\boldsymbol{n}) = a(\boldsymbol{l}') + \sum_{k=0}^{m'-1} \delta_k$. As the final output of $(\boldsymbol{l}', j')_F$ is defined to be $a(\boldsymbol{l}')$, we obtain $\mathcal{T}(\boldsymbol{n}) = a(\boldsymbol{l}') + \sum_{k=0}^{m'-1} \delta_k = a(\boldsymbol{n})$, as requested. $\qquad \square$

References

[1] Jean-Paul Allouche and Jeffrey Shallit, *Automatic sequences: Theory, applications, generalizations*, Cambridge University Press, Cambridge, 2003.

[2] Tom Apostol, *Modular functions and Dirichlet series in number theory*, Graduate Texts in Mathematics, vol. 41, Springer, New York, 1976.

[3] Guy Barat and Peter J. Grabner, *Distribution of binomial coefficients and digital functions*, J. London Math. Soc. (2) **64** (2001), no. 3, 523–547.

[4] Nader L. Bassily and Imre Kátai, *Distribution of the values of q-additive functions on polynomial sequences*, Acta Math. Hungar. **68** (1995), no. 4, 353–361.

[5] Valérie Berthé and Michel Rigo (eds.), *Combinatorics, automata and number theory*, Encyclopedia Math. Appl., vol. 135, Cambridge University Press, Cambridge, 2010.

[6] Emmanuel Cateland, *Suites digitales et suites k-régulières*, Ph.D. thesis, Université Bordeaux, 1992.

[7] Hubert Delange, *Sur la fonction sommatoire de la fonction "somme des chiffres"*, Enseignement Math. (2) **21** (1975), 31–47.

[8] *NIST Digital library of mathematical functions*, http://dlmf.nist.gov/, Release 1.0.9 of 2014-08-29, 2010, Online companion to [29].

[9] Michael Drmota and Peter J. Grabner, *Analysis of digital functions and applications*, Combinatorics, automata and number theory (Valérie Berthé and Michel Rigo, eds.), Encyclopedia Math. Appl., vol. 135, Cambridge University Press, Cambridge, 2010, pp. 452–504.

[10] Philippe Dumas, *Joint spectral radius, dilation equations, and asymptotic behavior of radix-rational sequences*, Linear Algebra Appl. **438** (2013), no. 5, 2107–2126.

[11] Philippe Dumas, *Asymptotic expansions for linear homogeneous divide-and-conquer recurrences: Algebraic and analytic approaches collated*, Theoret. Comput. Sci. **548** (2014), 25–53.

[12] Philippe Flajolet, Peter Grabner, Peter Kirschenhofer, Helmut Prodinger, and Robert F. Tichy, *Mellin transforms and asymptotics: digital sums*, Theoret. Comput. Sci. **123** (1994), 291–314.

[13] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[14] Chris D. Godsil and Gordon Royle, *Algebraic graph theory*, Graduate texts in mathematics, vol. 207, Springer Verlag (New York), 2001.

[15] Peter J. Grabner, Clemens Heuberger, and Helmut Prodinger, *Subblock occurrences in signed digit representations*, Glasg. Math. J. **45** (2003), 427–440.

[16] _____, *Distribution results for low-weight binary representations for pairs of integers*, Theoret. Comput. Sci. **319** (2004), 307–331.

[17] Peter J. Grabner and Hsien-Kuei Hwang, *Digital sums and divide-and-conquer recurrences: Fourier expansions and absolute convergence*, Constr. Approx. **21** (2005), 149–179.

[18] Peter J. Grabner and Jörg M. Thuswaldner, *On the sum of digits function for number systems with negative bases*, Ramanujan J. **4** (2000), no. 2, 201–220.

[19] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics. A foundation for computer science*, second ed., Addison-Wesley, 1994.

[20] Clemens Heuberger, Daniel Krenn, and Sara Kropf, *Automata and transducers in the computer algebra system Sage*, 2014, arXiv:1404.7458 [cs.FL].

[21] Clemens Heuberger and Sara Kropf, *Analysis of the binary asymmetric joint sparse form*, Combin. Probab. Comput. **23** (2014), 1087–1113.

[22] Clemens Heuberger, Sara Kropf, and Helmut Prodinger, *Asymptotic analysis of the sum of the output of transducers*, 25th International Conference on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'14), DMTCS-HAL Proceedings Series, vol. BA, 2014, pp. 145–156.

[23] Clemens Heuberger and James A. Muir, *Minimal weight and colexicographically minimal integer representations*, J. Math. Cryptol. **1** (2007), 297–328.

[24] Hsien-Kuei Hwang, *On convergence rates in the central limit theorems for combinatorial structures*, European J. Combin. **19** (1998), 329–343.

[25] Tosio Kato, *Perturbation theory for linear operators*, Springer, 1976.

[26] Peter Kirschenhofer, *Subblock occurrences in the q-ary representation of n*, SIAM J. Algebraic Discrete Methods **4** (1983), no. 2, 231–236.

[27] Peter Kirschenhofer and Helmut Prodinger, *Subblock occurrences in positional number systems and Gray code representation*, J. Inform. Optim. Sci. **5** (1984), no. 1, 29–42.

[28] Blake Madill and Narad Rampersad, *The abelian complexity of the paperfolding word*, Discrete Math. **313** (2013), no. 7, 831–838.

[29] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark (eds.), *NIST Handbook of mathematical functions*, Cambridge University Press, New York, 2010.

[30] Manfred Peter, *The asymptotic distribution of elements in automatic sequences*, Theoret. Comput. Sci. **301** (2003), 285–312.

[31] Marcel-Paul Schützenberger, *Sur une variante des fonctions sequentielles*, Theoret. Comput. Sci. **4** (1977), no. 1, 47–57.

[32] William A. Stein et al., *Sage Mathematics Software (Version 6.4.1)*, The Sage Development Team, 2014, http://www.sagemath.org.

[33] Gérard Tenenbaum, *Sur la non-dérivabilité de fonctions périodiques associées à certaines formules sommatoires*, The mathematics of Paul Erdős, I (Ronald L. Graham and Jaroslav Nešetřil, eds.), Algorithms Combin., vol. 13, Springer, Berlin, 1997, pp. 117–128.

[34] Jörg M. Thuswaldner, *Summatory functions of digital sums occurring in cryptography*, Period. Math. Hungar. **38** (1999), no. 1-2, 111–130.

[35] Edmund T. Whittaker and George N. Watson, *A course of modern analysis*, Cambridge University Press, Cambridge, 1963, Reprint of the fourth (1927) edition.

[36] Antoni Zygmund, *Trigonometric series*, vol. I & II combined, Cambridge University Press, Cambridge, 2002.