Abstract

Among all sequences that satisfy a divide-and-conquer recurrence, the sequences that are rational with respect to a numeration system are certainly the most immediate and most essential. Nevertheless, until recently they have not been studied from the asymptotic standpoint. We show how a mechanical process permits to compute their asymptotic expansion. It is based on linear algebra, with Jordan normal form, joint spectral radius, and dilation equations. The method is compared with the analytic number theory approach, based on Dirichlet series and residues, and new ways to compute the Fourier series of the periodic functions involved in the expansion are developed. The article comes with an extended bibliography.

Keywords: divide-and-conquer recurrence, radix-rational sequence, spectral radius, dilation equation, cascade algorithm, Dirichlet series, Fourier series

2010 MSC: 11A63, 41A60

The aim of this article is the asymptotic study of a special type of divide-and-conquer sequences, namely the sequences which are rational with respect to a numeration system or radix-rational sequences. In most cases where the numeration system is the binary system, such sequences satisfies a recursion which essentially links the values $u_{2n}$ and $u_{2n+1}$ to the value of $u_n$ by a linear homogeneous relationship. Section 1 is devoted to their precise definition. In Sections 2 to 5 we provide the reader with a method of easy use to determine the asymptotic behaviour of radix-rational sequences. The used tools all come from linear algebra and this is the distinctive feature of this article. We insist on ideas and proofs are only sketched. A formal and complete treatment of this part can be found in [1]. It results in Theorem 1 which asserts that a radix-rational sequence admits an asymptotic expansion with variable coefficients in the scale $N^\alpha \log^m N$, or equivalently $N^\alpha \log^m N$. The so-called dichopile algorithm, briefly reminded in Example 1 below, will be a thread in this paper and its asymptotic behaviour is given by

$$f_N \sim \frac{1}{2} N \log_2 N + N \Phi(\log_2 N) + O(N^\epsilon),$$

where $\Phi(t)$ is a 1-periodic function and $0 < \epsilon < 1$. We present in Section 6 some improvements of the method and we summarize it into an algorithm. Next in Section 7 we briefly mention the previous works about the asymptotic behaviour of divide-and-conquer sequences. Specifically, we show how our algebraic approach can help the analytic number theory approach, popularized by Philippe Flajolet. To end, in Section 8 we elaborate on the computation of the Fourier coefficients of the involved periodic functions and we provide the reader with several methods. This leads us to consider a Mellin transform.
1. Radix-rational sequences

Radix-rational sequences are a mere generalization of classical rational sequences, that is sequences which satisfy a linear homogeneous recurrence with constant coefficients. In the same manner they satisfy a linear homogeneous recurrence with constant coefficients but where the shift $n \mapsto n + 1$ is replaced by the pair of scaling transformations $n \mapsto 2n$ and $n \mapsto 2n + 1$, in case the radix is 2.

These sequences come to light in various domains of knowledge and the first example which comes in mind is the binary sum-of-digits function, that is the number of 1’s in the binary expansion of an integer $n$. The sequence satisfies $s_0 = 0$, $s_{2n} = s_n$, $s_{2n+1} = s_n + 1$. It may seem to the reader that such a sequence is of limited interest, but it appears in many problems like the study of the maximum of a determinant of a $n \times n$ matrix with entries $\pm 1$ or in a merging process occurring in graph theory. This example have been greatly generalized with the number of occurrences of some pattern in the binary code, or in the Gray code: Flajolet and Ramshaw study the average case of Batcher’s odd-even merge by using the sum-of-digits function of the Gray code. Among the sequences directly related to a numeration system the Thue-Morse sequence which writes $u_n = (-1)^{s_n}$ is certainly the one which has caused the greatest number of publications. There exist variants with some subsequences or with binary patterns other than the simple pattern 1: the Rudin-Shapiro sequence associated to the pattern satisfies $u_{2n} = u_{2n+1} = 1$, $u_{2n+1} = -u_{2n+1}$, $u_0 = u_1 = 1$, and was initially designed to minimize the $L^\infty$ norm of a sequence of trigonometric polynomials with coefficients $\pm 1$. The study of the complexity of algorithms is another source of radix-rational sequences. The cost $c_\ell$ of computing the $n$-th power of a matrix by binary powering satisfies $c_0 = 0$, $c_2 = c_n + 1$, $c_{2n+1} = c_n + 2$. The idea of binary powering has been re-employed by Morain and Olivos in the context of computations on an elliptic curve where the subtraction has the same cost than addition. Supowit and Reingold have used the divide-and-conquer strategy to give heuristics for the problem of Euclidean matching and this leads them to a 4-rational sequence in the worst case. The number of bases is another domain which provides examples like the number of odd binomial coefficients in row $n$ of Pascal’s triangle or the number of integers which are a sum of three squares in the first $n$ integers.

It is the merit of Allouche and Shallit to have put all these scattered examples in a common framework, that leads to the idea of a linear representation with matrices. For the examples above the matrix dimension $d$ is usually small, say from 1 to 4, but there exist examples where $d$ is larger, like in the work of Cassaigne which uses $d = 30$. In such cases a general method of study is necessary and we will elaborate one. But, it is first necessary to formalize the notion of a radix-rational sequence.

1.1. Forward direction

Let $S$ be a rational formal series over a finite alphabet $X$. For the sake of simplicity, we are assuming here and in all the sequel that the coefficients of the formal series are complex numbers, even if it is possible to consider in this first section the more general framework of a commutative field or even of a commutative semi-ring. If $X$ is the set of figures for the numeration system with radix $B$, the rational series defines a sequence $s_n$ in the following manner: for each non-negative integer $n$, we consider the word $w = n_\ell \ldots n_1 n_0$ which is the radix $B$ expansion of $n$, that is

$$n = (n_\ell \ldots n_1 n_0)_B = n_0 + n_1 B + \cdots + n_\ell -1 B^{\ell -1},$$

and the term $s_n$ of the sequence is the value $(S, w)$ of the rational series over the word $w$.

More concretely, if a linear representation of the rational series $S$ is at our disposal, we can compute the value $(S, w)$ for a word $w = w_1 \ldots w_\ell$ by the formula

$$(S, w) = LA_{w_1} A_{w_2} \cdots A_{w_\ell} C = LA_w C. \quad (2)$$

Here, $(L, (A_b)_{0 \leq b < B}, C)$ is a linear representation of $S$. This means that $L$ is a row vector, matrices $A_b$ are square matrices, and $C$ is a column vector, all with a coherent size $d$, which is the dimension of the representation. Formula (2) translates immediately into

$$s_n = LA_{n_{\ell-1}} \cdots A_1 A_0 C. \quad (3)$$
and gives a way to compute the successive values of the sequence $s$ associated with the rational series $S$.

The rational character of the formal series $S$ has the following meaning [21, 22]: there is a vector space which contains the series $S$; this vector space is of finite dimension $d$; and this vector space is left stable by the trimming operators $T_r$, with $0 \leq r < B$, defined by (the figure $r$ is viewed as a letter)

$$T_r S = \sum_{w=wr} (S, w) v.$$

The operator $T_r$ extracts the part of the formal series associated with the words which end with $r$ and trims the letter $r$. These operators translate immediately, via the numeration system, into the multisection operators $T_r s_n = s_{Bn+r}$.

The correspondence is more clear if we use the ordinary generating function of the sequence $s_n$, which parallels the formal series $S$.

$$T_r \sum_{n=0}^{+\infty} s_n z^n = \sum_{n=0}^{+\infty} s_{Bn+r} z^n \quad (4)$$

With this viewpoint, the linear representation of the formal series $S$ is interpreted as follows for the sequence $s_n$: we have a finite-dimensional vector space $V$ of sequences, equipped with a basis $(v^1, v^2, \ldots, v^d)$ (or more generally a generating system), which is left stable by the multisection operators and contains the sequence $s$. The column vector $C$ gives the coordinates of the sequence $s$ with respect to the basis $(v^i)_{1 \leq i \leq d}$; the matrix $A_k$ is the expression of the multisection operator $T_k$; and the row vector $L$ appears to be the linear form $u \mapsto s_{k0}$, which evaluates the sequences for $n = 0$. As an example, with $B = 2$, $n = 5 = (101)_2$, the column vectors $C_1, A_1 C, A_0 A_1 C$, and $A_1 A_0 A_1 C$ are the expressions of the sequences $s_n$, $s_{2n+1}$, $s_{4n+1}$, and $s_{8n+5}$, respectively. Taking the value for $n = 0$ of the last sequence provides the value $s_5$ and this explains the formula $s_5 = LA_1 A_0 A_1 C$.

We are led to the following definition, but with a slight change with respect to [18, 19]. We prefer the attributive adjective rational to regular, because rational is understood by both computer scientists and mathematicians, while regular is understood only by the former but remains hazy for the latter.

**Definition 1.** A (complex) sequence is said to be rational with respect to a numeration system, or more briefly radix-rational, or more precisely $B$-rational if its orbit under the action of the $B$-multisection operators remains in a finite dimensional vector space.

**1.2. Reverse direction**

Let us assume now that we have a sequence $s = (s_n)$ which is $B$-rational. We can build a linear representation of this sequence in the following way. If the sequence is the null sequence there is nothing to do, and we consider that the associated dimension $d$ is 0. Otherwise we put $v^1 = s$ as the first vector of the basis we want to build. Next we consider successively the sequences $(s_{B0})$, $(s_{Bn+1})$, ..., $(s_{Bn+B-1})$, $(s_{2n})$, $(s_{2n+1})$, and so on. For each of these sequences, either it is independent of the previous ones and we add it to the basis, say as $v^k$, or it can be expressed as a linear combination of the previous sequences that are in the basis. The process necessarily stops since all these sequences live in a finite dimensional space. When it stops, we have a free family $(v^1, v^2, \ldots, v^d)$ which is a basis of a vector space $V$. This vector space is left stable by the multisection operators and contains the sequence $s$. Moreover we have a linear representation of the sequence. First the sequence $s$ is the first vector of the basis, and this gives the vector column $C = E_1$, say. Next we have noted in passing the dependencies encountered in the process, between each of the sequence $v^j$ and its images by the multisection operators $T_0 v^j$, $T_1 v^j$, ..., $T_{B-1} v^j$, and this gives us the matrices $A_0, A_1, \ldots, A_{B-1}$. And last the values $v^0_0, \ldots, v^0_d$ at 0 give the row vector $L$.

By this process, we have associated with the $B$-rational sequence a rational formal series over the alphabet of the radix $B$ numeration system. Moreover, we see that the classical rational sequences, like the Fibonacci sequence, that is the sequences whose ordinary generating function is a rational function (for which 0 is not a pole), appear to be 1-rational because it is well known that such a sequence writes $s_n = LA^n C$ for some square matrix $A$. Hence, the $B$-rational sequences generalize the classical rational sequences.
1.3. Sensitivity to the leftmost zeroes

The above construction is not entirely satisfactory, for if a rational series determines a radix rational sequence, the converse is not true. The reason is that the sequence does not use the words which begin with some zeroes. (The \( B \)-ary word associated with the number 0 is the empty word.) A first idea to fill this gap is to extend the formal series by the value 0 for words that begin with zero. A more preferable method is to follow the process of the previous paragraph. It provides sequences \( v^j \) which satisfy obviously \( (T_0v^j)_0 = v^j_0 \), that is, in one formula, \( LA_0 = L \). This property determines completely the formal series if its values for the \( B \)-ary expansions of the integers are known.

**Definition 2.** A linear representation \((L, (A_b)_{0 \leq b < B}, C)\) of a \( B \)-rational sequence is said to be insensitive to the leftmost zeroes, or simply zero-insensitive, if it satisfies \( LA_0 = L \).

This definition implies two remarks. First, a radix-rational sequence admits always a linear representation which is insensitive to the leftmost zeroes. Above, we have proved this property. Second, because all the minimal linear representations of a rational formal series are isomorphic, the same property is satisfied for the radix-rational sequence if we use only zero-insensitive linear representations.

**Example 1 (Dichopile algorithm).** The dichopile algorithm is an attempt to find a balance between space and time in the random generation of words from a regular language \([23, 24]\) (the recursion (5) below is rather hidden in the last reference). To generate uniformly a word of length \( n \) the previously known methods use \( O(n) \) in space and \( O(n) \) in time, while the dichopile algorithm uses only \( O(\log n) \) in space (a big saving) but \( O(n \log n) \) in time (a small loss), as we will see. The time complexity is given by the sequence \( f_n \) defined as follows,

\[
f_n = n + g_n, \quad g_n = f_{\lceil n/2 \rceil} - 1 + g_{\lfloor n/2 \rfloor}, \quad \text{for} \quad n \geq 2,
\]

with \( f_1 = 1, \ g_1 = 0 \). We add \( f_0 = 0, \ g_0 = 0 \). Both sequences \( f_n \) and \( g_n \) are rational with respect to the radix \( B = 2 \).

As it will soon appear, we are not interested by a linear representation for the sequence \( f_n \) but for the sequence of backward differences \( u_n = \nabla f_n = f_n - f_{n-1} \) (with \( f_{-1} = 0 \)), which is 2-rational too according to Lemma 1 below. A linear representation of dimension 6 is the following

\[
L = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T, \quad A_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \end{pmatrix}.
\]

(6)

It uses the generating family \( u_n = \nabla f_n = 1 + g_n, \ \nabla f_{n-1} = 1 + \nabla g_{n-1}, \ \nabla f_{n+1} = \Delta g_n, \ \nabla g_n, \ 1, \) and \( 1 - \delta_{n,1} \) (here \( \delta_{n,m} \) is the usual Kronecker symbol). As a consequence it is not a reduced linear representation, because the first sequence is the sum of the fourth and the fifth. Nevertheless it is perfectly acceptable and, moreover, it is insensitive to the leftmost zeroes.

\( \triangle \)

In the rest of the article, we will use only zero-insensitive linear representations.

1.4. Backward differences

As it is pointed out in Example 1 we do not use a linear representation of the sequence \( s_n \), we want to estimate asymptotically, but a linear representation of the sequence \( u_n = \nabla s_n = s_n - s_{n-1} \) of its backward differences. The next assertion shows that the latter is radix-rational if the former is.

**Lemma 1.** If a sequence \( s_n \) is \( B \)-rational, then the sequence of its backward differences is \( B \)-rational too. Conversely, if a sequence \( u_n \) is \( B \)-rational, then the sequence of its partial sums \( s_n \) is \( B \)-rational too.
at last.

This leads us to the formula as follows

\[ \sum B_n x^n, \quad s(x) = \sum s_n x^n, \]

which are related by

\[ u(x) = (1-x)s(x), \quad s(x) = \frac{1}{1-x} u(x). \]

Allouche and Shallit have proved that the product of two \( B \)-rational generating functions is \( B \)-rational, hence the title *The ring of \( k \)-regular sequences* of their article [13]. Moreover \( 1-x \) is \( B \)-rational as a polynomial, and \( 1/(1-x) \) is \( B \)-rational as a rational function whose poles are all roots of unity [13 Th. 3.3]. Accordingly, if \( s(x) \) is \( B \)-rational, so is \( u(x) \), and conversely.

\[ \square \]

2. Basic ideas

Our aim is to expose a calculation method for an asymptotic expansion of a radix rational sequence \( s_n \). It takes as input a linear representation of the sequence of backward differences \( u_n = \nabla s_n = s_n - s_{n-1} \) (with \( s_{-1} = 0 \)). Its output is an asymptotic expansion for the sequence \( s_n \), which writes as a finite linear combination of terms

\[ n^\alpha \left( \frac{\log B}{m} \right) \Psi(\log_B n). \]  \hfill (7)

Here \( \alpha \) is a real number, \( B \) is the radix of the numeration system, \( m \) is a nonnegative integer, and \( \Psi(t) \) is an oscillatory function often \( 1 \)-periodic in practice.

In the rest of the article, we will constantly use the following notations.

**Notations.** Throughout, \( L \), \( (A_b)_{0 \leq b < B} \), \( C \) is a zero-insensitive linear representation for \( u_n = \nabla s_n \), which is \( d \)-dimensional. The matrix \( I_d \) is the identity matrix of size \( d \). The matrix \( Q \) is the sum of the square matrices in the representation,

\[ Q = \sum_{0 \leq b < B} A_b. \]  \hfill (8)

The integer \( K \) is the integer part of the logarithm base \( B \) of \( N \), and \( t \) is its fractional part,

\[ K = \lfloor \log_B N \rfloor, \quad t = \log_B N - K = \{ \log_B N \}. \]  \hfill (9)

At last \( (0.w)_B \) is the number, in \( [0, 1) \), whose \( B \)-ary expansion reads \( 0.w \).

The main idea is to link the partial sum of \( u_n \) up to \( N \), that is \( s_N \), with the sum of the rational series \( S \) behind \( u_n \) for some words of a given length.

**Lemma 2.** Both sums

\[ s_N = \sum_{n=0}^{N} u_n, \quad \text{and} \quad S_K(x) = \sum_{|w|=K} (S, w), \]  \hfill (10)

are related by

\[ s_N = S_{K+1}(B^{d-1}). \]  \hfill (11)

**Proof.** Formula (11) comes from cutting the interval \([0, N]\) by the powers of \( B \). The integers in the interval \([B^k, B^{k+1})\) have a \( B \)-ary expansion which is a length \( k+1 \) word. The sum over all length \( k+1 \) words expresses with the matrix \( Q \), as follows

\[ \sum_{|w|=k+1} LA_w C = \sum_{|w|=k+1} LA_{w_1} \cdots A_{w_{k+1}} C = L(A_0 + \cdots + A_{B-1})^{k+1} C = LQ^{k+1} C. \]

This leads us to the formula

\[ s_N = L(I_d - A_0) \sum_{k=0}^{K} Q^k C + S_{K+1}(B^{d-1}). \]  \hfill (12)
More precisely, the $B$-ary expansions of the integers in $[B^k, B^{k+1})$ do not begin with a 0, hence we have to subtract $LQ^k C$ from $LQ^{k+1} C$. This explains the first term in (12). The second term $S_{K+1}(B^{t-1})$ corresponds to the interval $[B^K, N]$. With the notations (9), the integer $N$ writes $N = B^{K+t} = B^{K+1}B^{t-1}$ and for a length $K + 1$ word $w$ the inequality $(w)_B \leq N$ is equivalent to $(0.w)_B \leq B^{t-1}$, hence the occurrence of the term $S_{K+1}(B^{t-1})$. At this point, we take advantage of our assumption that we consider only zero-insensitive linear representations and Formula (12) simplifies into (11).

It will turn out that it is not sufficient to consider the scalar-valued sum $S_K(x)$, and we reinforce the notations with a vector-valued sum.

**Notations.** The sum $S_K(x)$ writes $L \Sigma_K(x)$ where $\Sigma_K(x)$ is the vector-valued sum

$$\Sigma_K(x) = \sum_{|w|=K} A_w x$$  

(13)

We are interested by the behaviour of $\Sigma_K(x)$ for $K$ large. Our starting point is the next formula.

**Lemma 3.** The sum $\Sigma_K(x)$ satisfies the recursion

$$\Sigma_{K+1}(x) = \sum_{b_1 < x_1} A_{b_1} Q^{K} x_1 + A_{x_1} \Sigma_K(Bx - x_1),$$

valid for each $x = (0.x_1 x_2 \ldots)_B$ in $[0, 1)$.

*Proof.* The numbers from $[0, x]$ with a $B$-mantissa of length $K + 1$ are sorted according to the prefixes of their $B$-ary expansions. This emphasizes the intervals $[0, (0.x_1)_B], [(0.x_1)_B, (0.x_1 x_2)_B], \ldots, [(0.x_1 x_2 \ldots x_K)_B, (0.x_1 x_2 \ldots x_{K+1})_B]$, and leads to the formula

$$\Sigma_{K+1}(x) = \sum_{b_1 < x_1} A_{b_1} Q^{K} x_1 + \sum_{b_2 < x_2} A_{x_1} A_{b_2} Q^{K-1} x_1 + \cdots + \sum_{b_{K+1} \leq x_{K+1}} A_{x_1} A_{x_2} \cdots A_{b_{K+1}} C.$$

The above recursive Formula (14) is a mere consequence.

We will decompose the column vector $C$ of the linear representation on a basis of generalized eigenvectors for the matrix $Q$, because for such vectors Formula (14) will reveal a functional equation, known as a dilation equation. But this will be possible only for eigenvalues sufficiently large, and we will begin by what means ‘sufficiently large’ in the context.

3. **Joint spectral radius**

The computing of a rational series uses products of square matrices of an arbitrary length. We need to evaluate the size of such products. We consider all products $A_w$ for words $w$ of a given length $T$ and their norms, for some subordinate norm. Then the joint spectral radius of the set of matrices $A_b$, $0 \leq b < B$, is defined as [25, 26]

$$\rho_s = \lim_{T \to +\infty} \max_{|w|=T} \| A_w \|^{1/T}.$$  

(15)

It is independent of the used subordinate norm. Moreover the maximum in the right hand side term is a non increasing function of $T$. More information can be found in [27] and the papers which come with it.

We apply immediately this definition to the study of the error term in the asymptotic expansion.

**Lemma 4.** Let $V$ be a generalized eigenvector associated with the eigenvalue $\rho_2$, $\rho \geq 0$, $|w| = 1$, which satisfies $\rho \leq \rho_s$. Then for every $r > \rho_s$, the sequence $\Sigma_K$ associated with $V$ satisfies $\| \Sigma_K \| = O(r^K)$.
Sketch of the proof. Let us consider simply the case of an eigenvector and assume \( \|A_b\| \leq r \) for \( 1 \leq b < B \). With Formula (14) of Lemma 3, we readily obtain
\[
\|\Sigma_{K+1}(x)\| \leq \rho^K \sum_{b_1 < x_1} \|A_{b_1}\| \|V\| + \|A_{x_1}\| \|\Sigma_K(Bx - x_1)\|
\]
and, using the supremum norm,
\[
\|\Sigma_{K+1}\|_\infty \leq \rho^K Br \|V\| + r \|\Sigma_K\|_\infty.
\]
With \( \rho < r \), this recursion leads to \( \|\Sigma_K\|_\infty = O(r^K) \). More generally, we can find an integer \( T \) such that for all words \( w \) of length \( T \) we have \( \|A_w\| \leq r \) and we can deal with the subsequences \( \|\Sigma_{KT+r}(x)\|, 0 \leq r < T \), as previously.

Example 2 (Dichopile algorithm, continued). Here, we are using the maximum absolute column sum \( \|\|_1 \). Because \( A_0 \) and \( A_1 \) have nonnegative coefficients, computing \( \|A_w\|_1 \) amounts to compute the row vector \( UA_w \) and take the largest component, where \( U \) is the row vector with all components equal to 1. It turns out that all these products \( UA_w \) write
\[
W = (a \ b \ c \ d \ 1 \ 1)
\]
because this is true for the empty word, and this writing is preserved in the product by \( A_0 \) and \( A_1 \).
\[
WA_0 = (1+b \ 1+d \ c \ b \ 1 \ 1), \ WA_1 = (1+c \ 1+b \ a \ c \ 1 \ 1)
\]
The formulæ above show that all the components for a word of length \( T \) are bounded by \( T+1 \). Moreover if \( UA_1^T \) (here \( T \) is an exponent) writes \([a, b, c, d, 1, 1]\) where \( b \) is larger than all the other components, then \( UA_1^{T+1} \) writes \([1+c, 1+b, a, c, 1, 1]\) and \( 1+b \) is the largest component, so that \( \|UA_1^T\| = T+1 \). We obtain
\[
\max_{|w|=T} \|A_w\|_1 = T+1.
\]
As a consequence we conclude \( \rho_* = 1 \), and the error term will be \( O(r^K) \) for every \( r > 1 \).

4. Dilation equations

Assume for a while that we take for \( C \) an eigenvector \( V \) of \( Q \) associated with the eigenvalue \( \rho \omega \), with \( \rho > 0 \) and \( |\omega| = 1 \). With \( F_K(x) = \Sigma_K(x)/(\rho \omega)^K \), Formula (14) of Lemma 3 becomes
\[
\rho \omega F_{K+1}(x) = \sum_{b_1 < x_1} A_{b_1} V + A_{x_1} F_K(Bx - x_1).
\]
If the sequence \( F_K(x) \) has a limit \( F(x) \), the previous equation gives in the limit
\[
\rho \omega F(x) = \sum_{b_1 < x_1} A_{b_1} V + A_{x_1} F(Bx - x_1).
\]
According to Definition (13), we must have \( F(1) = V \), because of \( \Sigma_K(1) = \sum_{|w|=K} A_w V = Q^K V = (\rho \omega)^K V \).

More generally, if \( Q \) admits a generalized eigenvector associated with the eigenvalue \( \rho \omega \), there exists a free family \( V = (V(j))_{0 \leq j < \nu} \) of vectors in \( \mathbb{C}^d \), such that the subspace generated by \( V \) is left stable by \( Q \) and the matrix induced with respect to the basis \( V \) of this subspace is the usual Jordan cell with size \( \nu \)
\[
J = J_{\nu, \rho \omega} = \begin{pmatrix}
\rho \omega & 1 \\
\rho \omega & 1 \\
& \ddots & \ddots \\
& & \ddots & 1 \\
& & & \rho \omega
\end{pmatrix}.
\]
The same idea as in the case of an eigenvector leads us to consider the functional equation
\[
F(x)J = \sum_{b_1 < x_1} A_{b_1} V + A_{x_1} F(Bx - x_1).
\]
This time $V$ and $F(x)$ are matrix-valued, precisely their values are in $\mathbb{C}^{[1,d] \times [0,v]}$. Matrix $V$ is made from the column vectors $V^{(0)}, \ldots, V^{(v-1)}$ in that order. Anew we must have $F(1) = V$. As functions $\Sigma_K(x)$ and $F_K(x)$ are defined on the segment $[0,1]$, we are searching for solution of previous Equation (16) on $[0,1]$. But we extend function $F(x)$ by

$$F(x) = 0 \quad \text{for} \ x \leq 0, \quad F(x) = V \quad \text{for} \ x \geq 1.$$  

With this choice, Equation (16) rewrites as an homogeneous equation

$$F(x)J = \sum_{0 \leq b < B} A_b F(Bx - b).$$  

This equation is a dilation equation or two-scale difference equations. Such equations are common in the definition of wavelets [28] and refinement schemes [29], and have been studied in detail [30, Th. 2.2], [31, 32]. They appear also in the study of dynamical systems toy-examples [33, 34].

Let us recall that a function $f$ from the real line into a normed space is Hölder continuous with exponent $\alpha$ if it satisfies $\|f(y) - f(x)\| \leq c|y - x|^\alpha$ for some constant $c$.

**Lemma 5.** For an eigenvalue with modulus $\rho > \rho_*$, Equation (18) has a unique continuous solution, which is Hölder with exponent $\log B(\rho/r)$ for every $r > \rho_*$.

**Sketch of the proof.** The existence and uniqueness of the solution results from the Banach fixed-point theorem [28, §4]. More concretely, there is a power of the fixed-point operator $L\Phi(x) = \sum_{0 \leq b < B} A_b \Phi(Bx - b)J^{-1}$ which is a contraction mapping. The result about the Hölderian character of the solution is common in the study of dilation equations [29] p. 1039].

The solution of the dilation equation is not always explicit, but it can always be computed by the algorithm known as cascade algorithm [35, §6.5, p. 206], which enter into the family of interpolatory schemes [36]. All pictures in this article are computed by this algorithm. We will use it immediately in the next example.

**Example 3 (Dichopile algorithm, continued).** With the previous linear representation for the sequence $\delta_n$, the matrix

$$Q = A_0 + A_1 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 0 & 0 & 1 \end{pmatrix}.$$  

admits the Jordan normal form $\Lambda = P^{-1}QP$, with

$$\Lambda = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad P = \frac{1}{12} \begin{pmatrix} 0 & 2 & 0 & 6 & -2 & -6 \\ 0 & 4 & 0 & -6 & 2 & 0 \\ 0 & 4 & 0 & 6 & 2 & 0 \\ 0 & 2 & 0 & -6 & -2 & 6 \\ 12 & -16 & 6 & 15 & 1 & 0 \\ 0 & 10 & -6 & -15 & -1 & 6 \end{pmatrix}. \quad (19)$$

Let us consider only the Jordan block relative to 2, in the upper left corner. It has size 2, and we have to find two vector-valued functions $F^0(x)$ and $F^1(x)$, which we rename $f(x)$ and $g(x)$ to shorten. The equation for $f(x)$ is a homogeneous dilation equation. It is readily seen that it solves into $f_5(x) = x$ and $f_i(x) = 0$ for the other indices $i$.

(Obviously these expressions are valid only for $x$ in $[0,1]$.)
Let us consider more carefully the dilation equation for \( g(x) \). It is non-homogeneous and needs the knowledge of \( f(x) \) to be solved. It writes for \( 0 \leq x \leq 1 \)

\[
\begin{align*}
g_1(x) &= \frac{1}{2} g_3(2x - 1) \\
g_2(x) &= \frac{1}{2} g_1(2x) + \frac{1}{2} g_4(2x) + \frac{1}{2} g_2(2x - 1) \\
g_3(x) &= \frac{1}{2} g_3(2x) + \frac{1}{2} g_1(2x - 1) + \frac{1}{2} g_4(2x - 1) \\
g_4(x) &= \frac{1}{2} g_2(2x) \\
g_5(x) &= \frac{1}{2} g_5(2x) + \frac{1}{2} g_1(2x - 1) + \frac{1}{2} g_5(2x - 1) + \frac{1}{2} g_6(2x - 1) - \frac{1}{2} x \\
g_6(x) &= \frac{1}{2} g_1(2x) + \frac{1}{2} g_2(2x) + \frac{1}{2} g_6(2x) + \frac{1}{2} g_2(2x - 1)
\end{align*}
\]

With \( g(0) \) and \( g(1) \) as input, the cascade algorithm computes first \( g(1/2) \), next \( g(1/4) \) and \( g(3/4) \), and in the end all the values \( g(i/2^K) \) with \( 0 \leq i \leq 2^K \) up to some depth \( K \), using the formulæ above. In this way, we draw the pictures of Figure 1 and we see explicit expressions for \( g_1, g_2, g_3, g_4 \), like

\[
g_1(x) = \begin{cases} 
0 & \text{if } 0 \leq x \leq 1/2, \\
(x - 1/2)/3 & \text{if } 1/2 \leq x \leq 1.
\end{cases}
\]

Such empirical results are not a proof, but the system has a unique continuous solution and a mere substitution establishes the formulæ. Hence we know the four first components. Next we find elementarily

\[
g_6(x) = \begin{cases} 
x/3 + 1/2 & \text{if } 1/2 \leq x \leq 1, \\
(k + 2)x/3 + 1/(3 \times 2^k) & \text{if } 1/2^{k+1} \leq x \leq 1/2^k \text{ with } k \geq 1.
\end{cases}
\]

However we do not have an explicit expression for \( g_6(x) \). We only know that it is well and completely defined by the dilation equation

\[
g_6(x) = \frac{1}{2} g_5(2x) + \frac{1}{2} g_7(2x - 1) + h(x) \quad \text{for } 0 \leq x \leq 1,
\]

where \( h(x) \) is a known piecewise affine function, with the boundary conditions \( g_7(0) = 0, g_7(1) = -4/3 \). Moreover, it is Hölder continuous with exponent \( \log_2(2/r) \) for every \( r > \rho_s = 1 \), that is with exponent \( \varepsilon \) for every \( \varepsilon < 1 \). \( \triangle \)

5. Asymptotic expansion

5.1. Asymptotic expansion for the rational series

We proceed to a Jordan reduction of the matrix \( Q \), which is, let us call it to mind, the sum of the square matrices of the linear representation for the sequence \( u_n \). Next we expand the column vector \( C \) over the Jordan basis, so that we have to consider the sum \( \Sigma_K(x) \), defined by \( [43] \), relative to each vector of the Jordan basis. That is, using the same notations as in the previous section, we have to consider

\[
\Sigma_K(x) = \sum_{\substack{|w|=K \\atop (0,w) \neq e \leq x}} A_w V^{(\nu-1)}
\]

When \( x \) goes from 0 to 1, the sum goes from nearly the null vector in \( \mathbb{C}^d \) for \( x = 0 \) to the vector \( Q^K V^{(\nu-1)} \) for \( x = 1 \). This last quantity evaluates to

\[
Q^K V^{(\nu-1)} = \left( \begin{array}{c} K \\ \nu - 1 \end{array} \right) (\rho \omega)^K \nu + 1 V^{(0)} + \left( \begin{array}{c} K \\ \nu - 2 \end{array} \right) (\rho \omega)^K \nu + 2 V^{(1)} + \cdots \\
+ \left( \begin{array}{c} K \\ 1 \end{array} \right) (\rho \omega)^K \nu - 1 V^{(\nu-2)} + (\rho \omega)^K V^{(\nu-1)},
\]

(20)
with the help of the relationship $QV = VJ$ (as in the previous section, matrix $V$ has type $[1, d] \times [0, \nu]$ and its columns are $V^{(0)}, \ldots, V^{(\nu-1)}$). It is not too astonishing that the intermediate values may be described by the following expression

$$A_K(x) = \left( \frac{K}{\nu - 1} \right) (\rho \omega)^{K-\nu+1} F^{(0)}(x) + \left( \frac{K}{\nu - 2} \right) (\rho \omega)^{K-\nu+2} F^{(1)}(x) + \cdots$$

$$+ \left( \frac{K}{1} \right) (\rho \omega)^{K-1} F^{(\nu-2)}(x) + (\rho \omega)^{K} F^{(\nu-1)}(x),$$

at least asymptotically, since functions $F^{(k)}(x)$ are obtained by a refinement scheme whose input are both values at the ends of the interval.

**Lemma 6.** For a Jordan vector $V^{(\nu-1)}$ associated with an eigenvalue $\rho \omega$, whose modulus $\rho$ is larger than the joint spectral radius, that is $\rho > \rho_*$, the expression $A_K(x)$ above is an asymptotic expansion for the sum $\Sigma_K(x)$. The error term is $O(r^K)$ for every $r$ between $\rho$ and $\rho_*$, and it is uniform with respect to $x$.

**Sketch of the proof.** By substitution into the basic recursion Formula (14) of $\Sigma_K(x) = A_K(x) + \Delta_K(x)$, we find $\Delta_{K+1}(x) = A_{x_1} \Delta_K(Bx - x_1)$. Iterating this relationship, we bring out long products of matrices $A_b$ whose norm is about $\rho_*^K$ and consequently $O(r^K)$.

**Example 4 (Dichopile algorithm, continued).** Lemma 6 permits us to find an asymptotic expansion for $S_K(x)$ as follows. We expand the column vector $C$ of the linear representation (6) onto the Jordan basis. We find

$$C = V_2^0 + V_2^1 + V_1^1 - 2V_{-1}^0 - V_0^0,$$

Figure 1: The six components (from left to right, and from top to bottom) of the vector-valued function $g(x) = F^1(x)$ for the dilation equation associated with the dichopile algorithm.
Figure 2: Theory is closely akin to practice: for the dichopile algorithm, the normalized sequence \( N^{-1}(f_N - N \log_2(N)/2) \) approaches the 1-periodic function \( \Phi(t) \) as \( N \) goes towards infinity. Abscissae are in logarithmic scale for \( N \). The maximum value of the periodic function is \(-1/3\).

with natural notations: \( V^j \) is a generalized eigenvector for the eigenvalue \( \lambda \). According to Lemma 4, we have to take into account only the Jordan block relative to 2 in matrix \( \Lambda \) of Formula (19), and consequently only \( V^0 + V^1 \).

We apply Lemma 6 and specifically Formula (21) to obtain the contribution of each involved generalized eigenvector. With the notations of Example 6, the contribution of \( V^0 \) to the sum \( \Sigma_k(x) \), relative to \( C \), is \( 2^k F^0(x) = 2^k f(x) \). Similarly, the contribution of \( V^1 \) is \( 2^{k-1} K F^0(x) + 2^K F^1(x) = 2^{k-1} K f(x) + 2^K g(x) \), with \( v = 1 \). So that we obtain

\[
\Sigma_k(x) = 2^{k-1}(K + 2)f(x) + 2^K g(x) + O(r^K)
\]

for \( 1 < r < 2 \). Henceforth the sum \( S_k(x) = L \Sigma_k(x) \), defined in Formula (10), satisfies

\[
S_k(x) = 2^{k-1}(K + 2)x + 2^K g_5(x) + O(r^K). \tag{22}
\]

5.2. Asymptotic expansion for the radix-rational sequence

Equation (11) in Lemma 2, which relates \( s_N \) and \( S_{K+1} \), gives us the way to obtain the asymptotic expansion of the sequence \( s_N \), in other words the main result of this article.

**Theorem 1.** Let \( s_N \) be a radix-rational sequence, whose the sequence of backward differences is defined by a linear representation \( L \), \((A_b)_{0 \leq b < B}, C\), insensitive to the leftmost zeroes. Then the sequence \( s_N \) admits an asymptotic expansion which is a sum of terms

\[
N^{\log_B \rho} \rho \left( \frac{\log_B N}{m} \right) \times e^{i \Phi \log_B N} \times \Phi(\log_B N), \tag{23}
\]

indexed by the eigenvalues \( \rho e^{i \vartheta} \) of \( Q = A_0 + \cdots + A_{B-1} \) whose modulus \( \rho \) is larger than the joint spectral radius \( \rho_* \), with \( m \) a nonnegative integer and \( \Phi(t) \) a 1-periodic function. The error term is \( O(N^{\log_B \rho}) \) for every \( r > \rho_* \). Moreover the 1-periodic function \( \Phi(t) \) is Hölder continuous with exponent \( \log_B(\rho/r) \) for every \( r \) between \( \rho \) and \( \rho_* \).
It should be noted that the assumption of insensitivity is only a convenience that simplifies the computation. See \[\text{[1]}\] for a general version.

**Sketch of the proof.** The error term is \(O(r^{\log B}N) = O(N^{\log B}r)\) for every \(r > \rho\), according to Lemma 6.

A mere substitution in \(S_{K+1}(x)\) (not in \(S_K(x)\), however) translates the result about \(S_K(x)\) into the expected result about \(s_N\). First we write \(\omega = e^{i\vartheta}\) in (21) with \(\vartheta\) a real number. Next, for each term \((e^{i\vartheta})\rho \log B (x)\), we change \(K\) into \(K + 1\) and \(x\) into \(B^{(1)}\). Moreover we rewrite \(K = \lfloor \log_B N \rfloor\) as \(\log_B N - \{t\}\) (implicitly \(t = \log_B N\)) and we obtain the regular part of an asymptotic expansion for \(s_N\) as a sum of terms

\[
\left(\log_B N + 1 - \{t\}\right)\left(e^{i\vartheta}N^{-(\ell-1+\{t\})}\right)F(B^{(1)})^{-1}.
\]

With the Chu-Vandermonde identity and \(\rho^{\log B}N = N^{\log B}r\), this rewrites to obtain

\[
N^{\log B}B^{(1)}e^{-\ell+1-\{t\}}e^{-i\ell(\ell+1-\{t\})}F(B^{(1)})^{-1}\times \sum_{m=0}^{\ell} \left(\log_B N \right)\left(1 - \{t\}\right)
\]

It turns out that the expansion is a linear combination of the more elementary terms (23).

The proof is complete, except the point about the Hölderian character of \(\Phi(t)\). This property comes from Lemma 5 but the use of the fractional part \(\{t\}\) needs a complement: we have to verify the continuity at integers.

All in all, it amounts to the equality \(Q = VJ\) with the notations of Formula (16).

**Example 5** (Dichopile algorithm, the end). Applying the substitution explained above, we obtain readily

\[
f_N = \frac{1}{2} N \log_2 N + N\Phi(log_2 N) + O(N^\varepsilon),
\]

for \(0 < \varepsilon < 1\) and with a 1-periodic function

\[
\Phi(t) = \frac{3 - \{t\}}{2} + 2^{1-\{t\}}g_3(2^{(1)}-1),
\]

which is Hölder with exponent \(\varepsilon\) for every \(\varepsilon < 1\). Let us detail the change from one asymptotic expansion to the other. First, using Equation (22), we write \(S_{K+1}(x) = 2^K(K+3)x + 2^{K+1}g_3(x) + O(r^K)\). Next we replace \(K\) by \(\log_2 N - \{t\}\) and \(x\) by \(2^{(1)}-1\) to obtain

\[
f_N = 2^{\log_2 N - \{t\}}(2^{\log_2 N - \{t\}} + 3)2^{(1)}-1 + 2^{\log_2 N - \{t\}+1}g_3(2^{(1)}-1) + O(N^{\log r})
\]

and the writing above appears. Figure 2 shows the comparison between the sequence \(N^{-1}(f_N - N \log_2 (N)/2)\) and the periodic function \(\Phi(t)\).

**6. Improvements**

**6.1. Lazy approach**

Our first improvement is in fact a worsening. Perhaps the reader finds that all this algebraic machinery is too complicated for he wants only an asymptotic equivalent for the sequence under consideration. In that case, it suffices to simplify the process of computation.

**Example 6** (Dichopile algorithm, the true end). Let us assume that we are only interested by an equivalent for the cost \(f_N\) of the dichopile algorithm. We compute the spectral radius of the matrix \(Q\) and we find it is 2. Using the maximum absolute column sum norm, we find numerically for the length \(T = 10\)

\[
\max_{|w|=10} ||Aw||_1^{1/10} \approx 1.271 < 2.
\]

12
(As a matter of fact, the example is so simple that we find at hand the value \( \sqrt{3} \approx 1.732 \) for the length 2 and this is sufficient, but we try to be a little more generic.) We proceed as in Example 4 but we retain only the contribution of the vector \( V_2 \), and moreover we take only the first term in (21), that is \( \left( \frac{1}{\mu} \right)^{K-r} F(x) \). As often for the cost of an algorithm, the function \( F(x) = f(x) \) is explicit (Ex. 3) and we obtain

\[
S_K(x) \sim 2^{K-1} K x, \quad \text{henceforth} \quad f_N(x) \sim \frac{1}{2} N \log_2 N.
\]

\( \triangle \)

6.2. Improvement of the error term

It is slightly irritating that in using Theorem 4 we have always to write the error term is \( \mathcal{O}(N^{\log_B r}) \) for every \( r > \rho_* \). It would be simpler if we could write the error term is \( \mathcal{O}(N^{\log_B \rho_*}) \). This is not true in full generality, but there is a circumstance that guarantees this property. In the definition of the joint spectral radius \( 15 \), it can happen that there exists a subordinate norm and a length \( T \) such that the equality

\[
\rho_* = \max_{\|w\|=T} \|A_w\|^{1/T}
\]

(27)
takes place. In such a case, it is said the set of matrices has the finiteness property \( 37 \). We will said also that the linear representation has the finiteness property.

The coordinate vector \( C \) of the representation decomposes onto the Jordan basis used to reduce the matrix \( Q \). The eigenvalues associated to the generalized eigenvectors which occur in this decomposition can be sorted into two sets: first the set \( \Lambda_\infty \) of eigenvalues of \( Q \) larger than the joint spectral radius, next its complementary part \( \Lambda_\leq \). The first set \( \Lambda_\infty \) provides the regular part of the asymptotic expansion. The other set \( \Lambda_\leq \) provides the error term.

Lemma 7. If the linear representation has the finiteness property, the error term in the expansion announced by Theorem 4 writes

\[ - \mathcal{O}(N^{\log_B \rho_*}) \text{ if } \Lambda_\leq \text{ does not contain the joint spectral radius } \rho_*, \]

\[ - \mathcal{O}(N^{\log_B \rho_* \log m} N) \text{ if } \rho_* \text{ is a member of } \Lambda_\leq \text{ and } m \text{ is the maximal size of the Jordan cells associated to } \rho_* \text{ and involved in the decomposition of the coordinate vector } C \text{ over the Jordan basis.} \]

Sketch of the proof. We deal with every eigenvalue not larger than \( \rho_* \). We use the same notations as in Lemma 4. In the first case \( \rho < \rho_* \), the proof of this proposition works with \( \rho_* \) in place of \( r \). In the second case, let us assume for a while that we have \( T = 1 \) in (27). With the recursive formula (14) and \( Q(K) = \rho^K K^{\nu-1} \), we obtain

\[ \|\Sigma_{K+1}\|_{\infty} \leq c \rho^K K^{\nu-1} + \rho_* \|\Sigma_K\|_{\infty} \]

for some positive constant \( c \). This recursion solves into \( \|\Sigma_K\|_{\infty} = \mathcal{O}(\rho^K K^{\nu}) \). For the general case \( T \geq 1 \), we use the subsequences \( \Sigma_{KT+s} \) with \( 0 \leq s < T \). To have a bound at our disposal we employ \( m \) that is the maximal \( \nu \) encountered in the decomposition of \( C \). Substituting log \( B \) \( N \) for \( K \), we arrive at the announced result.

The next example enables us to see, with our eyes, that the order of the error term predicted by the above proposition is the right one.

Example 7 (Triangular tiling). This example is slightly different of the other examples in the article because first we do not consider a sequence but a rational series, second the linear representation is not insensitive to the leftmost zeroes. This last point is of no importance here because it was only useful in the change from a rational series to a radix-rational sequence. For a real \( \vartheta \), we consider the rotation matrix

\[
R_{\vartheta} = \begin{pmatrix}
\cos \vartheta & -\sin \vartheta \\
\sin \vartheta & \cos \vartheta
\end{pmatrix}.
\]

In the following we use \( V_0 = E_1 \), the first vector of the canonical basis and \( V_{\vartheta} = R_{\vartheta} V_0 \). The example is based on the linear representation, with radix 2,

\[
L = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad A_0 = R_{-\vartheta}, \quad A_1 = R_{\vartheta}, \quad C = V_0.
\]

13
Because we use orthogonal matrices, the joint spectral radius is $\rho = 1$. The matrix $Q$ is the diagonal matrix $2 \cos \vartheta I_2$. We particularize the case to $\vartheta = \pi/3$, so that the only eigenvalue is $\rho = \rho_\nu$ with $\nu = 1$, and $m = 1$.

We consider the words $u = 11$, $v = 01$, and the rational numbers $x_{2K + 2}$ and $y_{2K}$ whose binary expansions are $(0.u^K)_2$ and $(0.v^K)_2$ respectively. According to the functional equation (14) satisfied by $\Sigma_K(x)$, we have

$$\Sigma_{2K+2}(x_{2K+2}) = R_{-\pi/3}V_0 + R_{\pi/3}(R_{-\pi/3}V_0 + R_{\pi/3}\Sigma_{2K}(y_{2K})) = V_{-\pi/3} + V_0 + R_{\pi/3}\Sigma_{2K}(y_{2K}),$$

$$\Sigma_{2K}(y_{2K}) = R_{-\pi/3}(R_{-\pi/3}V_0 + R_{\pi/3}\Sigma_{2K-2}(y_{2K-2})) = V_{-2\pi/3} + \Sigma_{2K-2}(y_{2K-2}).$$

With $\Sigma_0(0) = V_0$, we obtain by induction $\Sigma_{2K}(y_{2K}) = KV_{-\pi/3} + V_0$ and

$$\Sigma_{2K+2}(x_{2K+2}) = (K + 1)V_0 + V_{-\pi/3} + V_{2\pi/3} = (K + 1)V_0.$$
Algorithm 1: LRtoAE

**Input**: A linear representation $L$, $(A_k)_{0 \leq k < b}$, $C$ insensitive to the leftmost zeroes for the sequence of backward differences $u_n = \nabla_s s_n$. 

**Output**: An asymptotic expansion of the sequence $s_N$ with respect to the scale $N^\alpha (\log B N)^\ell$, $\alpha \in \mathbb{R}$, $\ell \in \mathbb{N}_{\geq 0}$. 

1. $Q := \sum_{0 \leq k < b} A_k$; 
2. compute a Jordan basis $V$ for the matrix $Q$; 
3. compute the joint spectral radius $\rho_*$ of the linear representation; 
4. if the linear representation has the finiteness property then 
   5. $r := \rho_*$; 
   6. else 
   7. $r := \text{any number between } \rho_* \text{ and the infimum of the modulus of eigenvalues of } Q \text{ greater than } \rho_*$; 
   8. end 
9. expand the column vector $C$ of the linear representation over the Jordan basis, as $C = \sum_{V \in V} \gamma_V V$; 
10. $V_\gamma := \text{the set of generalized eigenvectors in } V \text{ such that } \gamma_V \neq 0 \text{ and the associated eigenvalue } \rho_\omega \text{ has a modulus } \rho > r$; 
11. for each vector $V$ in $V_\gamma$ do 
   12. $\rho_\omega := \text{the eigenvalue associated to } V$; 
   13. $\nu := \text{the size of the Jordan cell } J \text{ associated to } V$; 
   14. compute the solution $F(x)$ in $C[0,\nu) \times C[1,\nu]$ of the dilation system $F(x)J = \sum A_k F(Bx - b)$ with the boundary conditions $F(x) = 0$ for $x \leq 0$ and $F(x) = V$ for $x \geq 1$; 
   15. write down the expansion, with $F^{(-1)}(x) = 0$, $E_{V,N} := \sum_{t=0}^{\nu-1} \left( K + 1 \over \ell \right) (\rho_\omega)^{K+1-t} \left[ \rho_\omega F^{(\nu-t-1)}(B(t)-1) + F^{(\nu-t-2)}(B(t)-1) \right]$ substitute $\log B N - \{t\}$ for $K$ in $E_{V,N}$ (implicitly $t = \log B N$); 
   16. expand the binomial coefficients $\binom{\log B N + 1 - \{t\}}{\nu}$ by the Chu-Vandermonde formula; 
   17. collect $E_{V,N}$ according to the binomial coefficients $\binom{\log B N}{m}$ 
18. end 
19. $E_N := \sum_{V \in V_\gamma} \gamma_V E_{V,N}$; 
20. collect $E_N$ according to the powers of $N$ and next to the index $m$ of the binomial coefficients to obtain $E_N = \sum_{\rho,\theta,m} N^{\log B \rho} \binom{\log B N}{m} e^{i\theta \log B \rho \Phi_{\rho,\theta,m}(\log B N)}$ where $\Phi_{\rho,\theta,m}(t)$ is 1-periodic and Hölder $\log B (\rho/\rho_*)$; 
21. error $\text{term}_N := O(N^{\log B \rho} t)$; 
22. if the linear representation has the finiteness property then 
23. $V_\gamma := \text{the set of generalized eigenvectors in } V \text{ such that } \gamma_V \neq 0 \text{ and the associated eigenvalue } \rho_\omega \text{ has a modulus } \rho > r$; 
24. $\nu_{\max} := \text{the maximal size of the Jordan cell for the vectors } V \text{ in } V_\gamma$; 
25. error $\text{term}_N := O(N^{\log B \rho_* \log B_{\max} N})$ 
26. end 
27. return $s_N = LE_N + \text{error}_N$.
Example 8 (Biased coin distribution function). Billingsley [42, Ex. 31.1, p. 407] studied the random variable $X = \sum_{n \geq 0} X_n/2^n$ where $X_n$ is the result of a coin tossing with probabilities $p_0$ and $p_1$ for $X_n = 0$ and $X_n = 1$ respectively. This defines a rational series with dimension 1, radix 2 and a linear representation $L = (1)$, $A_0 = (p_0)$, $A_1 = (p_1)$, $C = (1)$ with $0 < p_0, p_1 < 1$, $p_0 + p_1 = 1$. From the standpoint of number theory, the associated sequence is a completely 2-multiplicative function [43, Def. 8.1.5]. We have $Q = (1)$, $\rho = 1$, $V = (1)$, $\rho^* = \max(p_0, p_1)$. The distribution function $F$ is the solution of the dilation equation (16) and it is Hölder with exponent $\log_2(1/\rho^*)$. We illustrate the example with $p_0 = 1/5$, $p_1 = 4/5$ and the exponent is $\log_2(4/5) \approx 0.322$, which provides us with Figure 4, already sketched out in [44, p. 268–269] or [45], where the distribution function is quoted as Lebesgue’s singular function. A similar picture appears also in [34, Fig. 6] about baker-type maps and in [46], which is a gentle introduction to the idea of box dimension.

The provided exponent is the best uniform exponent. The positive character of the representation permits us to show (see [47] for an example) that, assuming $p_0 \leq p_1$, at every dyadic point the best local Hölder exponent is $\log_2(1/p_0)$ on the right-hand side and $\log_2(1/\rho^*)$ on the left-hand side. Except in the case $p_0 = p_1 = 1/2$, this gives $\log_2(1/p_0) > 1$ and this explains the right-hand sided horizontal tangents in the pictures. See [45] for details about the derivative and a bibliography.

A recurrent question is about the bounds of the periodic functions. It is dealt with for example in [48, 11] and more recently in [49, 50].

An issue which does not seem to have been tackled yet concerns the symmetries of the solutions of dilations equations. It appears slightly in [41] with the symmetry $\tau(1-x) = \tau(x)$ of Takagi function. We exemplify it with the Rudin-Shapiro sequence.

Example 9 (Rudin–Shapiro sequence). The Rudin-Shapiro sequence may be defined as $u_n = (-1)^{e_2;11(n)}$ where $e_2;11(n)$ is the number of (possibly overlapping) occurrences of the pattern 11 in the binary expansion of the integer $n$ [51]. It is 2-rational: it admits the generating family $(u_n, u_{2n+1})$ and the reduced linear representation, insensitive to the leftmost zeroes,

$$L = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$ 

As an application of Theorem 1, we obtain the asymptotic expansion for the Rudin–Shapiro sequence [48]

$$\sum_{n \leq N} u_n = N \Phi(\log_4 N) + O(1)$$
where $\Phi$ is the 1-periodic function defined by $\Phi(t) = 2^{1−(t)}F(4^{t−1})$ and $F$ is defined through a dilation equation for radix 4. More precisely, we use the radix 4 linear representation $L' = L$, $A'_0 = A_0^2$, $A'_1 = A_0A_1$, $A'_2 = A_1A_0$, $A'_3 = A_1^2$, $C' = C$. Functions $F$ and $\Phi$ are illustrated in Figure 5. Let us denote $\pi_k$ the part of the graph of $F$ which corresponds to the interval $[k/8, (k + 1)/8]$ for $0 \leq k < 8$. It is evident that the parts of odd index on one hand and the part of even index on the other hand reproduce the same pattern (with a piece upside down). This can be proved elementarily, playing at the same time with the radix 2 representation and the radix 4 representation.

The symmetries of $F$ are translated to $\Phi$, but they lose their graphical evidence. The pieces $\pi_0$ and $\pi_1$ disappear and the pieces $\pi_2$, $2 \leq k < 8$ become the pieces $\pi'_k$ associated to the intervals $[\log_4(k/2), \log_4((k + 1)/2)]$. The links between the pieces become more intricate. For example the pieces $\pi_2'$ and $\pi_3'$ on one side and $\pi_6'$ and $\pi_7'$ on the other side are linked by the formula $2^t\Phi(s) + 2^s\Phi(t) = 4$ under the condition that both numbers $s \in [0, 1/2]$ and $t \in [\log_4 3, 1]$ are related by $4t - 4s = 2$.

7. Background

7.1. Context

Sequences related to a numeration system have a long history, but for what we are concerned the first studies about their asymptotic behaviour appear in the middle of the 20th century. The studies in question deal with one example at a time and use elementary methods. In our opinion, the most noteworthy article is the Brillhart, Erdős, and Morton’s work [48], for it contains the seeds of almost all ideas about the topic: dilation equation, asymptotic expansion with periodic coefficients, Hölder continuous and non-differentiable functions, bounds for the periodic functions, Fourier series, Fourier coefficients, Dirichlet series, residues, Mellin transform. It can be and needs to be read and re-read many times to extract all its very substance.

A more systematic study has begun in the late seventies, with two independent methods. One is related to the combinatorics of words [52, 53, 54, 55], and is summarized in [56, Sec. 4.1 and Sec. 6.1]. The other is based on analytic number theory [8, 57, 58, 59]. Its most evolved version is [60], even it is essentially limited to positive sequences, for it mixes several radices.

Until recently, the idea of radix-rational sequence was not emphasized, but this property sometimes rises to the surface [61 Corollaire, p. 13-07], [62 Satz 3 and Satz 4]. This is not surprising for it was defined only twenty years ago [18]. The sequences are viewed only as satisfying divide-and-conquer recurrences without this wording being necessarily employed. Nevertheless, a linear representation appear in [63], where
a new approach based on Fourier transforms and probability theory is used (this last point seems to limit
the method to nonnegative linear representations). Besides, dilation equations arise in [64, 19, 65] and are
systematically used in [66]. But these authors limit themselves to scalar-valued functions. In comparison,
[67] uses matrices, eigenvalues and dilation equations, but the study is very specific to the Thue-Morse
sequence.

Stolarsky [16] gives an extended bibliography (a little old-looking due to the date of publication, but rich)
about digital sequences, that is sequences based on a numeration system. More recent bibliographies may
be found in [1] or [43]. Moreover, Stolarsky notes: Whatever its mathematical virtues, the literature on sums
of digital sums reflects a lack of communication between researchers, a sentence which is yet topical. This
was a motivation for us to provide to the reader an extended bibliography with the hope that this domain of
research can get organized and that the basic results about that topic will be known of all interested people.

7.2. Linear algebra versus Analytic number theory

Our point of interest here is the comparison between the analytic number theory method and our linear
algebra method. The analytic method is based on meromorphic functions and computation of some residues.
A good and concise account of the method is given in [43, Sec. 8.2.3]. The underlying idea is geometrically
obvious but the application remains tricky. By contrast our approach is elementary but less intuitive.
Nevertheless, it would be a mistake to oppose both approaches. Each has its own merit and the algebraic
approach can greatly help the analytic approach. We will emphasize this point with our favorite example.

Example 10 (Dichopile algorithm, another way). The use of the analytic number theory approach begins with the
Dirichlet series

$$u(s) = \sum_{n=1}^{\infty} \frac{\nabla f_n}{n^s}$$

associated with the sequence $u_n = \nabla f_n$, or better with the family of Dirichlet series associated with each of the
sequences in the basis used to define the linear representation (Ex. 1). All these Dirichlet series are gathered into a
row-vector valued Dirichlet series $U(s)$, associated to the row-vector valued sequence $U_n$ whose components are the
sequences of the basis. Consequently the component of interest is recovered by $u(s) = U(s)C$.

The first question is to evaluate the abscissa of absolute convergence of $U(s)$. Usually this results from an ad hoc
computation, usually a bound obtained by elementary arguments in a former work [18, 55 Formule (6.7), (6.10)].
For us this results from the computation of the joint spectral radius. The value $\rho_n = 1$ shows that all sequences in
the basis are $O(n^\varepsilon)$ for every $\varepsilon > 0$ and this asymptotic relationship is the best possible (among the comparisons
with a power of $n$), so that the abscissa of absolute convergence is $\sigma_n = 1$.

Next, we filter the integer $n$ in the sum that defines $U(s)$ according to their parity. This provides us with

$$U(s) = \sum_{n=1}^{\infty} \frac{U_{2n}}{(2n)^s} + U_1 + \sum_{n=1}^{\infty} \frac{U_{2n+1}}{(2n+1)^s} = \frac{1}{2}U(s)A_0 + U_1 + \frac{1}{2^s}U(s)A_1 + \sum_{n=1}^{\infty} \left( \frac{1}{(2n+1)^s} - \frac{1}{(2n)^s} \right) U_n A_1,$$

or in other words

$$U(s)(I_6 - 2^{-s}Q) = \nabla U(s) \quad \text{with} \quad \nabla U(s) = U_1 + \sum_{n=1}^{\infty} \left( \frac{1}{(2n+1)^s} - \frac{1}{(2n)^s} \right) U_n A_1. \quad (28)$$

The last series converges absolutely for complex numbers $s$ with a positive real part, because of the difference
$(2n+1)^{-s} - (2n)^{-s}$, which is of order $n^{-s-1}$. Equation (28) shows that $U(s)$ extends on the half-plane $\sigma > 0$ as a
meromorphic function. More precisely the poles are the logarithms to base 2 of the eigenvalues of the matrix $Q$. As
a consequence, on the boundary of the half-plane of absolute convergence there is a vertical line of poles $1 + \chi_k$ with
$\chi_k = 2k\pi i / \ln 2$ and $k$ integer, associated to the dominant eigenvalue 2. There are no other poles in the vertical strip
between 0 and 1.

The next stage is the use of the Mellin-Perron summation formula [68, Th. 13]

$$\sum_{1 \leq k < N} U_k + \frac{1}{2} U_N = \frac{1}{2\pi i} \int_{(c)} U(s) N^s ds. \quad (29)$$
In this formula the integral is taken along a vertical line \((c)\) at an abscissa \(c\) larger than the abscissa of absolute convergence. We express the function \(U(s)\) using Formula \((25)\) and to emphasize the poles at abscissa 1, we write \((1_0 - 2^{-s}Q)^{-1} = R(s)/(1 - 2^{-1-s})^2\) with \(R(s)\) a meromorphic function, analytic on the right of 0. The function to be integrated has an expansion near \(s = 1 + \chi_k\) of the form

\[
\frac{R(s)U(s)}{(1 - 2^{-1-s})^2} = \frac{1}{s - 1 - \chi_k} c_{1,k}N^{1+\chi_k} + \frac{1}{s - 1 - \chi_k} (c_{1,k}N^{1+\chi_k} \ln(N) + c_{0,k}N^{1+\chi_k}) \frac{1}{s - 1 - \chi_k} + O(1).
\]

With Cauchy’s residue theorem we change Formula \((29)\) into

\[
\sum_{1 \leq k < N} U_k + \frac{1}{2} U_N = \sum_{k = -\infty}^{\infty} c_{1,k}N^{\chi_k} \ln(N) + \sum_{k = -\infty}^{\infty} c_{0,k}N^{\chi_k} + \frac{1}{2\pi i} \int_{(c)} U(s) N^s \frac{ds}{s},
\]

where this time \(\varepsilon\) is between 0 and 1. The term \(U_N/2\) and the integral are of order \(N^\varepsilon\). Overall \(N^\chi_k\) is nothing but \(\exp(2k\pi i \log_2 N)\) and some trigonometric series appear,

\[
\sum_{1 \leq k < N} U_k = N \ln(N) \sum_{k = -\infty}^{\infty} c_{1,k} \exp(2k\pi i \log_2 N) + N \sum_{k = -\infty}^{\infty} c_{0,k} \exp(2k\pi i \log_2 N) + O(N^\varepsilon).
\]

We have obtained an asymptotic expansion in the scale \(N^\varepsilon\) \(\ln^2(N)\) with variable coefficients. The method is simple, concrete, obvious, natural: we see the poles, we see the line of integration, we push the line to the left, we catch the residues, and we have the asymptotic expansion of the partial sum.

Unfortunately the radiant sun that illumines this method was soon overshadowed by thick clouds. First, we do not know if the points \(1 + \chi_k\) are really poles of \(U(s)\). In this example, the only point that is certainly a pole is the abscissa of absolute convergence \(a_n = 1\), according to Landau’s theorem [68, Th. 10], because the matrices \(L_{A_0}\) and \(A_1\) have nonnegative coefficients so that all the components of \(U_n\) are nonnegative. The same phenomenon occurs with the constant sequence of value 1, whose associated Dirichlet series is the Riemann zeta function \(\zeta(s)\).

In that case, Equation \((28)\) is the usual link between \(\zeta(s)\) and the alternate Riemann zeta function. Because of the factor \(1 - 2^{-1-s}\) it seems that \(\zeta(s)\) has a line of poles \(1 + \chi_k\), but we know that the only pole is 1. Here, comparing \((31)\) and our expansion \((29)\), which begins with a dominant term \(N \log_2(N)/2\), we see that 1 may be a double pole but that the other points \(1 + \chi_k\) are certainly at most simple poles.

Second, there is no reason (at this stage) for the trigonometric series above, which is the coefficient of \(N \ln(N)\), to be a convergent series and defines a periodic function. Usually, an extra argument proves the occurrence of a periodic function (see for example [8, Lemma II], [11, Sec. 2] or [48, Sec. 2]). For us, Formula \((26)\) defines a periodic function.

Third, the order of growth of \(U(s)\) along a vertical line does not permit the use of the Mellin-Perron formula. Equation \((25)\) shows, by the study of the right-hand side, that \([U(s + it)]\) does not grow at infinity more rapidly than \(|t|^\sigma\) for \(\sigma > 1\) and than \(|t|^\sigma\) for \(0 < \sigma < 1\), hence no more rapidly than \(|t|^{1-\sigma}\) for \(0 < \sigma < 1\). It is a consequence of the Lindelöf theorem [68, Sec. III.4], which asserts that the order of growth is a convex function with respect to \(\sigma\). The bound is not enough small to guarantee the absolute convergence of the integral on the line \((\varepsilon)\). However the Mellin-Perron formula, say of the second order,

\[
\frac{1}{N} \sum_{1 \leq k < \varepsilon N} U_k = \sum_{1 \leq k < N} U_k \left(1 - \frac{k}{N}\right) = \frac{1}{2\pi i} \int_{(c)} U(s) N^s \frac{ds}{s(s + 1)}
\]

can be used because if we change the line of integration \((c)\) into \((\varepsilon)\) the integrand becomes \(O(|t|^{-1-\varepsilon})\). We obtain an expansion

\[
\frac{1}{N} \sum_{1 \leq k < N} u_k = N \ln(N) \Psi_1(\log_2 N) + N \Psi_0(\log_2 N) + O(N^\varepsilon),
\]

where \(\Psi_1(t)\) and \(\Psi_0(t)\) are 1-periodic functions defined as sums of convergent trigonometric series. But Proposition 6.4 of [66] provides us, by summation of the expansion \((25)\), with

\[
\frac{1}{N} \sum_{1 \leq k < N} u_k = \frac{1}{4} N \ln(N) - \frac{1}{8} N + N \Psi(\log_2 N) + o(N).
\]

(32)
The uniqueness of the asymptotic expansion shows that $\Psi_1(t) = 1/4$ and $\Psi_0(t) = -1/8 + \Psi(t)$. Moreover Proposition 6.4 of [58] gives the link between the Fourier coefficients $c_k(\Phi)$ of the function $\Phi(t)$, which appears in the asymptotic expansion (25), and the Fourier coefficients $c_k(\Psi)$ of the function $\Psi(t)$ in (32),

$$c_k(\Psi) = (2 + \chi_k) c_k(\Psi).$$

This enables us to show that the Fourier series of $\Phi(t)$ is the trigonometric series which appears in (31) as the coefficient of $N$. In other words, to push the line of integration on the left and collect the residues provides us with the right Fourier series, even if this first seems to be a wrong process. Bernstein theorem [69, Vol. I, p. 240], quoted in [59, Prop. 6], guarantees the uniform convergence of the Fourier series for a Hölder continuous function with exponent $> 1/2$. We know by our algebraic approach that it is the case for the dichopile algorithm. In this example, the Fourier series converges, but it is not the general case, and it can be necessary to consider Féjer sums. \triangle

To summarize, we have two methods. One is brilliant but needs dexterity ([10] provides a good example). The other is more pedestrian and more accessible. Moreover, the algebraic approach provides arguments to sustain the analytic approach.

8. Fourier coefficients

Having made the link between the 1-periodic function and its Fourier series, we still have to compute its Fourier coefficients. At this point there are two possibilities. The first one is the fine case, where the Dirichlet series $U(s)$ (notation of Example 10) is explicitly known, as in many examples dealt with in [58] or [59], which use the Riemann zeta function. The second possibility is the generic one and the dichopile algorithm enters in this case. This is the case on which we will lay stress.

8.1. Residues

We have at our disposal several methods to compute numerically the Fourier coefficients. The first method is the direct application of the analytic number theory approach. The Fourier coefficients are obtained through some residues.

Example 11 (Dichopile algorithm, computation of the Fourier coefficients-1). The basic formula is (28), that is $U(s)(1 - 2^{-s}Q) = \nabla U(s)$. The idea is to compute the right member for a pole $1 + \chi_k$ of $U(s)$ to obtain the residue at $1 + \chi_k$. It is practically easier to use the change of basis of Example 3 that is to use a basis for which the matrix $Q$ is in Jordan form. This emphasizes the components of $U(s)$ which are really involved with the line of poles $1 + \chi_k$.

However, the computation of $\nabla U(s)$ is not so easy, because the convergence of the series in (28) is slow. Graber and Hwang [59] deal with examples where they speed up the convergence by considering differences of the second order in place of the first order. Their goal is to ease the process of pushing the line described in Example 10. Overall, they use their so-called 1/2-balancing principle. We will not insist on this point because it is well explained in [59]. Essentially it leads to consider series with a factor $1/4^m$ in place of series with a factor $1/2^m$, hence a clear gain for the convergence speed. It must be noticed that the description of the method with a factor $1/16^m$ is rather optimistic. It works only for very specific examples, and a factor $1/4^m$ is the general case. \triangle

8.2. Crude method

The second method is to return to the definition of the Fourier coefficients. We begin without subtlety.

Example 12 (Dichopile algorithm, computation of the Fourier coefficients-2). By definition, the Fourier coefficients of $\Phi(t)$ are

$$c_k = \int_0^1 \Phi(t) e^{-2\pi i k t} dt.$$

The change of variable $t = 1 + \log_2 x$ with $1/2 \leq x \leq 1$ transforms this formula into

$$c_k = \frac{1}{\ln 2} \int_{1/2}^1 \Phi(\log_2 x) e^{-2\pi i \log_2 x} \frac{dx}{x} = c_k = \frac{1}{\ln 2} \int_{1/2}^1 \left( 1 - \frac{1}{2} \log_2 x + \frac{g_2(x)}{x} \right) e^{-2\pi i \log_2 x} \frac{dx}{x}.$$
Distinguishing the case $k=0$, we have
\[ c_0 = \frac{5}{4} + \frac{1}{\ln 2} \int_{1/2}^{1} \frac{g_1(x)}{x^2} \, dx, \quad c_k = -\frac{i}{4k\pi} + \frac{1}{\ln 2} \int_{1/2}^{1} \frac{g_k(x)}{x^2} e^{-2k\pi i \log_2 x} \, dx \]  \tag{33}

A crude approach is to compute these integrals by the trapezoidal rule, with the nodes $j/2^K$ for $2^{K-1} \leq j \leq 2^K$ and a given $K$. Obviously, this uses the cascade algorithm. As $g_r(x)$ is Hölder with exponent $\alpha$ for every $0 < \alpha < 1$, the error is of order $1/2^{K\alpha}$ for every $0 < \alpha < 1$, that is essentially $1/2^K$. This method has the advantage of the simplicity, but we cannot speed up the computation à la Richardson because there is no asymptotic expansion of the error. It gives only a rough estimation. Below are the values obtained with $K = 10$ on the left-hand side and with $K = 12$ on the right-hand side. The computations have been made with 15 digits to avoid the rounding errors. The correct digits are written in bold.

\[
\begin{array}{ll}
c_0 \simeq -0.3626476334 & c_0 \simeq -0.362754935 \\
c_1 \simeq +0.0032858226 + 0.0019776043 i & c_1 \simeq +0.0032847914 + 0.0019880642 i \\
c_2 \simeq +0.0030605432 + 0.0006287349 i & c_2 \simeq +0.0030699145 + 0.000620152 i \\
c_3 \simeq +0.0016857421 + 0.0012276124 i & c_3 \simeq +0.0016854542 + 0.001235442 i \\
c_4 \simeq +0.0005424079 - 0.0011900529 i & c_4 \simeq +0.0005422498 - 0.0011836225 i \\
c_5 \simeq +0.0011328140 + 0.0005285686 i & c_5 \simeq +0.0011325212 + 0.0005350851 i \\
c_6 \simeq +0.0006169547 + 0.000375604 i & c_6 \simeq +0.0006166015 + 0.0003847058 i \\
c_7 \simeq +0.0006789418 - 0.0004749207 i & c_7 \simeq +0.0006785365 - 0.0004664313 i \\
c_8 \simeq -0.000312128 + 0.0002826099 i & c_8 \simeq -0.0003118626 + 0.0002912762 i \\
c_9 \simeq +0.0003387154 - 0.0004041520 i & c_9 \simeq +0.0003384147 - 0.0003914805 i \\
c_{10} \simeq +0.0005026412 + 0.0000947029 i & c_{10} \simeq +0.0005021373 + 0.0001051344 i \\
\end{array}
\]

This permits us to draw a picture (Fig. 6) of the truncated Fourier series against the periodic function. Because of the Hölderian character of $\Phi(t)$ the difference between the function and the partial Fourier sum of order $n$ is of order $O(\ln(n)/n^{\alpha})$ for every $0 < \alpha < 1$ and the convergence is rather slow \cite{th.10.8].

\[
8.3. \text{Moments}
\]

To go further, let us proceed to a detour. As the result we have in mind has some interest in itself, we consider a rather general framework. For a function $F(x)$ continuous on the interval $[0,1]$, we define its moments and partial moments, respectively
\[
M_{\ell} = \int_0^1 F(z)z^{\ell-1} \, dz, \quad M_{\ell,r} = \int_{r/B}^{(r+1)/B} F(z)z^{\ell-1} \, dz, \tag{34}
\]
Proof. Using the dilation equation (16) we find readily (36) by a mere change of variable and the binomial theorem.

\[
B^\ell M_{\ell}J - QM_{\ell} = \frac{1}{\ell} \sum_{b=0}^{B-2} B^\ell \left( B^\ell - (b+1)^\ell \right) W_b + \sum_{k=1}^{\ell-1} \left( \frac{\ell - 1}{k - 1} \right) \left( \sum_{r=0}^{B-1} r^{\ell-k} A_b \right) M_k,
\]

(35)

\[
M_{\ell,r}J = \sum_{b<r} \left( \frac{r+1}{\ell B^\ell} - r^{\ell} \right) W_b + \frac{1}{B^\ell} \sum_{k=1}^{\ell-1} \left( \frac{\ell - 1}{k - 1} \right) r^{\ell-k} M_k, \quad 0 \leq r < B,
\]

(36)

determine all the moments and partial moments of the matrix-valued function \(F(x)\) defined by Lemma 3.

Lemma 8. With the notations of Section 4 and under the assumption that none of the numbers \(B^\ell \rho, \ell \geq 1,\) is an eigenvalue of \(Q,\) the recursions \((W_b = A_b V^r \text{ for } 0 \leq b < B)\)

\[
B^\ell M_{\ell}J - QM_{\ell} = \frac{1}{\ell} \sum_{b=0}^{B-2} B^\ell \left( B^\ell - (b+1)^\ell \right) W_b + \sum_{k=1}^{\ell-1} \left( \frac{\ell - 1}{k - 1} \right) \left( \sum_{r=0}^{B-1} r^{\ell-k} A_b \right) M_k,
\]

(35)

\[
M_{\ell,r}J = \sum_{b<r} \left( \frac{r+1}{\ell B^\ell} - r^{\ell} \right) W_b + \frac{1}{B^\ell} \sum_{k=1}^{\ell-1} \left( \frac{\ell - 1}{k - 1} \right) r^{\ell-k} M_k, \quad 0 \leq r < B,
\]

(36)

determine all the moments and partial moments of the matrix-valued function \(F(x)\) defined by Lemma 3.

Proof. Using the dilation equation (16) we find readily (36) by a mere change of variable and the binomial theorem.

It may look troublesome to have to solve Equation (36), as it involves multiples of \(M_{\ell}\) on both the left hand and the right hand side. However, we can emphasize the structure of the Jordan cell by writing it \(J = \rho M_{j,0,r} + N\) and the structure of the \(d \times \nu\) matrix \(M_{\ell}\) by viewing it as the collection of its columns \(M_{0,\ell}, M_{1,\ell}, \ldots, M_{\nu-1,\ell}\). We add \(M_{-1,\ell} = 0\) for convenience. With these notations (36) rewrites

\[
(B^\ell \rho I_d - Q)M_{\ell}^j = -B^\ell M_{\ell}^{-1} + \frac{1}{\ell} \sum_{b=0}^{B-2} B^\ell \left( B^\ell - (b+1)^\ell \right) W_b + \sum_{k=1}^{\ell-1} \left( \frac{\ell - 1}{k - 1} \right) \left( \sum_{r=0}^{B-1} r^{\ell-k} A_b \right) M_k, \quad 0 \leq j < \nu,
\]

(37)

because of the equality \(M_{j,\nu} = (0, M_{0,\nu}, M_{1,\nu}, \ldots, M_{\nu-1,\nu})\). It is now clear that Equation (37) enables us to compute successively \(M_{0,\ell}, M_{1,\ell}, \ldots, M_{\nu-1,\ell}, M_0, \ldots, M_{\nu-1}\) and more generally all the moments, if the numbers \(B^\ell \rho\) are not eigenvalues of \(Q\). Last, Equation (36) provides us with the partial moments because \(J\) is invertible.

Example 13 (Dichopile algorithm, computation of the moments). For the dichopile algorithm, Lemma 8 translates into the formulæ

\[
(2^{\ell+1} I_b - Q)M_{\ell}^0 = \frac{2^\ell - 1}{\ell} A_0 V^2 + A_1 \sum_{j=1}^{\ell} \left( \frac{\ell - 1}{j - 1} \right) M_{j,\ell}^0,
\]

(38)

\[
(2^{\ell+1} I_b - Q)M_{\ell}^1 = \frac{2^\ell - 1}{\ell} A_0 V^2 + A_1 \sum_{j=1}^{\ell} \left( \frac{\ell - 1}{j - 1} \right) M_{j,\ell}^1 - 2^\ell M_{\ell,\ell}^0,
\]

(39)

\[
2M_{\ell,1}^0 = \frac{1 - 1/2^\ell}{\ell} A_0 V^2 + \frac{1}{2^\ell A_1} \sum_{j=1}^{\ell} \left( \frac{\ell - 1}{j - 1} \right) M_{j,\ell}^0,
\]

\[
2M_{\ell,0}^1 = \frac{1 - 1/2^\ell}{\ell} A_0 V^2 + \frac{1}{2^\ell A_1} \sum_{j=1}^{\ell} \left( \frac{\ell - 1}{j - 1} \right) M_{j,\ell}^1 - M_{\ell,\ell}^1.
\]

(40)

Since some components of \(f(x)\) and \(g(x)\) are known, we can verify the result of the computation.

8.4. Mellin transform

Let us return to the computation of the Fourier coefficients. If we look scrupulously to the proof of Theorem 1 and particularly to Formula (24), we see that the occurrence of periodic functions in the expansion of the sequence comes from the term

\[
\Phi(t) = \rho^{1-t} e^{i t (1-t)} F(B^{t-1}) \left( \frac{1 - \{t\}}{p} \right)
\]

(41)

for some nonnegative integer \(p\). This integer \(p\) is \(\ell - m\) in the Chu-Vandermonde type Formula (24).
Let us focus on the case $p = 0$. With the change of variable $t = 1 + \log_B x$, $1/B \leq x \leq 1$ (that already appears in Delange’s article [3, p. 37]), the Fourier coefficients of the 1-periodic function $\Phi(t)$ are

$$C_k = \int_0^1 \Phi(t) e^{-2\pi i k t} dt = \int_{1/B}^1 \rho \log_B x e^{i \theta \log_B x} F(x) e^{-2\pi i k \log_B x} \frac{dx}{x \ln B},$$

that is

$$C_k = \frac{1}{\ln B} \int_{1/B}^1 x^{-1 + \ln \log_B x} F(x) x^{-\chi_k} dx,$$

with $\chi_k = 2k\pi i / \ln B$. It turns out that these coefficients can be viewed as special values of a Mellin transform, namely values of

$$F^*(s) = \int_{1/B}^1 x^{s-1} F(x) dx,$$

at points $s = -\chi_k + \log_B \lambda$ with $\lambda = re^{i\vartheta}$ the eigenvalue of $Q$ under consideration. According to Lemma 8, we conversely know the value of the Mellin transform at positive integers,

$$F^* (\ell) = \sum_{r=1}^{B-1} M_{\ell,r},$$

with the notations of [34].

**Lemma 9.** The Mellin transform of the matrix-valued function $F(x)$ defined by Lemma 5,

$$F^*(s) = \int_{1/B}^1 x^{s-1} F(x) dx,$$

is an entire function, which can be expressed as the sum of a Newton series

$$F^*(s) = \sum_{n=0}^{+\infty} \left( \frac{s-1}{n} \right) \Delta^n F^*(1), \quad (42)$$

where $\Delta$ is the forward difference operator, acting on the variable $s$ of $F^*(s)$. A partial sum gives the value $F^*(s)$ with an error of order the first neglected term.

**Proof.** It must be noticed that this function $F^*(s)$ is analytic in the whole complex plane. Actually a Mellin transform is an integral from 0 to $+\infty$, and the behaviour of the function at the ends of the interval determines the vertical strip in the complex plane where the Mellin transform is defined and analytic. Here the integrand vanishes in the neighborhood of 0 and of $+\infty$, so the transform is an entire function. We write the term $x^{s-1}$ as $(1 - (1 - x)^n)^{-1}$ and we expand it by the binomial theorem. This gives

$$F^*(s) = \sum_{n=0}^{+\infty} \left( \frac{s-1}{n} \right) \int_{1/B}^1 F(x)(x-1)^n dx.$$

Again with the binomial theorem, the last integral appears as the $n$th order difference of the sequence $F^*(\ell)$ evaluated at $\ell = 1$. In other words we obtain Equation (42). Moreover the integral expression of $\Delta^n F^*(1)$ permits to show that it behaves as $(F(1 - 1/B) + o(1))(1 - 1/B)^n + 1/(n + 1)$. As a consequence, the series converges at least as fast as a geometric series of ratio $1 - 1/B$. Hence the assertion about the error term.

The above result may suggest that the calculation of the Fourier coefficients is particularly simple. This is not quite true, as the following example shows.
EXAMPLE 14 (Dichophiel algorithm, computation of the Fourier coefficients-3). According to (33), we have to compute the integrals
\[ e_k = \int_{1/2}^{1} g_2(x)x^{-\chi_k-2} \, dx = g_2^*(-1 - \chi_k) \]
and \( g_2^*(s) \) expresses as
\[ g_2^*(s) = \sum_{n=0}^{\infty} \frac{s-1}{n} \Delta^n g_2^*(1) = \sum_{n=0}^{\infty} \left( \frac{s-1}{n} \right) \Delta^n M_{2,1}^1, \]
where \( \Delta \) acts on the first lower index \( \ell \) of \( M_{2,1}^1 \). At this point, two problems arise. First the moments \( M_{2,1}^1 \) are of order \( 1/p \), while the differences \( \Delta^n M_{2,1}^1 \) are of order \( 2^n/n \) according to the proof of the above proposition. Hence there is a strong cancellation in the computation of differences \( \Delta^n M_{2,1}^1 \). To take this phenomenon into account, we do not compute these differences as float numbers but exactly as rational numbers they are. Second for \( s = -1 + \chi_k \) the modulus of \( \left( \frac{\chi_k}{2^n} \right) \) increases first up to a maximal value of order
\[ 10^{2.06|k| - 0.65 \ln |k| - 1.08} \]
obtained for \( n \approx 5|k| \), and next decreases towards 0. Henceforth, to obtain the sum within an error \( 10^{-m}/2 \) we have to sum the terms until they become smaller than \( 10^{-m}/2 \) and we use mantissa of length \( 2.06|k| - 0.65 \ln |k| - 1 + m + 5 \) (with 5 bits to control the rounding errors). Practically, we sum the series up to the \( n \)th term with \( n \approx (\pi/[2^{n(2)}])^2 \cdot |k| + m \log_2 10 - 3/2 \log_2 |k| \approx 20.5 |k| + 3.3 m - 2 \ln |k| \). This analysis is not valid for the case \( k = 0 \), where the absolute value of the sequence is decreasing from the beginning. Below are the first few values of the Fourier coefficients for \( m = 50 \).

\[ c_0 \approx -0.3627648321990552373394157911628179554357682624599 \]
\[ c_1 \approx 0.03284440242836895765428343957045278659679317794466 + 0.00199132027449197796100439035667083426695959747 i \]
\[ c_2 \approx 0.00909100501732717791345773939400386962994937936 + 0.0006207605748476162777330653757375401619891385299 i \]
\[ c_3 \approx 0.0016854033169851520752684372516526158637135512757 + 0.000123480138622010180313152996424913777847524841664 i \]
\[ c_4 \approx 0.0005422693373394980444215803891721804491134682 - 0.000118546883363459902535629724150607450996 i \]
\[ c_5 \approx 0.0001124887958587730781507675577758329883342952392 - 0.000053603714268075111588486643817573977322189555252 i \]
\[ c_6 \approx 0.00006156599797162490158550584690901186842150761797 + 0.0000385617421526202076710596427624390201335494056798 i \]
\[ c_7 \approx 0.000067856380906817663938277151623545389083080834 + 0.000046549574081761810249121821229989781189632338 i \]
\[ c_8 \approx -0.00003118497820623170409286440861244052559475337566921 + 0.00002921772708707377904875974566743920965525668 i \]
\[ c_9 \approx 0.0000338391083941378853968616993689633927151766089 - 0.00003905285208456028985072135158517264916825572241 i \]
\[ c_{10} \approx 0.00005021008394950064916235510500678844805781175404 + 0.00006087090960591513462957275163259840464824986 i \]

We have dealt with the case \( p = 0 \) in Equation (41). The general case of a nonnegative integer \( p \) would lead us to consider derivatives of the Mellin transform \( F^*(s) \).

References
[1] P. Dumas, Joint spectral radius, dilatation equations, and asymptotic behavior of radix-rational sequences, Linear Algebra and its Applications 438 (2013) 2107 – 2126.
[2] J. R. Trollope, An explicit expression for binary digital sums, Math. Mag. 41 (1968) 21–25.
[3] H. Delange, Sur la fonction sommatoire de la fonction “somme des chiffres”, Enseignement Math. (2) 21 (1975) 31–47.
[4] M. Drmota, J. Gajdosik, The distribution of the sum-of-digits function, J. Théor. Nombres Bordeaux 10 (1998) 17–32.
[5] G. F. Clements, B. Lindström, A sequence of \((\pm 1)\)--determinants with large values, Proc. Amer. Math. Soc. 16 (1965) 548–550.
[6] M. D. McIroy, The number of 1’s in binary integers: Bounds and extremal properties, SIAM J. Computing 3 (1974) 255–261.
[7] D. W. Boyd, J. Cook, P. Morton, On sequences of \( \pm 1 \)'s defined by binary patterns, Dissertationes Math. (Rozprawy Mat.) 283 (1989) 64.
[8] P. Flajolet, L. Ramshaw, A note on Gray code and odd-even merge, SIAM J. Comput. 9 (1980) 142–158.
[9] J.-P. Allouche, J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in: Sequences and their applications (Singapore, 1998), Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999, pp. 1–16.
[10] D. J. Newman, On the number of binary digits in a multiple of three, Proc. Amer. Math. Soc. 21 (1969) 719–721.
[11] J. Coquet, A summation formula related to the binary digits, Invent. Math. 73 (1983) 107–115.
[12] W. Rudin, Some theorems on Fourier coefficients, Proc. Amer. Math. Soc. 10 (1959) 855–859.
[13] H. S. Shapiro, Extremal problems for polynomials and power series, Master’s thesis, Massachusetts Institute of Technology, 1951.
