



**HAL**  
open science

## Shifts of k-nested sequences

Marie-Pierre Béal, Pavel Heller

► **To cite this version:**

Marie-Pierre Béal, Pavel Heller. Shifts of k-nested sequences . Theoretical Computer Science, 2017, 658 (A), pp.18-26. 10.1016/j.tcs.2016.06.041 . hal-01790002

**HAL Id: hal-01790002**

**<https://hal.science/hal-01790002>**

Submitted on 11 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Shifts of $k$ -nested sequences <sup>☆</sup>

Marie-Pierre Béal

*Université Paris-Est, Laboratoire d'informatique Gaspard-Monge, UMR 8049 CNRS*

Pavel Heller

*Université Paris-Est, Laboratoire d'informatique Gaspard-Monge, UMR 8049 CNRS*

---

## Abstract

We introduce a new class of subshifts of sequences, called  $k$ -graph shifts, which expresses nested constraints on  $k$  symbols instead of on two symbols like for Dyck shifts. These shifts share many properties with Markov-Dyck shifts but are generally not conjugate to them. We prove that they are conjugate to sofic-Dyck shifts. We give a computation of the multivariate zeta function for this class of shifts.

*Keywords:* Dyck shifts, Markov-Dyck shifts, sofic-Dyck shifts, sofic shifts, symbolic dynamics, visibly-pushdown languages, zeta function.

---

## 1. Introduction

Dyck shifts were introduced by Krieger in [12]. They are sets of bi-infinite sequences over symbols of opening and closing parentheses where no mismatching appears, *i.e.* where each finite factor is a factor of a well-parenthesized word. These shifts are examples of coded systems defined by Blanchard and Hansel [8].

Dyck shifts were generalized to Markov-Dyck shifts by Matsumoto [15] and Krieger and Matsumoto [13] (see also [16], [10]), and to sofic-Dyck shifts (see [2], [3]) which are exactly the sets of sequences avoiding a visibly pushdown language (or a regular language of nested-words) [1]. All these shifts express nesting constraints of arity 2.

In this paper we consider nesting constraints of higher arity. We consider expressions of the form  $(a; (b; c))$ , where the symbol  $;$  is a middle tag separating two parts enclosed by parentheses, a generalization of Dyck-like expressions

---

<sup>☆</sup>This work is supported by the French National Agency (ANR) through “Programme d’Investissements d’Avenir” (Project ACRONYME no. ANR-10-LABX-58), through the ANR EQINOCS (Project ACRONYME no. ANR-11-BS02-004), and by the region of Île-de-France through the DIM RDM-IdF.

*Email addresses:* `beal@u-pem.fr` (Marie-Pierre Béal), `pavel.heller@u-pem.fr` (Pavel Heller)

of the form  $(a(b))$ . We define a class of shifts of sequences called shifts of  $k$ -nested sequences, where  $k$  is the nesting arity, and subshifts called  $k$ -graph shifts defined by structures called  $k$ -graphs. The sets of forbidden factors of such shifts are clearly deterministic context-free languages and Markov-Dyck shifts form a subclass of  $k$ -graph shifts.

We prove that  $k$ -graph shifts are not conjugate to Markov-Dyck shifts but are conjugate to sofic-Dyck shifts. Two conjugate shifts are considered to be essentially the same after some recoding but it is not known whether conjugacy is decidable or not even for shifts of finite type.

We investigate the computation of the zeta function of  $k$ -graph shifts. The zeta function of a shift is a formal series allowing to count its periodic sequences. It is a powerful invariant of conjugacy.

The zeta function has been computed for Dyck shifts by Keller in [11] and for Markov-Dyck shifts by Krieger and Matsumoto in [13]. A formula for the zeta function of sofic-Dyck shifts is given in [3]. All computations are based on an encoding of the periodic patterns of the shift. It was proved by Reutenauer in [17] that the multivariate and ordinary zeta functions of a sofic shift are  $\mathbb{N}$ -rational series. The multivariate and ordinary zeta functions of sofic-Dyck shifts (and thus Markov-Dyck shifts) are  $\mathbb{N}$ -algebraic series [4].

Since  $k$ -graph shifts are conjugate to sofic-Dyck shifts, a formula of their zeta function can be obtained from [3]. Nevertheless, as a consequence of the recoding into sofic-Dyck shifts, the formula would involve  $(m \times m)$ -matrices where  $m$  is exponential in the number of states of the  $k$ -graph. We give here a computation of the multivariate and ordinary zeta functions of a  $k$ -graph shift based directly on the  $k$ -graph structure, the size of the computed matrices staying equal to  $n \times n$  where  $n$  is the number of states of the  $k$ -graph. The proof is obtained using Keller's results and an encoding of periodic patterns similar to one used for sofic-Dyck shifts in [4].

In Section 2 we give a quick background on shifts of sequences. We define the class of shifts of  $k$ -nested sequences in Section 3 and the  $k$ -graph shifts in Section 4. Section 5 contains the computation of the zeta functions. In Section 6 we generalize the  $k$ -graph to  $\mathbf{v}$ -graph shifts to express constraints mixing several nesting arities.

## 2. Background on shifts

We refer to [14] for basic notions in symbolic dynamics. Let  $A$  be a finite alphabet. A *shift* of sequences  $X$  is defined as the set of bi-infinite sequences of symbols of  $A$  avoiding some set  $F$  of finite words (*i.e.* having no finite factor in  $F$ ). The set  $F$  is called a set of *forbidden factors* of  $X$ . The shift  $X$  is denoted  $X = X_F$ . The set of finite factors of a shift  $X$  of bi-infinite sequences is denoted by  $\mathcal{B}(X)$ . It is also called the set of *blocks* of the shift. The set of blocks of  $X$  of length  $l$  is denoted by  $\mathcal{B}_l(X)$ . A shift  $X$  is *irreducible* if whenever  $u, v \in \mathcal{B}(X)$  there is a block  $w$  such that  $uwv \in \mathcal{B}(X)$ .

When  $F$  can be chosen finite (resp. regular, visibly pushdown),  $X$  is called a *shift of finite type* (resp. a *sofic shift*, a *sofic-Dyck shift*). The *full shift* over  $A$  is the set  $A^{\mathbb{Z}}$ .

Shifts of sequences may be defined as closed subsets of  $A^{\mathbb{Z}}$  invariant by the *shift transformation*  $\sigma$ , where  $\sigma((x_i)_{i \in \mathbb{Z}}) = (x_{i+1})_{i \in \mathbb{Z}}$ . Sets of bi-infinite sequences which are invariant by the shift transformation without being necessarily closed subsets of  $A^{\mathbb{Z}}$  are called  *$\sigma$ -invariant sets*. The *orbit* of a sequence  $x \in A^{\mathbb{Z}}$  is the set of all  $\sigma^i(x)$  for  $i \in \mathbb{Z}$ . A *period* of a sequence  $x \in A^{\mathbb{Z}}$  is a positive integer  $p$  such that  $\sigma^p(x) = x$ .

Let  $A, B$  be two finite alphabets and  $X$  a shift over  $A$ . A sliding block code is a map  $\Phi: X \rightarrow B^{\mathbb{Z}}$  for which there is a positive integer  $R = 2r + 1$  and a map  $\phi: \mathcal{B}_R(X) \rightarrow B$  such that  $\Phi(x)_i = \phi(x_{i-r} \cdots x_i \cdots x_{i+r})$ . Such a map is called an  *$R$ -block code*.

A (*topological*) *conjugacy* from  $X \subseteq A^{\mathbb{Z}}$  to  $Y \subseteq B^{\mathbb{Z}}$  is a bijective sliding block code from  $X$  onto  $Y$ . It is known that its inverse is also a sliding block code [14]. Observe that a conjugacy preserves the periods of a sequence.

### 3. Shifts of $k$ -nested sequences

In this section we define shifts of sequences of symbols satisfying some nesting constraints of arity greater than 2. We will denote by  $k$  a positive integer representing a nesting arity as explained below.

Let  $A$  be a finite set of symbols of size  $n$ . We consider the alphabets  $A^{(i)}$  for  $1 \leq i \leq k$  defined by  $A^{(i)} = \{e^{(i)} \mid e \in A\}$ . We set  $A_k = \sqcup_{1 \leq i \leq k} A^{(i)}$ . We consider the free monoid generated by  $A_k$  with a zero  $0$  quotiented by the following relations

$$\begin{aligned} e^{(i)} f^{(j)} &= 0, \text{ if } i \neq k, j \neq 1, e \neq f \\ e^{(i)} e^{(j)} &= 0, \text{ if } i \neq k, j \neq 1, j \neq i + 1 \\ e^{(1)} e^{(2)} \dots e^{(k)} &= \mathbf{1}, \end{aligned}$$

where  $e, f \in A$  and  $\mathbf{1}$  is the unity of the monoid. The third relation expresses a nesting relation of arity  $k$  on the symbols  $e^{(1)}, e^{(2)}, \dots, e^{(k)}$ . The symbols  $e^{(1)}$  may be seen as opening parentheses, the symbols  $e^{(k)}$  as closing parenthesis, and the other symbols  $e^{(i)}$  as middle tags.

For a word  $w = w_0 \cdots w_{n-1}$  over  $A_k$ , we denote by  $\tilde{w} \in A_k^* \cup \{0, \mathbf{1}\}$  its *reduced form* which is the unique word obtained by applying the above relations.

We denote by  $F$  the set of finite words over the alphabet  $A_k$  whose reduced form is  $0$ . We set  $X_{A_k} = X_F$ , which is the shift avoiding all words in  $F$ . The shift  $X_{A_k}$  is called the shift of  *$k$ -nested sequences* over  $A_k$ .

Note that two such shifts  $X_{A_k}$  and  $X_{B_k}$  are conjugate if and only if  $A$  and  $B$  have the same cardinal. This justifies the notation of  $X_{(k,n)}$  standing for the shift  $X_{A_k}$  up to renaming of the symbols of the alphabet  $A$  of size  $n$ .

We denote by  $\text{Dyck}(A_k)$  the set of finite words whose reduced form is  $\mathbf{1}$  and we call these words the *Dyck words* over  $A_k$ . Finite and infinite words avoiding the above set  $F$  of forbidden words have a natural nested structure.

*Example 1.* Let  $A = \{e, f\}$  and  $k = 1$ . The shift  $X_{A_1}$  is the full shift over  $A_1$ , that is, the set  $A_1^{\mathbb{Z}}$ .

*Example 2.* Let  $A = \{e, f\}$  and  $k = 2$ . The shift  $X_{A_2}$  is the *Dyck shift* with two types of parentheses introduced by Krieger in [12] where  $e^{(1)}, f^{(1)}$  represent the opening parentheses  $(, [$ , and  $e^{(2)}, f^{(2)}$  represent the closing parentheses  $), ]$ . The Dyck shift is the set of bi-infinite sequences of symbols in  $\{(, [, ], )\}$  where each finite factor is a factor of some well-parenthesized word, *i.e.* contains no mismatching. For instance the bi-infinite sequence  $\cdots( ( ) [ [ ] ( ( \cdots$  belongs to the Dyck shift while  $\cdots( [ [ ] ) ( \cdots$  does not since it contains a mismatching.

*Example 3.* Let  $A = \{e, f\}$  and  $k = 3$ . The nontrivial monoid relations are the following:

$$\begin{aligned} e^{(i)} f^{(j)} &= f^{(i)} e^{(j)} = 0, \text{ if } i \neq 3, j \neq 1 \\ e^{(1)} e^{(3)} &= e^{(2)} e^{(2)} = f^{(1)} f^{(3)} = f^{(2)} f^{(2)} = 0, \\ e^{(1)} e^{(2)} e^{(3)} &= f^{(1)} f^{(2)} f^{(3)} = \mathbf{1}, \end{aligned}$$

Imagine that  $e^{(1)}, e^{(2)}, e^{(3)}$  correspond to the parenthesis symbols  $(_1, |_1, )_1$  and  $f^{(1)}, f^{(2)}, f^{(3)}$  correspond to the parenthesis symbols  $(_2, |_2, )_2$ . The monoid relations express a natural constraint on composition of parentheses representing two kinds of pairs of objects which have to be separated by some symbol in the middle.

Let  $G = (V, E)$  be the directed multigraph where  $V$  is the set of vertices containing a unique vertex 1 and  $E = \{e^{(1)}, e^{(2)}, e^{(3)}, f^{(1)}, f^{(2)}, f^{(3)}\}$  is the set of edges from 1 to 1. The shift  $X_{A_3}$  may be seen as the set of bi-infinite paths in  $G$  avoiding factors whose reduced form is 0, the alphabet being the set of edges  $E$ .

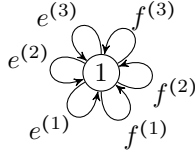


Figure 1: The graph  $G$  defining the shift  $X_{A_3}$ .

For instance, the reduced form of  $w = e^{(1)} e^{(2)} f^{(1)} f^{(2)} f^{(3)} e^{(3)} e^{(1)} e^{(2)} e^{(3)}$  is  $\mathbf{1}$ . The word  $w$  is thus a Dyck word over  $A_3$ . The reduced form of the word  $z = e^{(1)} f^{(1)} f^{(2)} f^{(3)} e^{(3)}$  is 0. The word  $z$  is hence a forbidden factor of the shift  $X_{A_3}$ .

The reduced form of a finite word has always the form 0,  $\mathbf{1}$ ,  $u$ ,  $v$ , or  $u \cdot v$  with

$$\begin{aligned} u &= (e_1^{(j_1)} \dots e_1^{(k)}) \dots (e_n^{(j_n)} \dots e_n^{(k)}) \\ v &= (f_1^{(1)} \dots f_1^{(i_1)}) \dots (f_m^{(1)} \dots f_m^{(i_m)}), \end{aligned}$$

where  $e_r, f_{r'} \in A$ , for  $1 \leq r \leq n$ ,  $1 \leq r' \leq m$  and  $1 < j_r \leq k$  and  $1 \leq i_{r'} < k$ . A word whose reduced form is  $v$  only or  $\mathbf{1}$  is said to be *matched-return*. A word whose reduced form is  $u$  only or  $\mathbf{1}$  is said to be *matched-call*. The set of matched-return (resp. matched-call) words is denoted by  $\text{MR}(A_k)$  (resp.  $\text{MC}(A_k)$ ). The set of matched-return (resp. matched-call) blocks of a subshift  $X$  of  $\mathbb{X}_{A_k}$  is denoted by  $\text{MR}(X)$  (resp.  $\text{MC}(X)$ ).

The following lemma shows that any block of  $\mathbb{X}_{A_k}$  can be extended on the right to get some matched-call block. Similarly any block can be extended on the left to get some matched-return block.

**Lemma 1.** *Let  $X = \mathbb{X}_{A_k}$  and  $u$  a block of  $X$ . There is a block  $v$  such that  $uv$  is a matched-call block of  $X$ .*

*Proof.* Without loss of generality we may assume that  $u \in \text{MR}(X)$ . Let us show that there is a word  $v$  such that  $uv \in \text{Dyck}(X)$ . Let  $\tilde{u}$  be the reduced form of  $u$  in  $S$ . If  $u$  is not already in  $\text{Dyck}(X)$  we have  $\tilde{u} = we^{(1)} \dots e^{(i)}$  where  $1 \leq i < k$  and  $w \in \text{MR}(X)$ . Then  $ue^{(i+1)} \dots e^{(k)}$  reduces to  $w$  shorter than  $\tilde{u}$ . The claim is obtained by iterating this process.  $\square$

Observe that the shift  $\mathbb{X}_{A_k}$  is irreducible. Indeed, let  $X = \mathbb{X}_{A_k}$  and  $u, v \in \mathcal{B}(X)$ . By Lemma 1 there is a word  $w$  such that  $uw \in \text{MC}(X)$ . Similarly there is a word  $z$  such that  $zv \in \text{MR}(X)$ . As a consequence  $uwzv \in \mathcal{B}(X)$ .

#### 4. Shifts defined by $k$ -graphs

In this section we define a class of shifts of nested sequences defined by finite directed multigraphs equipped with a special structure.

Let  $A$  be a finite set of symbols and  $k$  a positive integer. We denote by  $[k]$  the integer interval  $\{1, 2, \dots, k\}$ .

We consider structures called  *$k$ -graphs* denoted by  $\mathcal{G} = (V, E, A, k)$ , where  $(V, E)$  is a multigraph (called simply graph) with a finite set of vertices  $V$  and a finite set of edges  $E$ , and where  $A$  is a finite set of symbols. For each edge  $e$  we denote by  $s(e)$  its starting state and by  $t(e)$  its target state.

Further it is required that the set of edges is partitioned into  $k$  parts of equal size  $E^{(i)} = \{e^{(i)} \mid e \in A\}$  such that for each  $e \in A$ , each  $0 \leq i < k$ ,  $t(e^{(i)}) = s(e^{(i+1)})$  and  $t(e^{(k)}) = s(e^{(0)})$ . Hence each path  $e^{(1)}e^{(2)} \dots e^{(k)}$  is a cyclic path in  $\mathcal{G}$ .

The set of bi-infinite paths in  $\mathcal{G}$  belonging to  $\mathbb{X}_{A_k}$  is a shift denoted by  $\mathbb{X}_{\mathcal{G}}$  and is called a  *$k$ -graph shift*. A *Dyck path* of  $\mathcal{G}$  is a finite path of  $\mathcal{G}$  belonging to  $\text{Dyck}(A_k)$ . Observe that by construction each Dyck path of  $\mathcal{G}$  is a cyclic path of  $\mathcal{G}$ .

*Example 4.* The  $k$ -nested shifts defined in Section 3 are  $k$ -graph shifts defined by graphs containing a single state.

*Example 5.* Let  $\mathcal{G} = (V, E, A, k)$  be the 3-graph defined by  $A = \{e, f, g\}$ ,  $k = 3$  and the edges described in Figure 2. The word  $e^{(3)}e^{(1)}e^{(2)}f^{(1)}f^{(2)}f^{(3)}e^{(3)}f^{(1)}$  is a block whose reduced form is  $e^{(3)}f^{(1)}$ .

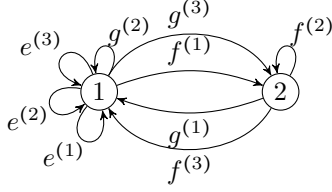


Figure 2: A 3-graph  $\mathcal{G}$  defining the shift  $X_{\mathcal{G}}$  with  $A = \{e, f, g\}$ .

In this setting, the Markov-Dyck shifts introduced by Matsumoto [15] and Krieger and Matsumoto [13] (see also [16], [10]) may be seen as 2-graph shifts as follows. The Markov-Dyck shift defined by some directed graph  $H = (V, E)$  is the 2-graph shift  $X_{\mathcal{G}}$  with  $\mathcal{G} = (V, F, E, 2)$  where  $F = F^{(1)} \sqcup F^{(2)}$ ,  $F^{(1)}$  being a set of copies of the edges of  $H$ , and  $F^{(2)}$  being a set of copies of the edges of  $H$  in the reverse sense. Each edge  $e^{(2)}$  in  $F^{(2)}$  is thus a backward edge of the  $e^{(1)}$  in  $F^{(1)}$  as in Markov-Dyck shifts.

The class of  $k$ -graph shifts is very close to the class of Markov-Dyck shifts but it is a strictly larger class as is shown in the following proposition.

**Proposition 1.** *The 3-graph shift  $X_{(3,n)}$  is not conjugate to a Markov-Dyck shift.*

*Proof.* Assume that there is a 2-graph  $\mathcal{G} = (V, E, B, 2)$  defining a Markov-Dyck shift  $X_{\mathcal{G}}$  which is conjugate to  $X_{A_3}$ . Let  $\Phi : X_{\mathcal{G}} \rightarrow X_{A_3}$  be a conjugacy from  $X_{\mathcal{G}}$  onto  $X_{A_3}$ . Let us assume that  $\Phi$  is an  $R$ -block code. Since  $X_{A_3}$  is an irreducible shift and irreducibility is invariant by conjugacy,  $X_{\mathcal{G}}$  is irreducible. The graph of  $\mathcal{G}$  is thus strongly connected.

If  $a$  is a symbol, we denote by  $\omega a$  (resp.  $a\omega$ ) the left (resp. right) infinite sequence  $\cdots aaa$  (resp.  $aaa\cdots$ ), and by  $\omega a\omega$  the bi-infinite sequence  $\omega a\omega$ .

Since  $X_{A_3}$  has sequences of period 1,  $X_{\mathcal{G}}$  also and thus  $\mathcal{G}$  has at least one loop edge. Hence there is at least one symbol  $e \in B$  such that  $e^{(1)}$  and  $e^{(2)}$  are loop edges of  $\mathcal{G}$ . Let  $B'$  be the subset of symbols  $e$  in  $B$  such that  $e^{(1)}$  (and thus  $e^{(2)}$ ) are loop edges. Let  $e \in B'$ . Since  $\omega e^{(1)\omega}$  has period 1 we have  $\Phi(\omega e^{(1)\omega})$  has period 1. Hence, as  $a^{(2)}a^{(2)}$  is forbidden in  $X_{A_3}$  for any  $a \in A$ ,

$$\begin{aligned} \Phi(\omega e^{(1)\omega}) &= \omega a^{(1)\omega} \text{ or } \Phi(\omega e^{(1)\omega}) = \omega a^{(3)\omega} \text{ for some } a \in A, \\ \Phi(\omega e^{(2)\omega}) &= \omega b^{(1)\omega} \text{ or } \Phi(\omega e^{(2)\omega}) = \omega b^{(3)\omega} \text{ for some } b \in A. \end{aligned}$$

If  $\Phi(\omega e^{(1)\omega}) = \omega a^{(1)\omega}$  and  $\Phi(\omega e^{(2)\omega}) = \omega b^{(3)\omega}$  for some  $a, b \in A$ , then, since  $\Phi$  is a sliding block code,

$$\Phi(\omega e^{(1)}.e^{(2)\omega}) = \omega a^{(1)}w.w'b^{(3)\omega},$$

where  $w, w'$  are finite words, a contradiction since this  $\omega a^{(1)}w.w'b^{(3)\omega}$  is not in  $X_{A_3}$  even if  $a = b$ , as  $ww'$  contains only a finite number of symbols in  $A^{(2)}$ .

If  $\Phi(\omega e^{(1)\omega}) = \omega a^{(3)\omega}$  and  $\Phi(\omega e^{(2)\omega}) = \omega b^{(1)\omega}$  for some  $a, b \in A$ , then, since  $\Phi$  is a sliding block code,

$$\Phi(\omega e^{(2)}.e^{(1)\omega}) = \omega b^{(1)}w.w'a^{(3)\omega},$$

where  $w, w'$  are finite words, a contradiction since this  $\omega b^{(1)}w.w'a^{(3)\omega}$  is not in  $X_{A_3}$  even if  $a = b$ .

It follows that

$$\Phi(\omega e^{(1)\omega}) = \omega a^{(1)\omega} \text{ and } \Phi(\omega e^{(2)\omega}) = \omega b^{(1)\omega} \text{ for some } a, b \in A, \quad (1)$$

or

$$\Phi(\omega e^{(1)\omega}) = \omega a^{(3)\omega} \text{ and } \Phi(\omega e^{(2)\omega}) = \omega b^{(3)\omega} \text{ for some } a, b \in A. \quad (2)$$

Let us denote by  $B'_1$  the subset of symbols  $e$  in  $B'$  satisfying Equations 1, and by  $B'_3$  the subset of symbols  $e$  in  $B'$  satisfying Equations 2. Let  $e \in B'_1$  and  $f \in B'_3$ . Since  $X_G$  is irreducible, there is a finite path  $w \cdot w'$  such that  $e^{(2)}ww'f^{(1)}$  is a block of  $X_G$ . Then

$$\Phi(\omega e^{(2)}w.w'f^{(1)\omega}) = \omega a^{(1)}w_1.w_2b^{(3)\omega},$$

where  $a, b \in A$  and where  $w_1, w_2$  are finite words. This gives a contradiction since  $\omega a^{(1)}w_1.w_2b^{(3)\omega}$  is not in  $X_{A_3}$  even if  $a = b$ .

As a consequence  $B' = B'_1$  or  $B' = B'_3$ . In the former case  $\omega a^{(3)\omega}$  has no pre-image by  $\Phi$  for  $a \in A$ . In the latter case  $\omega a^{(1)\omega}$  has no pre-image by  $\Phi$  for  $a \in A$ , which ends the proof.  $\square$

The proof can be generalized to show that  $X_{(k,n)}$  is not conjugate to a Markov-Dyck shift when  $k > 2$ .

We now show that  $k$ -graph shifts can be seen as sofic-Dyck shifts. This result is obtained by expressing the nesting constraints of arity  $k$  as another regular nesting constraints of arity 2.

Sofic-Dyck shifts may be defined as follows. We consider an alphabet  $B$  which is a disjoint union of three finite sets of letters, the set  $B_c$  of *call letters*, the set  $B_r$  of *return letters*, and the set  $B_i$  of *internal letters*. The set  $B = B_c \sqcup B_r \sqcup B_i$  is called a *pushdown alphabet*. A *Dyck word over  $B$*  is a word  $w$  generated by the grammar  $D \rightarrow \varepsilon \mid iD \mid cDrD$ , where  $D$  is a variable and  $c \in B_c, r \in B_r, i \in B_i$  are terminal symbols of the grammar. We denote by  $\text{Dyck}(B)$  the set of Dyck words over  $B$ .

A (*finite*) *Dyck automaton*  $\mathcal{A}$  over  $B$  is a pair  $(G, M)$  of a directed labeled graph  $G = (V, E)$  over  $B$  where  $V$  is the finite set of states,  $E \subseteq V \times B \times V$  is the set of edges, and of a set  $M$  of pairs of edges  $((p, a, q), (r, b, s))$  such that  $a \in B_c$  and  $b \in B_r$ . The set  $M$  is called the set of *matched edges*.

A finite path  $\pi$  of  $\mathcal{A}$  is said to be an *admissible path* if for any factor  $(p, a, q) \cdot \pi_1 \cdot (r, b, s)$  of  $\pi$  with  $a \in B_c, b \in B_r$  and the label of  $\pi_1$  being a Dyck word over  $B$ ,  $((p, a, q), (r, b, s))$  is a matched pair. Hence any path of length zero is



admissible and factors of finite admissible paths are admissible. A bi-infinite path is *admissible* if all its finite factors are admissible.

A *sofic-Dyck shift* over  $B$  is the set labels of bi-infinite admissible paths of a Dyck automaton over  $B$ .

**Proposition 2.** *Every  $k$ -graph shift is conjugate to a sofic-Dyck shift over some pushdown alphabet.*

*Proof.* Let  $X \subseteq E^{\mathbb{Z}}$  be a  $k$ -graph shift defined by a  $k$ -graph  $\mathcal{G} = (V, E, A, k)$ . We define the pushdown alphabet  $B = B_c \sqcup B_i \sqcup B_r$  with

$$\begin{aligned} B_c &= \{e^{(1)} \mid e \in A\}, \\ B_i &= \{e^{(i)} \mid e \in A, 1 < i < k\}, \\ B_r &= \{e^{(k)} \mid e \in A\}. \end{aligned}$$

We define a Dyck automaton  $\mathcal{A} = (G', M)$  over  $B$  where  $G' = (V', E')$  as follows. The set  $V'$  is the set  $\{q_{e^{(i)}} \mid e \in A, i \in [k]\}$ . For any  $e, f \in A, i \in [k], 1 \leq j < k$ , we set

$$\begin{aligned} q_{e^{(i)}} &\xrightarrow{f^{(1)}} q_{f^{(1)}} \in E' \text{ if and only if } e^{(i)} f^{(1)} \text{ is a path in } \mathcal{G} \\ q_{e^{(j)}} &\xrightarrow{e^{(j+1)}} q_{e^{(j+1)}} \in E' \quad (\text{as } e^{(j)} e^{(j+1)} \text{ is a path in } \mathcal{G}) \\ q_{e^{(k-1)}} &\xrightarrow{e^{(k)}} q_{f^{(i)}} \in E' \text{ if and only if } e^{(k)} f^{(i)} \text{ is a path in } \mathcal{G} \end{aligned}$$

Each edge  $(q, e^{(1)}, q_{e^{(1)}})$  is matched with  $(q_{e^{(k-1)}}, e^{(k)}, q)$  for each  $q \in V'$ .

Let  $Y$  be the sofic-Dyck shift defined by  $\mathcal{A}$ . It may be checked that the construction implies  $X = Y$ .  $\square$

## 5. Zeta function of shifts defined by $k$ -graphs

### 5.1. Multivariate zeta functions

Recall the notion of multivariate zeta function introduced by Berstel and Reutenauer in [7].

For  $K = \mathbb{Z}$  or  $K = \mathbb{N}$  we denote by  $K\langle\langle A \rangle\rangle$  the set of noncommutative formal power series over the alphabet  $A$  with coefficients in  $K$ . For each language  $L$  of finite words over a finite alphabet  $A$  we define the *characteristic series* of  $L$  as the series  $\underline{L} = \sum_{u \in L} u$  in  $\mathbb{N}\langle\langle A \rangle\rangle$ .

Let  $K[[A]]$  be the usual commutative algebra of formal power series in the variables of  $A$  and  $\pi: K\langle\langle A \rangle\rangle \rightarrow K[[A]]$  be the natural homomorphism. Let  $S$  be a commutative or noncommutative series. One can write  $S = \sum_{n \geq 0} [S]_n$  where each  $[S]_n$  is the homogeneous part of  $S$  of degree  $n$ . The notation extends to matrices  $H$  with coefficients in  $K\langle\langle A \rangle\rangle$  or  $K[[A]]$  with  $([H]_n)_{pq} = [H_{pq}]_n$ , where  $p, q$  are indices of  $H$ .

Call *periodic pattern* of a shift  $X$  a word  $u$  such that the bi-infinite concatenation of  $u$  belongs to  $X$  and denote  $\mathcal{P}(X)$  the set of periodic patterns of  $X$ .

These definitions are extended to  $\sigma$ -invariant sets of bi-infinite sequences which may not be shifts.

The *multivariate zeta function*  $Z(X)$  of a  $\sigma$ -invariant set  $X$  is the commutative series in  $\mathbb{Z}[[A]]$

$$Z(X) = \exp \sum_{n \geq 1} \frac{[\mathcal{P}(X)]_n}{n}.$$

The (*ordinary*) *zeta function* of a language  $X$  is

$$\zeta_X(z) = \exp \sum_{n \geq 1} p_n \frac{z^n}{n},$$

where  $p_n$  is the number of sequences of  $X$  of period  $n$ , *i.e.* of sequences  $x$  such that  $\sigma^n(x) = x$ .

Let  $\theta: \mathbb{Z}[[A]] \rightarrow \mathbb{Z}[[z]]$  be the homomorphism such that  $\theta(a) = z$  for any letter  $a \in A$ . If  $S \in \mathbb{Z}[[A]]$ ,  $\theta(S)$  will also be denoted by  $S(z)$ . Note that  $\zeta_X(z) = \theta(Z(X))$ .

It is known that the multivariate zeta function of a shift has nonnegative integer coefficients. The *entropy* of a language  $L$  is  $h(L) = \limsup_{n \rightarrow \infty} 1/n \log |L \cap A^n|$ . The *entropy* of a shift  $X$  is  $h(\mathcal{B}(X))$ . The entropy  $h(\mathcal{P}(X))$  of the set of periodic patterns of a shift  $X$  is  $\log(1/\rho)$  where  $\rho$  is the radius of convergence of  $\zeta_X(z)$ .

## 5.2. Encoding of periodic sequences of a $k$ -graph shift

Let  $A$  be a finite alphabet and  $A^{(i)} = \{e^{(i)} \mid e \in A\}$  for  $1 \leq i \leq k$ .

We say that a Dyck word  $w$  over  $A_k$  is *prime* if it cannot be decomposed into a product of strictly shorter nonempty Dyck words. Note that the empty word is a Dyck word but not prime. We denote by  $\text{Prime}(A_k)$  the set of prime Dyck words over  $A_k$  and by  $\text{Prime}(X)$  the set of prime Dyck factors of a shift  $X$ .

Let  $\mathcal{G} = (V, E, A, k)$  be a  $k$ -graph defining  $X = X_{\mathcal{G}}$ . The paths in  $\text{Prime}(X)$  are called *prime Dyck paths* of  $\mathcal{G}$ . For each  $p \in V$ , let us denote by  $D_p$  the set of Dyck paths of  $\mathcal{G}$  going from  $p$  to  $p$ . Note that there are no Dyck paths going from  $p$  to  $q$  if  $p \neq q$ . We define the two languages of finite paths of  $\mathcal{G}$

$$L_c = \bigsqcup_{e \in A, 1 \leq i < k} e^{(1)} D_{t(e^{(1)})} \cdots e^{(i-1)} D_{t(e^{(i-1)})} e^{(i)},$$

$$L_r = \bigsqcup_{e \in A, 1 < i \leq k} e^{(i)} D_{t(e^{(i)})} e^{(i+1)} \cdots D_{t(e^{(k-1)})} e^{(k)},$$

and the following  $(V \times V)$ -matrices

- $C = (C_{pq})$ , where  $C_{pq}$  is the set of prime Dyck paths going from  $p$  to  $q$  in  $\mathcal{G}$ . Note that  $C_{pq} = \emptyset$  if  $p \neq q$ . We set  $C_p = C_{pp}$ .
- $C^* = (C_{pq}^*)$ , where  $C_{pq}^*$  is the set of paths going from  $p$  to  $q$  in  $\mathcal{G}$  being concatenation of prime Dyck paths (*i.e.* being Dyck paths).
- $M_c = (M_{c,pq})$ , (resp.  $M_r$ ) where  $M_{c,pq}$  is the set of paths in  $L_c$  (resp. in  $L_r$ ) going from  $p$  to  $q$  in  $\mathcal{G}$ .

Let  $H$  be one of the matrices  $C$ ,  $C^*M_c$ ,  $M_r + C$ . We call an  $H$ -path of  $\mathcal{G}$  a path  $(e_i)_{i \in I}$  of  $\mathcal{G}$ , where  $I$  is an interval or  $\mathbb{Z}$ , which can be factorized into segments in a way that each segment  $e_j \cdots e_{j'} \in H_{s(e_j)t(e_{j'})}$ . We denote by  $X_H$  the  $\sigma$ -invariant set containing all orbits of bi-infinite  $H$ -paths of  $\mathcal{G}$ .

We say that two finite words  $x, y$  are *conjugate* if  $x = uv$  and  $y = vu$  for some words  $u, v$ .

The following proposition gives an encoding of the periodic patterns of a shift defined by a  $k$ -graph. The formula is similar to the one obtained for sofic-Dyck shifts in [4].

**Proposition 3.** *Let  $X$  be the shift defined by a  $k$ -graph  $\mathcal{G}$ . The set of periodic patterns  $\mathcal{P}(X)$  of  $X$  is*

$$\mathcal{P}(X) = \mathcal{P}(X_{C^*M_c}) \sqcup \mathcal{P}(X_{M_r+C}).$$

*Proof.* Let  $z$  be a periodic pattern. Then  $x = \cdots z z z z \cdots$  is a periodic sequence of  $X$ . If  $z$  is not already in  $\text{MC}(X)$  or in  $\text{MR}(X)$ , then its reduced form is

$$\tilde{z} = u \cdot v = (e_1^{(j_1)} \cdots e_1^{(k)}) \cdots (e_n^{(j_n)} \cdots e_n^{(k)}) \cdot (f_1^{(1)} \cdots f_1^{(i_1)}) \cdots (f_m^{(1)} \cdots f_m^{(i_m)}),$$

where  $e_r, f_r \in A$ ,  $1 < j_r \leq k$  and  $1 \leq i_r < k$ . In that case  $z$  has a conjugate  $z'$  whose reduced form is the reduced form of  $v \cdot u$  where

$$v \cdot u = (f_1^{(1)} \cdots f_1^{(i_1)}) \cdots (f_m^{(1)} \cdots f_m^{(i_m)}) \cdot (e_1^{(j_1)} \cdots e_1^{(k)}) \cdots (e_n^{(j_n)} \cdots e_n^{(k)}).$$

This product is either in  $\text{MC}(X)$  or in  $\text{MR}(X)$  since  $\cdots z' z' z' z' \cdots \in X$ . So  $z$  always has a conjugate  $z'$  in  $\text{MC}(X)$  or in  $\text{MR}(X)$ .

If  $z'$  is matched-call, then it is a product of words in  $\text{Prime}(X)$  or of words in  $L_r$ . In that case  $z$  is conjugate to an  $(M_r + C)$ -path of  $\mathcal{G}$ .

If  $z'$  is matched-return and not matched-call, *i.e.*  $z' \notin \text{Prime}(X)^*$ , we can assume that it does not end with a Dyck word (if  $z' = uw$  with  $w$  Dyck, we could consider  $wu$  instead). In that case it is a product of words in  $\text{Prime}(X)^* L_c$  and  $z$  is conjugate to an  $(C^*M_c)$ -path of  $\mathcal{G}$ . As a consequence  $\mathcal{P}(X) = \mathcal{P}(X_{C^*M_c}) \cup \mathcal{P}(X_{M_r+C})$ .

Let us finally show that  $\mathcal{P}(X_{C^*M_c}) \cap \mathcal{P}(X_{M_r+C}) = \emptyset$ . Assume the contrary. Then there are nonempty conjugate words  $w, w'$  such that  $w$  is an  $(C^*M_c)$ -path of  $\mathcal{G}$  and  $w'$  is an  $(M_r + C)$ -path of  $\mathcal{G}$ . This implies that the number of letters in  $A^{(k)}$  minus the number of letters in  $A^{(1)}$  is positive in  $w$  and nonpositive in  $w'$ , contradicting the conjugacy of  $w$  and  $w'$ .  $\square$

### 5.3. Computation of the zeta function

As before, let  $X$  be a  $k$ -graph shift defined by the  $k$ -graph  $\mathcal{G} = (V, E, A, k)$ .

We recall below the notion of circular codes (see for instance [6]). We say that a subset  $S$  of nonempty words over  $A_k$  is a *circular code* if for all  $n, m \geq 1$  and  $x_1, x_2, \dots, x_n \in S$ ,  $y_1, y_2, \dots, y_m \in S$  and  $p \in A_k^*$  and  $s \in A_k^+$ , the equalities  $s x_2 x_3 \cdots x_n p = y_1 y_2 \cdots y_m$  and  $x_1 = ps$  imply  $n = m$ ,  $p = \varepsilon$  and  $x_i = y_i$  for each  $1 \leq i \leq n$ .

This following notion of circular-Markov matrix extends the classical notion of circular codes. It corresponds to the notion of *circular Markov codes* of Keller [11].

We say that the matrix  $(H_{pq})_{p,q \in V}$ , where each  $H_{pq}$  is a set of nonempty words over  $A_k$ , is *circular-Markov* if for all  $n, m \geq 1$  and  $x_i \in H_{p_{i-1}, p_i}$  for  $1 \leq i < n$ ,  $x_n \in H_{p_{n-1}, p_0}$ ,  $y_i \in H_{q_{i-1}, q_i}$  for  $1 \leq i < m$ ,  $y_m \in H_{q_{m-1}, q_0}$  and  $p \in A_k^*$  and  $s \in A_k^+$ , the equalities  $sx_2x_3 \cdots x_n p = y_1y_2 \cdots y_m$  and  $x_1 = ps$  imply  $n = m$ ,  $p = \varepsilon$ ,  $x_i = y_i$  and  $p_i = q_i$  for each  $1 \leq i \leq n$ .

**Proposition 4.** *The sets  $\text{Prime}(A_k)^*L_c$  and  $L_r \sqcup \text{Prime}(A_k)$  are circular codes.*

*Proof.* We first show that  $L_r \sqcup \text{Prime}(A_k)$  is circular. Keeping the notation of the definition, let  $x_1, x_2, \dots, x_n \in S$ ,  $y_1, y_2, \dots, y_m \in S$ ,  $p \in A_k^*$  and  $s \in A_k^+$ . We prove the claim by induction on  $n + m$ . Suppose that  $sx_2x_3 \cdots x_n p = y_1y_2 \cdots y_m$  and  $x_1 = ps$  imply  $n = m$  and  $x_i = y_i$  when  $n + m < N$ . Assume now that  $sx_2x_3 \cdots x_n p = y_1y_2 \cdots y_m$  and  $x_1 = ps$  for some  $n, m$  with  $n + m = N$ .

If  $p$  were nonempty, then, since  $x_1 = ps$  where  $s \neq \varepsilon$ ,  $p$  would have some word  $e^{(i)}w$  with  $i < k$  and  $w \in \text{Dyck}(A_k)$  as a suffix. This would contradict  $p$  being a suffix of  $y_1y_2 \cdots y_m$ , which is clearly matched-return, hence we get  $p = \varepsilon$ . It follows that  $x_1$  is a prefix of  $y_1$  or the converse, implying  $x_1 = y_1$ . By induction hypothesis we obtain that  $n = m$  and  $x_i = y_i$ .

We now show that  $\text{Prime}(A_k)^*L_c$  is circular. Let us assume that  $s \neq x_1$ . Since  $s$  is a prefix of  $y_1y_2 \cdots y_m$  and is a suffix of  $x_1$ , we have  $s \in \text{Prime}(A_k)^*L_c$  and  $p \in \text{Prime}(A_k)^*$ . As  $p \neq \varepsilon$ ,  $p \in \text{Prime}(A_k)^+$  and thus  $p \notin (A_k)^*L_c$ . This contradicts the fact that  $p$  is a suffix of  $y_1 \cdots y_m$ . Hence  $s = x_1$  and  $p = \varepsilon$ . Now  $x_1 \cdots x_n = y_1y_2 \cdots y_m$  implies  $x_1 = y_1$  since  $x_i, y_i \in \text{Prime}(A_k)^*L_c$ . By induction hypothesis we get  $n = m$  and  $x_i = y_i$ .  $\square$

**Corollary 1.** *Let  $\mathcal{G} = (V, E, A, k)$  be a  $k$ -graph defining a shift  $X$ . Let  $C$ ,  $M_c$ ,  $M_r$  be the matrices defined from  $\mathcal{G}$  as above. The matrices  $C$ ,  $C^*M_c$  and  $(M_r + C)$  are circular-Markov.*

*Proof.* Let  $H = C^*M_c$  or  $(M_r + C)$ . Let  $x_i \in H_{p_{i-1}, p_i}$  for  $1 \leq i < n$ ,  $x_n \in H_{p_{n-1}, p_0}$ ,  $y_i \in H_{q_{i-1}, q_i}$  for  $1 \leq i < m$ ,  $y_m \in H_{q_{m-1}, q_0}$  and  $p \in A_k^*$  and  $s \in A_k^+$ . Since  $\text{Prime}(A_k)^*L_c$  and  $L_r \sqcup \text{Prime}(A_k)$  are circular codes, the equalities  $sx_2x_3 \cdots x_n p = y_1y_2 \cdots y_m$  and  $x_1 = ps$  imply  $n = m$ ,  $p = \varepsilon$ ,  $x_i = y_i$  for  $1 \leq i \leq n$ . This implies  $p_i = q_i$  for  $1 \leq i \leq n$  since  $p_i$  is the target of the path  $x_i$  and  $q_i$  is the target of the path  $y_i$ . The matrices  $C^*M_c$  and  $(M_r + C)$  are thus circular-Markov. Since  $M_c + C$  is circular-Markov  $C$  is also circular-Markov.  $\square$

Let  $H$  be a circular-Markov matrix whose coefficients are sets of words over  $A_k$ . Then for any positive integer  $n$  we have  $(\underline{H})^n = \underline{H^n}$  and  $(\underline{H})^* = \underline{H^*} = 1/(1 - \underline{H})$  (see for instance [6]).

The following proposition is a consequence of Keller's formula of the zeta for circular Markov-codes [11].

**Proposition 5.** Let  $\mathcal{G} = (V, E, A, k)$  be a  $k$ -graph defining a shift  $X$ . The multivariate zeta function of  $X$  has the following expression.

$$Z(X) = \frac{1}{\det(I - \pi(\underline{C^*M_c})) \det(I - \pi(\underline{M_r + C}))}.$$

*Proof.* From Proposition 3 we get that the multivariate zeta function of  $X$  is  $Z(X) = Z(X_{C^*M_c})Z(X_{M_r+C})$ . Since  $C^*M_c$  and  $(M_r + C)$  are circular-Markov matrices we get from [11, Theorem 1]

$$Z(X_{C^*M_c}) = \frac{1}{\det(I - \pi(\underline{C^*M_c}))}, \quad Z(X_{C+M_r}) = \frac{1}{\det(I - \pi(\underline{M_r + C}))},$$

hence the proposition.  $\square$

*Example 6.* Let  $X = X_{(k,n)}$  be the shift of  $k$ -nested sequences over  $A_k$  of size  $n$ . Hence  $D_1$  the set of Dyck paths going from 1 to 1 in the unique-vertex graph defining  $X_{A_k}$ . We get

$$\begin{aligned} C &= [C_1], \\ M_r &= [M_{r,11}] = \left[ \sum_{e \in A} e^{(2)} D_1 e^{(3)} D_1 \cdots D_1 e^{(k)} + \cdots + e^{(k-1)} D_1 e^{(k)} + e^{(k)} \right], \\ M_c &= [M_{c,11}] = \left[ \sum_{e \in A} e^{(1)} + e^{(1)} D_1 e^{(2)} + \cdots + e^{(1)} D_1 e^{(2)} D_1 e^{(3)} \cdots D_1 e^{(k-1)} \right], \end{aligned}$$

with

$$\begin{aligned} C_1 &= \sum_{e \in A} e^{(1)} D_1 e^{(2)} D_1 e^{(3)} \cdots D_1 e^{(k)}, \\ D_1 &= \sum_{e \in A} e^{(1)} D_1 e^{(2)} D_1 e^{(3)} \cdots D_1 e^{(k)} D_1 + \varepsilon \end{aligned}$$

Hence

$$Z(X) = (C_1^* M_c)^* (M_r + C_1)^*.$$

We get

$$\begin{aligned} \zeta_X(z) &= \frac{1}{(1 - D_1(z) M_c(z)) (1 - (M_r(z) + C_1(z)))} \\ &= \frac{1}{\left(1 - \frac{nzD_1(z)+1-D_1(z)}{D_1(z)(1-zD_1(z))}\right) \left(1 - z \frac{(n+1)-D_1(z)}{(1-zD_1(z))}\right)}, \end{aligned}$$

where  $D_1(z)$  is the  $\mathbb{N}$ -algebraic series defined by the above equations.

Let  $\rho$  be the radius of convergence of  $\zeta_X(z)$ . We have  $h(\mathcal{P}(X)) = \log(1/\rho)$ . It can be shown that  $h(\mathcal{P}(X)) = h(\mathcal{B}(X))$  and the entropy of  $X$  is thus equal to  $\log(1/\rho)$ . The positive real value  $\rho$  satisfies  $D_1(\rho)M_c(\rho) = 1$  implying  $\rho = 1/(n+1)$ . It follows that the entropy of  $X$  is  $\log(n+1)$  which is independent of  $k$ . We recover the entropy of the Markov-Dyck shift with  $n$  types of parentheses (see [13]).

*Example 7.* Let  $X = X_G$  be the shift defined by the 3-graph  $\mathcal{G}$  of Figure 2. With the above notation we get

$$\begin{aligned} C_1 &= e^{(1)}D_1e^{(2)}D_1e^{(3)} + f^{(1)}D_2f^{(2)}D_2f^{(3)} \\ C_2 &= g^{(1)}D_1g^{(2)}D_1g^{(3)} \\ D_1 &= 1 + e^{(1)}D_1e^{(2)}D_1e^{(3)}D_1 + f^{(1)}D_2f^{(2)}D_2f^{(3)}D_1 \\ D_2 &= 1 + g^{(1)}D_1g^{(2)}D_1g^{(3)}D_2 \end{aligned}$$

$$\begin{aligned} M_c &= \begin{bmatrix} e^{(1)} + e^{(1)}D_1e^{(2)} & f^{(1)} + f^{(1)}D_2f^{(2)} \\ g^{(1)} + g^{(1)}D_1g^{(2)} & 0 \end{bmatrix} \\ M_r &= \begin{bmatrix} e^{(2)}D_1e^{(3)} + e^{(3)} & g^{(2)}D_1g^{(3)} + g^{(3)} \\ f^{(2)}D_2f^{(3)} + f^{(3)} & 0 \end{bmatrix} \end{aligned}$$

Thus  $\zeta_X(z) = \det(I - DM_c(z))^{-1} \det(I - (M_r + C)(z))^{-1}$ , where

$$\begin{aligned} \det(I - DM_c(z)) &= (1 - A_1(1 + A_1) - A_1A_2(1 + A_2)(1 + A_1))(z), \\ \det(I - (M_r + C)(z)) &= (1 - z(1 + A_1 + A_1^2 + A_2^2)(z))(1 - zA_1^2(z)) \\ &\quad - z^2((1 + A_1)(1 + A_2)(z)), \end{aligned}$$

and  $A_1(z) = zD_1(z)$ ,  $A_2(z) = zD_2(z)$ . Since  $D_1(z)$  and  $D_2(z)$  are defined by the system of equations

$$\begin{aligned} D_1(z) &= 1 + z^3D_1(z)(D_1^2(z) + D_2^2(z)) \\ D_2(z) &= 1 + z^3D_1^2(z)D_2(z) \end{aligned}$$

There is indeed a unique pair of series  $D_1(z), D_2(z)$  with nonnegative integer coefficients solution of the above system (see for instance [9] or [5]).

## 6. Shifts defined by $\mathbf{v}$ -graphs

In this section we generalize the notion of  $k$ -graphs by allowing different symbols from the alphabet to have different arities of the associated nesting constraints. We define  $\mathbf{v}$ -graphs where  $\mathbf{v}$  is a finite vector giving the number of symbols with any given arity of nesting constraint. We define shifts presented by these structures.

Let  $I = [m]$  where  $m$  is a positive integer. Let  $\mathcal{G}_i = (V, E_i, A_i, k_i)$  be  $k_i$ -graphs on the same set of vertices  $V$  for  $i \in I$  with  $|A_i| = n_i$ . Note that the notation differs slightly from the previous sections.  $A_i$  here corresponds to what was typically called  $A$  before (not  $A_k$ ). We set  $\mathcal{G} = (V, E, A, \mathbf{v})$  where  $E = \sqcup_{i \in I} E_i$ ,  $A = \sqcup_{i \in I} A_i$ ,  $\mathbf{v} = ((k_i, n_i))_{i \in I}$ .

We consider the free monoid  $S$  generated by  $E$  with a zero  $0$  quotiented by the following relations

- For  $e \in A_i, f \in A_{i'}$  and  $e \neq f$ ,  $e^{(j)}f^{(j')} = 0$  if  $j < k_i$  and  $j' \neq 1$ .
- For  $e \in A_i$ ,  $e^{(j)}e^{(j')} = 0$  if  $j < k_i, j' \neq j + 1$  and  $j' \neq 1$ .
- For  $e \in A_i$ ,  $e^{(1)}e^{(2)}\dots e^{(k_i)} = \mathbf{1}$ .

In the last relation  $\mathbf{1}$  stands for the unity of the monoid.

A *forbidden finite path* of  $\mathcal{G}$  is a finite path which is zero in  $S$ . A *Dyck path* of  $\mathcal{G}$  is a finite path which reduces to  $\mathbf{1}$ . The set of bi-infinite paths in  $\mathcal{G}$  avoiding finite forbidden paths as factors is a shift denoted by  $X_{\mathcal{G}}$  and is called a *v-graph shift*. The multigraph  $\mathcal{G}$  defines  $X_{\mathcal{G}}$ . If  $\mathbf{v} = ((k_i, n_i))_{i \in I}$ , we denote by  $X_{\mathbf{v}}$  the *v-graph shift* defined by a graph having a unique state 1 and edges  $e^{(j)}$  from 1 to 1 for each  $j \in [k_i]$  and each  $e \in A_i$ .

The computation of zeta functions of *v-graph shifts* is similar to the computation of zeta functions of *k-graph shifts*. The topological entropy of  $X_{\mathbf{v}}$  is  $\log(N + 1)$  where  $N = \sum_{i \in I} n_i$ .

*Example 8.* The shift  $X_{\mathbf{v}}$  with  $\mathbf{v} = ((2, 1), (3, 1))$  is defined by the graph  $\mathcal{G}$  of Figure 3. For instance if we denote  $A_1 = \{e\}$  and  $A_2 = \{f\}$  then  $e^{(1)}f^{(1)}f^{(2)}f^{(3)}e^{(2)}$  is Dyck path of  $X_{\mathbf{v}}$ .

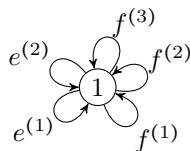


Figure 3: The *v-graph shift* graph  $X_{\mathbf{v}}$  where  $\mathbf{v} = ((2, 1), (3, 1))$ .

## 7. Conclusion

We have defined a class of shifts of sequences which satisfies nesting constraints of arity greater than 2 and is close to the class of Markov-Dyck shifts. Since these shifts are conjugate to sofic-Dyck shifts they may also be characterized by a forbidden regular set of nested words with a nesting arity equal to 2. We have shown how to compute the zeta function for this class of shifts avoiding the increase of complexity due to their sofic-Dyck nature.

## Acknowledgements

The authors thank Wolfgang Krieger and anonymous reviewers for very helpful comments.

## References

- [1] Alur, R., Madhusudan, P., 2009. Adding nesting structure to words. *J. ACM* 56 (3).
- [2] Béal, M.-P., Blockelet, M., Dima, C., 2014. Sofic-Dyck shifts. In: 39th International Symposium on Mathematical Foundations of Computer Science MFCS 2014, (MFCS 2014). Lecture Notes in Computer Science. Springer-Verlag.
- [3] Béal, M.-P., Blockelet, M., Dima, C., 2016. Sofic-Dyck shifts. *Theor. Comput. Sci.* 609, 226–244.
- [4] Béal, M.-P., Dima, C., Heller, P., 2015. N-algebraicity of zeta functions of sofic-Dyck shifts. CoRR <http://arxiv.org/1501.05843>.
- [5] Bell, J. P., Burris, S., Yeats, K. A., 2010. Characteristic points of recursive systems. *Electr. J. Comb.* 17 (1).
- [6] Berstel, J., Perrin, D., Reutenauer, C., 2010. Codes and Automata. Vol. 129 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge.
- [7] Berstel, J., Reutenauer, C., 1990. Zeta functions of formal languages. *Trans. Amer. Math. Soc.* 321, 533–546.
- [8] Blanchard, F., Hansel, G., 1986. Systèmes codés. *Theor. Comput. Sci.* 44, 17–49.
- [9] Flajolet, P., Sedgewick, R., 2009. Analytic Combinatorics. Cambridge University Press, Cambridge.
- [10] Inoue, K., Krieger, W., 2010. Subshifts from sofic shifts and Dyck shifts, zeta functions and topological entropy. CoRRAbs/1001.1839.
- [11] Keller, G., 1991. Circular codes, loop counting, and zeta-functions. *J. Combin. Theory Ser. A* 56 (1), 75–83.
- [12] Krieger, W., 1974/75. On the uniqueness of the equilibrium state. *Math. Systems Theory* 8 (2), 97–104.
- [13] Krieger, W., Matsumoto, K., 2011. Zeta functions and topological entropy of the Markov-Dyck shifts. *Münster J. Math.* 4, 171–183.
- [14] Lind, D., Marcus, B., 1995. An Introduction to Symbolic Dynamics and Coding. Cambridge University Press, Cambridge.
- [15] Matsumoto, K., 2007. On the simple  $C^*$ -algebras arising from Dyck systems. *J. Operator Theory* 58 (1), 205–226.
- [16] Matsumoto, K., 2016. On the Markov-Dyck shifts of vertex type. *Discrete Contin. Dyn. Syst.* 36 (1), 403–422.



- [17] Reutenauer, C., 1997.  $\mathbb{N}$ -rationality of zeta functions. *Adv. in Appl. Math.* 18 (1), 1–17.