Generalized Dyck Shifts

Marie-Pierre Béal, Pavel Heller

Generalized Dyck Shifts ⋆

Marie-Pierre Béal[1] and Pavel Heller[1]

Université Paris-Est, Laboratoire d'informatique Gaspard-Monge, UMR 8049 CNRS
beal@u-pem.frpavel.heller@u-pem.fr

Abstract. We introduce a new class of subshifts of sequences, called generalized Dyck shifts, which extends the class of Dyck shifts introduced by Krieger. The finite factors of these shifts are factors of generalized Dyck words. Generalized Dyck words were introduced by Labelle and Yeh who exhibited unambiguous algebraic grammars generating these context-free languages. Other unambiguous algebraic grammars for generalized Dyck languages were found by Duchon. We define a coding of periodic patterns of generalized Dyck shifts which allows to compute their zeta function. We prove that the zeta function of a generalized Dyck shift is the commutative image of the generating function of an unambiguous context-free language and is thus an ℕ-algebraic series.

1 Introduction

The Dyck shift introduced by Krieger in [9] is the set of bi-infinite sequences of symbols whose finite factors are factors of Dyck words, or well-parenthesized words. To be well-parenthesized, a word needs to have exactly as many opening parentheses (represented here by the letter a) as closing parentheses (represented by the letter b) with the added condition that each opening parenthesis is matched with a closing parenthesis. If one gives the height value $+1$ to the letter a and the height value -1 to the letter b, this condition means that the total height of a Dyck word is 0 and the height of each prefix of a Dyck word is nonnegative.

Dyck shifts are symbolic dynamical systems which are not sofic and belong to larger classes of shifts like Markov-Dyck shifts (see [13], [10]), or sofic-Dyck shifts (see [1]).

In [11] Labelle and Yeh introduced the notion of generalized Dyck words where potentially a larger set of height values are used. They proved the unambiguous context-free nature of generalized Dyck words and exhibited unambiguous context-free grammars for these languages. In [6], Duchon gave new unambiguous context-free grammars for them. Generalized Dyck words were also studied from the point of view of Lyndon words by Melançon and Jacquet in [7].

⋆ This work is supported by the French National Agency (ANR) through "Programme d'Investissements d'Avenir" (Project ACRONYME no. ANR-10-LABX-58) and by the region of Île-de-France through the DIM RDM-IdF

In this paper we show how to define a shift from generalized Dyck words. This shift is called a generalized Dyck shift. We assign class values to letters (for instance, we assign the class value α to the letters a and b). In order to get a nontrivial shift we need to have at least two class values. Generalized Dyck words and factor-free generalized Dyck words are defined recursively as follows. Factor-free generalized Dyck words are the nonempty sequences w of letters *of a same class* such that $h(w) = 0, h(w_1) > 0$ for any proper prefix w_1 of w, and which have no proper factor with these properties. This includes the letters of height 0. Generalized Dyck words are defined as either the empty word or sequences $a_1 d_1 \cdots a_k d_k$ where d_i are generalized Dyck words and $a_1 \cdots a_k$ is a factor-free generalized Dyck word. The generalized Dyck shift is the set of bi-infinite sequences of symbols whose finite factors are factors of generalized Dyck words. These shifts extend the Dyck shift of Krieger.

We give a computation of the zeta function of generalized Dyck shifts which counts the periodic sequences of the shift. We prove that the multivariate zeta function of a generalized Dyck shift is the commutative image of a product of the generating series of the stars of unambiguous context-free circular codes, the codes being cyclically disjoint. The result is based on an encoding of the periodic patterns of the shift. As a consequence the zeta function of a generalized Dyck shift is an $\mathbb{N}$-algebraic series.

Section 2 provides some background on shifts. In Section 3 we define the notions of generalized Dyck words and generalized Dyck shifts. We give unambiguous context-free grammars generating several languages linked to generalized Dyck words. The computation of the multivariate and ordinary zeta functions of a generalized Dyck shift is given in Section 4. This section contains the decomposition of the multivariate zeta function of a generalized Dyck shift into the commutative image of a product of the generating series of the stars of two unambiguous context-free circular codes.

2 Background on shifts

We refer to [12] for basic notions in symbolic dynamics. Let A be a finite alphabet. We denote by A^* the set of words over A and by A^+ the set of nonempty words over A.

A *factor* of a word w is a word u such that $w = vuz$ for some words v, z. A *proper factor* of a word w is a factor distinct from w and the empty word.

A *shift* of sequences X is defined as the set X_F 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. We denote by $\mathcal{B}(X)$ the set of finite blocks of X, that is the set of allowed finite factors of 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* σ, 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 *σ-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 $\sigma^p(x) = x$.

A *(topological) conjugacy* from $X \subseteq A^{\mathbb{Z}}$ to $Y \subseteq B^{\mathbb{Z}}$ is a bijective continuous map from X onto Y which commutes with the shift transformation. Observe that a conjugacy preserves the periods of a sequence.

3 Generalized Dyck words

In this paper, we consider a finite alphabet $A \subset \mathbb{Z} \times \Sigma$, where Σ is a finite alphabet, equipped with two functions: a *height* function h from A to $\mathbb{Z}$ and a *class* function c from A to Σ. Letters with positive height will be denoted by A_+ and letters with negative height by A_-. The set of letters of class α will be denoted by A_α. The set of letters of class α with positive (resp. negative) heigth is denoted by $A_{\alpha,+}$ (resp. $A_{\alpha,-}$). We assume that all sets A_α have both letters with a positive and with a negative height. We set $(i_\alpha, \alpha) \in A_{\alpha,+}$ and $(-j_\alpha, \alpha) \in A_{\alpha,-}$. The height of a nonempty word is the sum of the height of its letters. The height of the empty word is 0.

A *factor-free generalized Dyck word* is a nonempty sequence w of letters *of a same class* such that $h(w) = 0, h(w_1) > 0$ for any proper prefix w_1 of w, and which has no proper factor with these properties. This includes the letters of height 0. Note that it is a sequence of letters in a same class. We denote by $\tilde{D}_\alpha$ the set of factor-free generalized Dyck words in A_α^+ and $\tilde{D} = \sqcup_\alpha \tilde{D}_\alpha$. A *generalized Dyck word* is defined recursively as follows. It is either the empty word or a sequence $a_1 d_1 \cdots a_k d_k$ where each d_i is a generalized Dyck word and $a_1 \cdots a_k$ is a factor-free generalized Dyck word, or a concatenation of generalized Dyck words. We denote by D_α the set of generalized Dyck words built from factor-free sequences $a_1 \cdots a_k$ in A_α^+. We denote by D the set of generalized Dyck words. Note that $D = \cup_\alpha D_\alpha \cup \{\varepsilon\}$.

Hence a nonempty generalized Dyck word can be obtained by inserting after each letter of a factor-free generalized Dyck word, other generalized Dyck words. Further (see [6, Theorem 7]) this decomposition is unique.

A word is *factor-free* if no proper factor of this word belongs to D. The set of factor-free words of a language L is denoted by $\tilde{L}$. Generalized Dyck words (resp. factor-free generalized Dyck words) will be simply called Dyck words (resp. factor-free Dyck words).

A *prime Dyck word* over A is a Dyck word which is not empty and not the product of shorter Dyck words. Note that the empty word is a Dyck word which is not prime. We denote by P the set of prime Dyck words.

Observe that P is a prefix and suffix code. A factor-free Dyck word is prime but the converse is not true. If w is a Dyck word over A then $h(w) = 0$ and $h(w_1) \geq 0$ for each prefix w_1 of w. If w is a prime Dyck word over A then $h(w) = 0$ and $h(w_1) > 0$ for each proper prefix w_1 of w.

Example 1. Let $\Sigma = \{\alpha, \beta\}$ and $A = \{a = (+3, \alpha), b = (-2, \alpha), a' = (+3, \beta), b' = (-2, \beta)\}$. The word $ababb$ is a factor-free Dyck word over A, $a(a'b'a'b'b')b\,(a'b'a'b'b')\,abb$ (see Figure 1) is a prime Dyck word over A which is not factor-free.

The *generalized Dyck shift* over A is the set of bi-infinite sequences whose blocks are factors of a Dyck word over A. We denote this shift by X_A. It is thus a coded system as defined by Blanchard and Hansel [4].

Example 2. If Σ is a singleton the generalized Dyck shift is just the full shift, *i.e.* the set of all bi-infinite sequences over A. So the notion of generalized Dyck shift is interesting only for alphabets Σ of size at least two.

Example 3. If $\Sigma = \{\alpha, \beta\}$ and $A = \{"(" = (+1, \alpha), ")" = (-1, \alpha), "[" = (+1, \beta), "]" = (-1, \beta)\}$, the shift X_A is the Dyck shift with two kinds of parentheses.

Example 4. If $\Sigma = \{\alpha, \beta\}$ and $A = \{a = (+3, \alpha), b = (-2, \alpha), a' = (+3, \beta), b' = (-2, \beta)\}$, for instance the sequences $\cdots bbb.aaaa\cdots$, ${}^{\omega}(aba'b'abbaba'b'b'abb)^{\omega}$ belong to X_A.

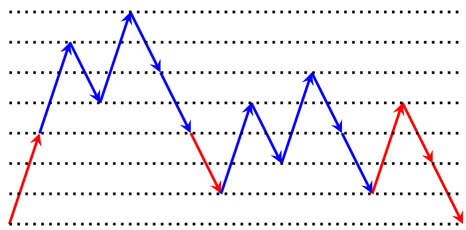

Fig. 1. The prime Dyck word $a(a'b'a'b'b')b(a'b'a'b'b')abb$ of Example 1. Symbols a or a' are represented by up edges while symbols b or b' by down edges according to the height of the symbols. Symbols in A_α (resp. A_β) are represented by red (resp. blue) edges.

Following Duchon [6] we set $m = \max_{a \in A_+} h(a)$, $n = -\min_{a \in A_-} h(a)$. We define for $\alpha \in \Sigma$, $i > 0$, $j > 0$,

- $\tilde{L}_{i,\alpha}$ the set of factor-free words $w \in A_\alpha^+$ with height i such that each proper prefix w_1 of w has a height $h(w_1) > i$.
- $\tilde{R}_{j,\alpha}$ the set of factor-free words $w \in A_\alpha^+$ with height $-j$ such that each proper prefix w_1 of w has a height $h(w_1) > 0$.

- $L_{i,\alpha}$[1] the set of nonempty words $w = a_1 d_1 a_2 \cdots d_{k-1} a_k$ with $k \geq 1$, $d_i \in D$, $a_1 a_2 \ldots a_k \in \tilde{L}_{i,\alpha}$.

[1] The definition of L_i differs here from the one given in [6].

– $R_{j,\alpha}$ the set of nonempty words $w = a_1 d_1 a_2 \cdots d_{k-1} a_k$ with $k \geq 1$, $d_i \in D$, $a_1 a_2 \dots a_k \in \tilde{R}_{j,\alpha}$.
– P_α the set of nonempty words $w = a_1 d_1 a_2 \cdots d_{k-1} a_k$ with $k \geq 1$, $d_i \in D$, $a_1 a_2 \dots a_k \in \tilde{D}_\alpha$.

We set $L_i = \bigcup_\alpha L_{i,\alpha}$, $R_j = \bigcup_\alpha R_{j,\alpha}$, $L = \bigcup_{i=1}^m L_i$, $R = \bigcup_{j=1}^n R_j$. Note that a word in $L_{i,\alpha}$ or $R_{j,\beta}$ does not end nor start with a nonempty Dyck word by definition.

In terms of lattice paths, $L_{i,\alpha}$ is a set of paths that start in $(0,0)$ and end on the line $y = i$ without having a step ending on or under this line before the last step. The set $R_{j,\alpha}$ is a set of paths that start in $(0,0)$ and end on the level $-j$ without having a step ending on or going below the line $y = 0$ before the last step (see Figure 2).

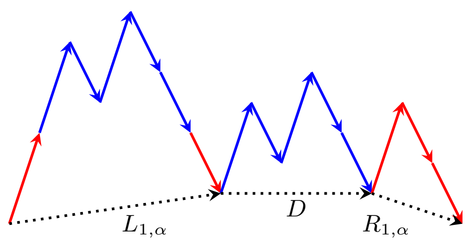

Fig. 2. The prime Dyck word $a(a'b'a'b'b')b(a'b'a'b'b')abb$ of Example 1 in $L_{1,\alpha} D R_{1,\alpha}$.

By definition $L_{i,\alpha}$ and $R_{j,\alpha}$ are codes which are both prefix and suffix. The set $L_{i,\alpha}$ does not overlap strictly with $L_{j,\alpha}$. Indeed, if $uv \in L_{i,\alpha}$ and $vw \in L_{j,\alpha}$, $h(v) < 0$ unless v is the empty word or $v = uv$. A word of $L_{i,\alpha}$ is neither prefix nor suffix of a word in $L_{j,\beta}$ with $\alpha \neq \beta$. Observe also that $L_{i,\alpha}$ is empty for $i > m$ and $R_{j,\alpha}$ is empty for $j > n$.

Lemma 1. *We have the following unambiguous grammars for L_i, R_j, P:*

$$\tilde{L}_{i,\alpha} = \sum_{h(a)=i, c(a)=\alpha} a + \sum_{k>i} \tilde{L}_{k,\alpha} \tilde{R}_{k-i,\alpha} \quad (1)$$

$$\tilde{R}_{j,\alpha} = \sum_{h(a)=-j, c(a)=\alpha} a + \sum_{k} \tilde{L}_{k,\alpha} \tilde{R}_{k+j,\alpha} \quad (2)$$

$$\tilde{D}_\alpha = \sum_{h(a)=0, c(a)=\alpha} a + \sum_{k} \tilde{L}_{k,\alpha} \tilde{R}_{k,\alpha} \quad (3)$$

$$L_{i,\alpha} = \sum_{h(a)=i, c(a)=\alpha} a + \sum_{k>i} L_{k,\alpha} D R_{k-i,\alpha} \quad (4)$$

$$R_{j,\alpha} = \sum_{h(a)=-j, c(a)=\alpha} a + \sum_{k} L_{k,\alpha} D R_{k+j,\alpha} \quad (5)$$

$$P_\alpha = \sum_{h(a)=0} a + \sum_k L_{k,\alpha} D R_{k,\alpha} \tag{6}$$

$$P = \sum_\alpha P_\alpha,\ \ D = P^*,\ \ L_i = \sum_\alpha L_{i,\alpha},\ \ R_j = \sum_\alpha R_{j,\alpha}. \tag{7}$$

Proof. We have $\tilde{L}_{i,\alpha}\tilde{R}_{j,\beta}$ with $\alpha \neq \beta$ forbidden in $\tilde{L}_{i,\alpha}$, $\tilde{R}_{j,\alpha}$, $\tilde{D}_\alpha$. We have $\tilde{L}_{k,\alpha}\tilde{R}_{k-i,\alpha} \subseteq \tilde{L}_{i,\alpha}$ for any $k > i$. If $w \in \tilde{L}_{i,\alpha}$, if $|w| > 1$, let u be the unique proper prefix of w such that $h(u)$ is minimal. Let $h(u) = k > i$ and $w = uv$. Then $u \in \tilde{L}_{k,\alpha}$ and $v \in \tilde{R}_{k-i,\alpha}$. Further, if $uv = u'v'$ with $u \in \tilde{L}_{k,\alpha}$, $v \in \tilde{R}_{k-i,\alpha}$, $u' \in \tilde{L}_{k',\alpha}$, $v' \in \tilde{R}_{k'-i,\alpha}$. One has for instance u prefix of u'. Let $u' = uu$". Then $k \geq k'$. If $k > k'$, then $v \notin \tilde{R}_{k-i,\alpha}$. Thus $k = k'$, $u" \in D$, implying $u" = \varepsilon$. This proves Equation (1). Equations (2), (3) are obtained similarly.

We have $L_{k,\alpha} D R_{k-i,\alpha} \subseteq L_{i,\alpha}$. If $w \in L_{i,\alpha}$ and if $|w| > 1$ let u be the smallest proper prefix of w such that $h(u)$ is minimal and t be the largest proper prefix of w such that $h(t)$ is minimal (see Figure 2). We have $t = uv$ with $v \in D$ and $w = uvz$. Then $u \in L_{k,\alpha}$ and $z \in R_{k-i,\alpha}$. Further, if $uvz = u'v'z'$ with $u \in L_{k,\alpha}$, $z \in R_{k-i,\alpha}$, $u' \in L_{k',\alpha}$, $z' \in R_{k'-i,\alpha}$, $v, v' \in D$, then for instance u is a prefix of u'. Assume that u is a strict prefix of u'. Let $u' = uu''$. Then $k \geq k'$. If $k > k'$, then $vz \notin DR_{k-i,\alpha}$. Thus $k = k'$. Then $u" \in D \setminus \{\varepsilon\}$, a contradiction since $u' \in L_{k',\alpha}$ dos not end with a nonempty word of D. Thus $u = u'$. Similarly, $z = z'$ and thus $v = v'$. This proves Equation (4). Equations (5), (6) are obtained similarly.

We consider the free monoid generated by A with a zero quotiented by the following relations

$$a_1 \cdots a_k = \mathbf{1} \text{ if } a_1 \cdots a_k \text{ is a factor-free Dyck word}$$
$$w = 0,\ \text{if } w \in \tilde{L}_{i,\alpha}\tilde{R}_{j,\beta} \text{ with } \alpha \neq \beta \text{ and } i, j > 0.$$

where $\mathbf{1}$ is the unity of the monoid.

For a word w over A, we denote by $\overline{w} \in A^* \cup \{0, \mathbf{1}\}$ its *reduced form* which is the unique word obtained by applying the above relations.

For instance ( ] reduces to 0 in the Dyck shift.

Observe that a word z in $\tilde{L}_{i,\alpha}\tilde{R}_{j,\beta}$ with $\alpha \neq \beta$ is not factor of a Dyck word. Indeed, if z is a factor of a Dyck word d, it is a factor uv, where $u \in \tilde{L}_{i,\alpha}$, $v \in \tilde{R}_{j,\beta}$, of $a_1 d_1 \cdots a_k d_k$ where $d_i \in D$ and $a_1 \cdots a_k$ is a factor-free Dyck word. If none a_i is a factor of uv, then uv is a factor of some d_i whose length is shorter than d. By recurrence on the size of d we obtain that uv is factor of a factor-free Dyck word. Since $uv \in A_\alpha^+ A_\beta^+$, we get a contradiction. Observe that, since $u \in \tilde{L}_{i,\alpha}$, $v \in \tilde{R}_{j,\beta}$ and $d_i \in D$ for $1 \leq i \leq k$, d_i cannot overlap nor be a factor of uv unless d_i is the empty word. Thus if a_i is factor of u or v for some $1 \leq i \leq k$, then $uv = a_1 \cdots a_k$ and $d_1 = d_2 = \cdots d_{k-1} = \varepsilon$. This gives a contradiction since $uv \in A_\alpha^+ A_\beta^+$ with $\alpha \neq \beta$ and $a_1 \cdots a_k \in A_\gamma^+$ for some γ.

The set of words reducing to $\mathbf{1}$ is the set of Dyck words. If two Dyck words overlap, the overlapping word is a Dyck word: if uv and vw are Dyck words, then u, v, w also. Dyck words. Thus if uv and vw reduce to $\mathbf{1}$, then u, v, w also. Further, a factor-free word has no suffix being a prefix of a word in $\tilde{L}_{i,\alpha}\tilde{R}_{j,\beta}$ with $\alpha \neq \beta$. Hence a word reducing to $\mathbf{1}$ has no non trivial overlap with a word reducing to 0. As a consequence the reduced form is unique.

Proposition 1. *The reduced form of a word w is either 0, $\mathbf{1}$, u, v or uv where*

$$u \in \tilde{R}_{j_s,\beta_s} \cdots \tilde{R}_{j_1,\beta_1}$$
$$v \in \tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r}$$

for some $j_1,\ldots,j_s$, $i_1,\ldots,i_r > 0$ and $\beta_1,\ldots,\beta_s$, $\alpha_1,\ldots,\alpha_r \in \Sigma$.

Proof. Assume that $\overline{w} \neq 0, \mathbf{1}$. Let u be the unique prefix of $\overline{w}$ of minimal height, the unicity coming from the fact that $\overline{w}$ is reduced. This prefix may be the empty word. We set $\overline{w} = uv$. Then v has a unique decomposition into $\tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r}$ where $i_1 + \cdots + i_s = i$ with $h(v) = i > 0$. Indeed let z be the unique nonempty prefix of v such that $h(z) = i_1$ is minimal. The prefix z being reduced, it is a factor-free word. As z does not contain any factor in $\tilde{L}_{i,\alpha}\tilde{R}_{j,\beta}$ with $\alpha \neq \beta$, $i, j > 0$, $z \in \tilde{L}_{i_1,\alpha_1}$ for some $\alpha_1 \in \Sigma$. The whole decomposition of v thus belongs to $\tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r}$ where $i_1 + \cdots + i_r = i$. A symmetrical property holds for u.

Let $uv \in \tilde{R}_{j_s,\beta_s} \cdots \tilde{R}_{j_1,\beta_1}\tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r}$. Then $uv \neq 0$. Indeed, if uv contains a factor z in $\tilde{L}_{i,\alpha}\tilde{R}_{j,\beta}$ with $\alpha \neq \beta$ and $i, j > 0$, then $z = tt'$ where $t \in \tilde{L}_{i,\alpha}$ and $t' \in \tilde{R}_{j,\beta}$. Since t cannot be a suffix of some word in $\tilde{R}_{j_k,\beta_k} \cdots \tilde{R}_{j_{k'},\beta_{k'}}$ it is a suffix of some word in $\tilde{R}_{j_s,\beta_s} \cdots \tilde{R}_{j_1,\beta_1}\tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_{k'},\alpha_{k'}}$ and t' cannot be a prefix of a word in $\tilde{L}_{i_{k'+1},\alpha_{k'+1}} \cdots \tilde{L}_{i_r,\alpha_r}$. Thus $uv \neq 0$. Further a word in $\tilde{L}_{i,\alpha}\tilde{R}_{j,\beta}$ with $\alpha \neq \beta$ and $i, j > 0$ is not factor of a Dyck word, thus $0 \neq \mathbf{1}$.

Words whose reduced form is either 1 or u (resp. v) as above are called *matched-call* (resp. *matched-return*). The set of matched-call (resp. matched-return) words is denoted by $\mathrm{MC}(X)$ (resp. $\mathrm{MR}(X)$). Thus matched-call words are sequences of words in $R+D$ and mached-return words are sequences of words in $L + D$.

Example 5. For instance, the reduced form of the word ") (( ) [[ ]] [" in the Dyck shift with two kinds of parentheses is ") ([". We have $u =$) in $\tilde{R}_{1,\alpha}$ and $v =$ ([in $\tilde{L}_{1,\alpha}\tilde{L}_{1,\beta}$. In the shift of Example 1, the word $ba(a'b'a'b'b')ba$ has the reduced form $b(ab)a \in \tilde{R}_{2,\alpha}\tilde{L}_{1,\alpha}\tilde{L}_{3,\alpha}$.

Proposition 2. *The generalized Dyck shift is the set of sequences avoiding the factors whose reduced form is* 0.

Proof. First a word whose reduced form is 0 is not factor of a Dyck word. Conversely let w be a word over A. Its reduced form is either 0, 1 or u, v, uv with $u \in X = \tilde{R}_{j_s,\beta_s} \cdots \tilde{R}_{j_1,\beta_1}$ and $v \in Y = \tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r}$. Hence it is nonnull if and only if it is a factor of a Dyck word. Indeed any word in X, Y or XY is factor of a Dyck word since $((\tilde{L}_{i_{\beta_1},\beta_1})^{j_1}(\tilde{R}_{j_1,\beta_s})^{i_{\beta_1}-1}) \cdots ((\tilde{L}_{i_{\beta_s},\beta_s})^{j_s}\ (\tilde{R}_{j_s,\beta_s})^{i_{\beta_s}-1})\ X \subseteq D$, $Y((\tilde{L}_{i_r,\alpha_r})^{j_{\alpha_r}-1}\ (\tilde{R}_{j_{\alpha_r},\alpha_r})^{i_r}) \cdots ((\tilde{L}_{i_1,\alpha_1})^{j_{\alpha_1}-1}(\tilde{R}_{j_{\alpha_1},\alpha_1})^{i_1}) \subseteq D$ and $\tilde{L}_{i_{\beta_1},\beta_1}, \ldots, \tilde{R}_{j_{\alpha_1},\alpha_1}$ are non empty. Thus w itself is factor of a Dyck word.

Example 6. We continue with Example 4. Setting in this example $L_i = L_{i,\alpha}$, $R_i = R_{i,\alpha}$, and $L'_i = L_{i,\beta}$, $R'_i = R_{i,\beta}$, $r_i = R_i D$, $\ell_i = L_i D$, we have

$$
\begin{aligned}
L_3 &= a & L'_3 &= a' \\
L_2 &= L_3 D R_1 = a D R_1 & L'_2 &= a' D R'_1 \\
L_1 &= L_2 D R_1 + L_3 D R_2 & L'_1 &= L'_2 D R'_1 + L'_3 D R'_2 \\
R_2 &= b & R'_2 &= b' \\
R_1 &= L_1 D R_2 = L_1 D b & R'_1 &= L'_1 D b' \\
P &= L_1 D R_1 + L_2 D R_2 + L'_1 D R'_1 + L'_2 D R'_2
\end{aligned}
$$

Thus $R_1 = L_1 D b = (a D R_1 D R_1 + a D b) D b = a D((R_1 D)^2 + bD)b$. We can set $U b D = (R_1 D)^2 + bD$ since $R_1 \in A^* b$. We get $r_1 = R_1 D = \ell_1 b D = (aD)U(bD)(bD)$, $\ell_1 = (aD)U(bD)$.

We have

$$
\begin{aligned}
R_1 D &= (aD)U(bD)(bD) = (aD)((R_1 D)^2 + bD)bD \\
&= (aD)(aDUbDbDaDUbDbD + bD)bD.
\end{aligned}
$$

Thus

$$U = \varepsilon + (aD)U(bD)(bD)(aD)U(bD).$$

Similarly,

$$
\begin{aligned}
R'_1 D &= (a'D)V(b'D)(b'D) \\
V &= \varepsilon + (a'D)V(b'D)(b'D)(a'D)V(b'D).
\end{aligned}
$$

We have

$$
\begin{aligned}
PD = {} & (aD)U(bD)(aD)U(bD)(bD) + (aD)(aD)U(bD)(bD)(bD) + \\
& (a'D)V(b'D)(a'D)V(b'D)(b'D) + (a'D)(a'D)V(b'D)(b'D)(b'D).
\end{aligned}
$$

The right symbol D in all the above equations may be removed by right multiplication of both sides by $(1 - P)$ which is the inverse of D.

4 Zeta function of generalized Dyck shifts

4.1 Multivariate zeta functions

We recall the notion of multivariate zeta function introduced by Berstel and Reutenauer in [3], [14].

For $K = \mathbb{Z}$ or $K = \mathbb{N}$ (containing 0) 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 $\pi\colon K\langle\langle A\rangle\rangle \to K[\![A]\!]$ 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 σ-invariant sets of bi-infinite sequences which may not be shifts.

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

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

The *(ordinary) zeta function* of a σ-invariant set 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 $\theta\colon \mathbb{Z}[\![A]\!] \to \mathbb{Z}[\![z]\!]$ 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 [12].

4.2 Encoding of periodic sequences of a generalized Dyck shift

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

If C is a code, we denote by X_C the σ-invariant set containing all bi-infinite concatenation of words in C. This set is not a shift since it may not be closed.

The following proposition gives an encoding of the periodic patterns of a generalized Dyck shift.

Proposition 3. *Let X be the generalized Dyck shift over A. The set of periodic patterns $\mathcal{P}(X)$ of X is*

$$\mathcal{P}(X) = \mathcal{P}(\mathsf{X}_{DL}) \sqcup \mathcal{P}(\mathsf{X}_{R+P}).$$

Proof. Let z be a periodic pattern. Then $x = \cdots zz.zz \cdots$ is a periodic sequence of X. The reduced form of z is $\overline{z} = \mathbf{1}, u, v$ or $u \cdot v$ where

$$u \in \tilde{R}_{j_s,\beta_s} \cdots \tilde{R}_{j_1,\beta_1}$$
$$v \in \tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r}$$

If z is not already in MC(X) or in MR(X), then its reduced form is uv. In this case z has a conjugate z' whose reduced form is the reduced form of vu. We have

$$vu \in \tilde{L}_{i_1,\alpha_1} \cdots \tilde{L}_{i_r,\alpha_r} \tilde{R}_{j_s,\beta_s} \cdots \tilde{R}_{j_1,\beta_1}.$$

Since $\cdots z'z'.z'z' \cdots \in X$, we have $\alpha_r = \beta_s$, and $\tilde{L}_{i_r,\alpha_r}\tilde{R}_{j_s,\beta_s}$ is included in either $L_{i_r-j_s,\alpha_r}$ or $R_{j_s-i_r,\alpha_r}$ or D. In the first case, $\alpha_r = \beta_{s-1}$ and $L_{i_r-j_s,\alpha_r}\tilde{R}_{j_{s-1},\beta_{s-1}}$ is then included in either $L_{i_r-j_s-j_{s-1},\alpha_r}$ or $R_{j_{s-1}-(i_r-j_s),\alpha_r}$ or D. In the second case $\alpha_{r-1} = \beta_s$ and $\tilde{L}_{i_{r-1},\alpha_{r-1}}R_{j_s-i_r,\alpha_r}$ is then included in either $L_{j_s-i_r-i_{r-1},\alpha_r}$ or $R_{i_{r-1}+i_r-j_s,\alpha_r}$ or D. In the third case, $\alpha_{r-1} = \beta_{s-1}$ and $\tilde{L}_{i_{r-1},\alpha_{r-1}}D\tilde{R}_{j_{s-1},\alpha_{r-1}}$ is then included in either $L_{i_{r-1}-j_{s-1},\alpha_{r-1}}$ or $R_{j_{s-1}-i_{r-1},\alpha_{r-1}}$ or D. By iterating the reduction, we get that vu is included in some product equal to either $L_{k_1,\gamma_1} \ldots L_{k_n,\gamma_n}$ or $L_{k_1,\gamma_1} \ldots L_{k_n,\gamma_n}D$ or $R_{k_n,\gamma_n} \ldots R_{k_1,\gamma_1}$ or $DR_{k_n,\gamma_n} \ldots R_{k_1,\gamma_1}$ or D. This product vu is thus either in MC(X) or in MR(X).

If z' is matched-call, then it is a product of words in P or in R. In this case z is conjugate to a word in $(P+R)^*$.

If z' is matched-return and not matched-call, *i.e.* $z' \notin D$, 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 $P^*L = DL$ and z is conjugate to a word in $(DL)^*$. As a consequence $\mathcal{P}(X) = \mathcal{P}(\mathsf{X}_{DL}) \sqcup \mathcal{P}(\mathsf{X}_{R+P})$.

Let us finally show that $\mathcal{P}(\mathsf{X}_{DL}) \cap \mathcal{P}(\mathsf{X}_{R+P}) = \emptyset$. Assume the contrary. Then there are nonempty conjugate words w, w' such that w is in $(DL)^*$ and w' is in $(R+P)^*$.

This implies that the height of w is positive and the height of w' is nonpositive, contradicting the conjugacy of w and w'.

4.3 Computation of the zeta function

We recall below the notion of circular codes (see for instance [2]). We say that a subset S of nonempty words over A is *a circular code* if for all $n, m \geq 1$ and $x_1, x_2, \ldots, x_n \in S$, $y_1, y_2, \ldots, y_m \in S$ and $p \in A^*$ and $s \in A^+$, the equalities $sx_2x_3 \cdots x_np = y_1y_2 \cdots y_m$ and $x_1 = ps$ imply $n = m$, $p = \varepsilon$ and $x_i = y_i$ for each $1 \leq i \leq n$.

Two codes C_1 and C_2 are *cyclically disjoint* if a word of C_1^* which is conjugate to a word of C_2^*, is empty.

Proposition 4. *The sets DL and $P \sqcup R$ are cyclically disjoint circular codes.*

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

If p was nonempty, then, since $x_1 = ps$ where $s \neq \varepsilon$, we have $h(p) > 0$. This would contradict p being a suffix of $y_1y_2 \cdots y_m$, which is clearly matched-call, 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 $x_i = y_i$.

Let us show that P^*L is circular. Let us assume that $s \neq x_1$. Since s is a prefix of $y_1 y_2 \cdots y_m$ and is a suffix of x_1, we have $s \in P^*L$ and $p \in P^*$. As $p \neq \varepsilon$, $p \in P^+$. This contradicts the fact that p is a suffix of $y_1 \cdots y_m$. Hence $s = x_1$ and $p = \varepsilon$. Now $x_1 \cdots x_n = y_1 y_2 \cdots y_m$ implies $x_1 = y_1$ since $x_i, y_i \in P^*L$. By induction hypothesis we get $n = m$ and $x_i = y_i$.

We now show that DL and $P+R$ are cyclically disjoint. Let $u \in (DL)^*$ and $v \in (P+R)^*$ such that u and v are two nonempty conjugate words. This implies that the height of u is positive and the height of v is nonpositive, contradicting the conjugacy of u and v.

Proposition 5. *Let X be a generalized Dyck shift over A. The multivariate zeta function of X has the following expression.*

$$Z(X) = \pi(\underline{(DL)^*(P+R)^*}).$$

Proof. From Proposition 3 we get that the multivariate zeta function of X is $Z(X) = Z(\mathsf{X}_{DL})Z(\mathsf{X}_{P+R})$.

From [15, Proposition 4.7.11] (see also [2, Proposition 3.1],[8]), if C is a circular code $Z(\mathsf{X}_C) = \pi(\underline{C^*})$. The result follows from the fact that DL and $P+R$ are circular codes.

Example 7. We consider the Dyck shift X with two kinds of parentheses of Example 3 defined by $\Sigma = \{\alpha, \beta\}$ and $A = \{\text{"("} = (+1, \alpha), \text{")"} = (-1, \alpha), \text{"["} = (+1, \beta), \text{"]"} = (-1, \beta)\}$.

Setting $a = \text{"("}$, $b = \text{")"}$, $a' = \text{"["}$, $b' = \text{"]"}$, $L_1 = L_{1,\alpha}$, $L'_1 = L_{1,\beta}$, $R_1 = R_{1,\alpha}$, $R'_1 = R_{1,\beta}$, we have

$$L_1 = a \quad L'_1 = a' \tag{8}$$
$$R_1 = b \quad R'_1 = b' \tag{9}$$
$$P = L_1 D R_1 + L'_1 D R'_1 = aDb + a'Db' \tag{10}$$
$$D = 1 + PD = 1 + aDbD + a'Db'D \tag{11}$$

Thus

$$Z(X) = \pi((D(a+a'))^* \ (b + b' + aDb + a'Db')^*),$$

where D is defined by Equation 11. A computation gives the formula of Keller for $\zeta_X(z)$ [8]:

$$\zeta_X(z) = \frac{2(1+\sqrt{1-8z^2})}{(1-4z+\sqrt{1-8z^2})^2}.$$

Example 8. We consider the shift X_A defined by $\Sigma = \{\alpha, \beta\}$ and $A = \{a = (+2, \alpha), b = (-1, \alpha), a' = (+2, \beta), b' = (-1, \beta)\}$.

Setting $L_i = L_{i,\alpha}$, $L'_i = L_{i,\beta}$, and $R_i = R_{i,\alpha}$, $R'_i = R_{i,\beta}$, we have

$$L_2 = a \quad L'_2 = a' \tag{12}$$
$$R_1 = b \quad R'_1 = b' \tag{13}$$
$$L_1 = L_2 D R_1 = aDb \quad L'_1 = a'Db' \tag{14}$$

$$P = L_1 D R_1 + L'_1 D R'_1 = aDbDb + a'Db'Db' \quad (15)$$
$$D = 1 + PD = 1 + aDbDbD + a'Db'Db'D \quad (16)$$

Thus

$$Z(X) = \pi((D(aDb + a'Db' + a + a'))^* \ (b + b' + aDbDb + a'Db'Db')^*),$$

where D is defined by Equation 16.

Let S be a multivariate series in $\mathbb{N}\langle\langle A\rangle\rangle$. We denote by $< S, u >$ the coefficient of a word u in S. We say that S is $\mathbb{N}$-*algebraic* if $S- < S, \varepsilon > \varepsilon$ is the multivariate generating series of some unambiguous context-free language. The multivariate zeta function of a shift is $\mathbb{N}$-*algebraic* if it is the commutative image of some multivariate $\mathbb{N}$-algebraic series. In one variable, a series $S(z)$ is $\mathbb{N}$-algebraic if it is the first component $(S_1(z))$ of a system of equations $S_i(z) = P_i(z, S_1(z), .., S_r(z))$, where $1 \leq i \leq r$ and P_i are multivariate polynomials with coefficients in $\mathbb{N}$ (see for instance [5]).

Corollary 1. *The multivariate zeta function of a generalized Dyck shift is the commutative image of a product of the generating series of the stars of unambiguous context-free circular codes, the codes being cyclically disjoint. The multivariate and ordinary zeta functions of a generalized Dyck shift are* $\mathbb{N}$*-algebraic series.*

Proof. The result follows from Proposition 5 and the fact that DL and $P \cup R$ are unambiguous context-free circular codes since the languages $P, L_{i,\alpha}, R_{j,\beta}$ are unambiguous context-free. Further DL and $P \cup R$ are cyclically disjoint.

References

1. M.-P. Béal, M. Blockelet, and C. Dima. Sofic-Dyck shifts. *Theor. Comput. Sci.*, 609:226–244, 2016.
2. J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*, volume 129 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.
3. J. Berstel and C. Reutenauer. Zeta functions of formal languages. *Trans. Amer. Math. Soc.*, 321:533–546, 1990.
4. F. Blanchard and G. Hansel. Systèmes codés. *Theor. Comput. Sci.*, 44:17–49, 1986.
5. M. Bousquet-Mélou. Rational and algebraic series in combinatorial enumeration. In *International Congress of Mathematicians. Vol. III*, pages 789–826. Eur. Math. Soc., Zürich, 2006.
6. P. Duchon. On the enumeration and generation of generalized Dyck words. *Discrete Mathematics*, 225(1-3):121–135, 2000.
7. H. Jacquet and G. Mélançon. Langages de Dyck généralisés et factorisations du monoide libre. *Ann. Sci. Math. Québec*, (97):103–122, 1997.
8. G. Keller. Circular codes, loop counting, and zeta-functions. *J. Combin. Theory Ser. A*, 56(1):75–83, 1991.

9. W. Krieger. On the uniqueness of the equilibrium state. *Math. Systems Theory*, 8(2):97–104, 1974/75.
10. W. Krieger and K. Matsumoto. Zeta functions and topological entropy of the Markov-Dyck shifts. *Münster J. Math.*, 4:171–183, 2011.
11. J. Labelle and Y. Yeh. Generalized Dyck paths. *Discrete Mathematics*, 82(1):1–6, 1990.
12. D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.
13. K. Matsumoto. On the simple C^*-algebras arising from Dyck systems. *J. Operator Theory*, 58(1):205–226, 2007.
14. C. Reutenauer. $\mathbb{N}$-rationality of zeta functions. *Adv. in Appl. Math.*, 18(1):1–17, 1997.
15. R. P. Stanley. *Enumerative Combinatorics*. Wadsworth Publ. Co., Belmont, CA, USA, 1986.