

ON FACTOR-FREE DYCK WORDS WITH HALF-INTEGER SLOPE

DANIEL BIRMAJER, JUAN B. GIL, AND MICHAEL D. WEINER

ABSTRACT. We study a class of rational Dyck paths with slope $\frac{2m+1}{2}$ corresponding to factor-free Dyck words, as introduced by P. Duchon. We show that, for the slopes considered in this paper, the language of factor-free Dyck words is generated by an auxiliary language that we examine from the algebraic and combinatorial points of view. We provide a lattice path description of this language, and give an explicit enumeration formula in terms of partial Bell polynomials. As a corollary, we obtain new formulas for the number of associated factor-free generalized Dyck words.

1. INTRODUCTION

In these notes we consider the set of factor-free words belonging to the generalized Dyck language constructed from the alphabet $\mathcal{A} = \{a, b\}$ with valuations $h(a) = 2m + 1$, $m \in \mathbb{N}$, and $h(b) = -2$. This is an instance of the two-letter Dyck language studied by P. Duchon for which words correspond to Dyck paths with rational slope. In the case at hand, the slope is $\frac{2m+1}{2}$.

In [9], Duchon provided an algebraic grammar for generalized Dyck languages (as introduced by Labelle and Yeh [10]) and proved that words in such a language can be obtained uniquely by inserting words of the language into factor-free words of the same language.

Recall that a *factor* of a word w is any word w' such that $w = w_1 w' w_2$. If w_1 and w_2 are not both the empty word, w' is a *proper factor* of w . For a given alphabet \mathcal{A} we let \mathcal{A}^* denote the set of all words made from \mathcal{A} , together with the empty word ε . A word w in \mathcal{A}^* is said to be a *generalized Dyck word* if it satisfies the conditions that $h(w) = 0$, and for each left factor w_1 of w , $h(w_1) \geq 0$. We denote by $\mathcal{D}_{\mathcal{A},h}$ the set of generalized Dyck words over the alphabet \mathcal{A} with valuation given by h . Moreover, we say that a word in \mathcal{A}^* is $\mathcal{D}_{\mathcal{A},h}$ -factor-free (or simply factor-free if the underlying Dyck language is clear) if it has no proper factor belonging to $\mathcal{D}_{\mathcal{A},h}$. The set of factor-free words in $\mathcal{D}_{\mathcal{A},h}$ will be denoted by $D_{\mathcal{A},h}$.

As shown in [9, Section 5], the algebraic grammars for $\mathcal{D}_{\mathcal{A},h}$ and $D_{\mathcal{A},h}$ can be described by a system of derivation rules in terms of certain auxiliary languages with restrictions on their total and partial valuations. As we will see in Section 2, for Dyck words with slope $\frac{2m+1}{2}$, the aforementioned derivation rules may be reduced to a single core language that we denote by U , or $U_{\frac{2m+1}{2}}$ if we wish to emphasize the slope.

The main focus of this paper is to study the auxiliary language U from the algebraic and combinatorial points of view. We provide a description of U in terms of lattice paths and, based on a polynomial equation satisfied by the generating function, we give an explicit enumeration formula involving partial Bell polynomials. As a corollary, we obtain new formulas for the enumeration of the corresponding factor-free generalized Dyck words. In

Section 4 we illustrate our results for slopes $\frac{3}{2}$ and $\frac{5}{2}$. In particular, we discuss a bijection between the elements of $U_{\frac{5}{2}}$ and certain colored trees having nonleaf nodes of outdegrees 2 or 4. In the last section of the paper, we provide the building blocks needed to create factor-free words with slope $\frac{7}{2}$, we give an interesting connection between the auxiliary language U and certain colored Dyck paths, and we briefly discuss the use of factor-free Dyck words to generate cross-bifix-free (non-overlapping) codes of binary words with variable length.

2. GRAMMAR FOR FACTOR-FREE WORDS AND THE AUXILIARY LANGUAGE U

In this section, we will review some of the terminology introduced by Duchon in [9] and will discuss the reduced algebraic grammar for the two-letter sublanguage of factor-free Dyck words with slope $\frac{2m+1}{2}$. In addition, we define the language U alluded to in the introduction and describe its elements in terms of lattice paths.

First, consider the auxiliary languages L_i and R_j defined as follows:

- L_i is the set of factor-free words $w \in \mathcal{A}^*$ with total valuation i , such that each nonempty left factor w_1 of w has $h(w_1) > i$,
- R_j is the set of factor-free words $w \in \mathcal{A}^*$ with total valuation $-j$, such that each nonempty left factor w_1 of w has $h(w_1) > 0$.

Using Duchon's results [9, Section 5], we conclude that if $\mathcal{A} = \{a, b\}$ is a two-letter alphabet with valuations $h(a) = 2m + 1$ and $h(b) = -2$, then the set $D = D_{\mathcal{A}, h}$ of factor-free Dyck words in \mathcal{A}^* may be described via the derivation rules:

$$\begin{aligned} D &= \varepsilon + L_1 R_1 + L_2 R_2, \\ L_i &= L_{i+1} R_1 + L_{i+2} R_2 \quad \text{for } 1 \leq i \leq 2m - 1, \\ L_{2m} &= L_{2m+1} R_1, \quad L_{2m+1} = a, \\ R_1 &= L_1 R_2, \quad R_2 = b, \end{aligned}$$

which can be reduced to

$$\begin{aligned} (1) \quad D &= \varepsilon + L_1 L_1 b + L_2 b, \\ L_{2m+1} &= a, \quad L_{2m} = a L_1 b, \\ L_i &= L_{i+1} L_1 b + L_{i+2} b \quad \text{for } 1 \leq i \leq 2m - 1. \end{aligned}$$

This is an unambiguous context-free grammar.

Definition 1. Let $U = U_{\frac{2m+1}{2}}$ be the set consisting of the empty word ε together with all factor-free words $w \in \mathcal{A}^*$ with total valuation 0, having at least one left factor with negative valuation, and such that each left factor w_1 of w has $h(w_1) > -2m$.

By identifying the letter a with an east-step $(1, 0)$ and the letter b with a north-step $(0, 1)$, each nonempty word in U corresponds to an east-north lattice path from $(0, 0)$ to the line $y = \frac{2m+1}{2}x$ with the following properties:

- ▷ it has no two lattice points on a line with slope $\frac{2m+1}{2}$ such that the path connecting them lies completely below that line (factor-free),
- ▷ it crosses the line $y = \frac{2m+1}{2}x$ at least once,
- ▷ it stays strongly below the line $y = \frac{2m+1}{2}x + m$.

Observe that, factoring $L_1 = aUb^m$, the derivation rules (1) give

$$(2) \quad \begin{aligned} D &= \varepsilon + aUb^m aUb^{m+1} + L_2b, \\ L_{2m+1} &= a, \quad L_{2m} = aaUb^{m+1}, \\ L_i &= L_{i+1}aUb^{m+1} + L_{i+2}b \text{ for } 1 \leq i \leq 2m - 1. \end{aligned}$$

In other words, the set of factor-free Dyck words with slope $\frac{2m+1}{2}$ is completely determined by the auxiliary language U . Note that the length of a word in U is necessarily a multiple of $2m + 3$.

Example 2 (slope $\frac{3}{2}$). If $m = 1$, then the derivation rules (2) become

$$\begin{aligned} D &= \varepsilon + aUbaUbb + L_2b, \\ L_3 &= a, \quad L_2 = aaUbb, \\ L_1 &= L_2aUbb + L_3b, \end{aligned}$$

which yield the equations

$$(3) \quad \begin{aligned} D &= \varepsilon + aUbaUbb + aaUbbb, \\ U &= \varepsilon + aUbbaUb. \end{aligned}$$

In particular, $abbab$ is the only word in U of length 5, and the words in U of length 10 are obtained by inserting $abbab$ into itself according to (3). Thus we obtain the two words $aabbabbab$ and $abbaabbabb$, see Figure 1.

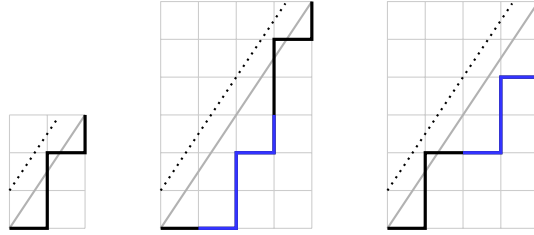


FIGURE 1. Lattice paths corresponding to $abbab$, $aabbabbab$, $abbaabbabb \in U_{\frac{3}{2}}$.

As already pointed out in [9], in the case of slope $\frac{3}{2}$ the elements of U are enumerated by the Catalan numbers.

Example 3 (slope $\frac{5}{2}$). If $m = 2$, then (2) gives

$$\begin{aligned} D &= \varepsilon + aUbbaUbbb + L_2b, \\ L_5 &= a, \quad L_4 = aaUbbb, \quad L_3 = L_4aUbbb + L_5b, \\ L_2 &= L_3aUbbb + L_4b, \quad L_1 = L_2aUbbb + L_3b, \end{aligned}$$

which yield

$$(4) \quad \begin{aligned} D &= \varepsilon + aUbbaUbbb + a(aUbbb)^3b + abaUbbbb + aaUbbbbb, \\ U &= \varepsilon + (aUbbb)^3aUb + baUbbbaUb + aUbbbaUb + aUbbbaUbb. \end{aligned}$$

In particular, $babbbab$, $abbbbab$, and $abbbabb$ are the only words of length 7 in U , and $abbabbbbabbab$ is the only word of length 14 in U that cannot be derived from the three primitive words of length 7, see Figure 2.

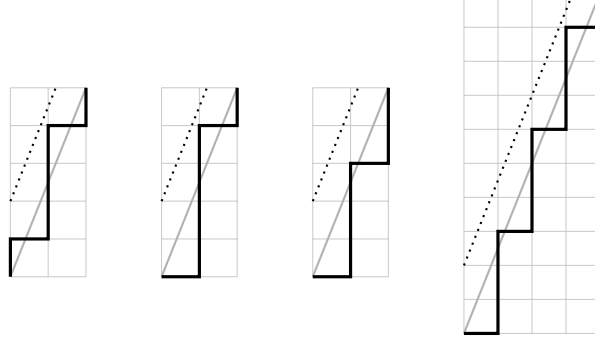


FIGURE 2. Lattice paths for $babbbab$, $abbbbab$, $abbbabb$, $abbabbbbabbab \in U_{\frac{5}{2}}$.

For some enumerative and asymptotic results for rational Dyck paths with slope $\frac{5}{2}$, see [1].

3. ENUMERATION OF U AND FACTOR-FREE $\frac{2m+1}{2}$ -DYCK WORDS

In this section we prove a polynomial equation satisfied by the generating function $\mathcal{U}(\tau)$ of U , and use it to find an enumeration formula in terms of partial Bell polynomials. As a corollary, we obtain formulas for the enumeration of the corresponding factor-free generalized Dyck words.

Recall that the auxiliary languages L_i satisfy:

$$\begin{aligned} L_{2m+1} &= a, \quad L_{2m} = aL_1b, \\ L_i &= L_{i+1}L_1b + L_{i+2}b \quad \text{for } 1 \leq i \leq 2m-1. \end{aligned}$$

This is an unambiguous grammar, so the above equations map to algebraic equations for the corresponding generating functions $\mathcal{L}_i(\tau)$:

$$(5) \quad \begin{aligned} \mathcal{L}_{2m+1}(\tau) &= \tau, \quad \mathcal{L}_{2m}(\tau) = \tau^2 \mathcal{L}_1(\tau), \\ \mathcal{L}_i(\tau) &= \tau \mathcal{L}_1(\tau) \mathcal{L}_{i+1}(\tau) + \tau \mathcal{L}_{i+2}(\tau) \quad \text{for } 1 \leq i \leq 2m-1. \end{aligned}$$

Proposition 4. *The following relations hold:*

$$\mathcal{L}_1(\tau) = \sum_{j=0}^m \binom{m+j}{m-j} \tau^{j+m+1} \mathcal{L}_1^{2j}(\tau) \quad \text{and} \quad \mathcal{L}_2(\tau) = \sum_{j=0}^{m-1} \binom{m+j}{m-j-1} \tau^{j+m+1} \mathcal{L}_1^{2j+1}(\tau).$$

Proof. Let $\ell_i = \mathcal{L}_{2m+1-i}(\tau)$ for $i = 0, \dots, 2m$. Then (5) gives the recurrence relation:

$$\begin{aligned} \ell_0 &= \tau, \quad \ell_1 = \tau^2 \mathcal{L}_1, \\ \ell_i &= (\tau \mathcal{L}_1) \ell_{i-1} + \tau \ell_{i-2} \quad \text{for } 2 \leq i \leq 2m. \end{aligned}$$

Using [5, Prop. 1] and the fact that $B_{n,j}(c_1, 2c_2, 0, \dots) = \frac{n!}{j!} \binom{j}{n-j} c_1^{2j-n} c_2^{n-j}$, we find a closed formula for ℓ_n :

$$\ell_n = \sum_{j \geq n/2}^n \binom{j}{n-j} \tau^{j+1} \mathcal{L}_1^{2j-n}.$$

This implies

$$\mathcal{L}_1 = \ell_{2m} = \sum_{j=m}^{2m} \binom{j}{2m-j} \tau^{j+1} \mathcal{L}_1^{2j-2m} = \sum_{j=0}^m \binom{m+j}{m-j} \tau^{j+m+1} \mathcal{L}_1^{2j},$$

and similarly,

$$\mathcal{L}_2 = \ell_{2m-1} = \sum_{j=0}^{m-1} \binom{m+j}{m-j-1} \tau^{j+m+1} \mathcal{L}_1^{2j+1}.$$

□

Since $L_1 = aUb^m$, we have $\mathcal{L}_1(\tau) = \tau^{m+1}\mathcal{U}(\tau)$, and Proposition 4 gives

$$\tau^{m+1}\mathcal{U}(\tau) = \sum_{j=0}^m \binom{m+j}{m-j} \tau^{j+m+1} (\tau^{m+1}\mathcal{U}(\tau))^{2j}.$$

Thus

$$\mathcal{U}(\tau) = \sum_{j=0}^m \binom{m+j}{m-j} (\tau^{2m+3}\mathcal{U}^2(\tau))^j.$$

Moreover, since the length of any nonempty word of valuation 0 with slope $\frac{2m+1}{2}$ must be a multiple of $2m+3$, the generating function \mathcal{U} is of the form $1 + \sum_{n=1}^{\infty} u_n \tau^{(2m+3)n}$, where u_n denotes the number of words in U of length $(2m+3)n$.

Therefore, with the change of variables $t = \tau^{2m+3}$, and denoting the generating function again by \mathcal{U} , we obtain

$$(6) \quad \mathcal{U}(t) = 1 + \sum_{j=1}^m \binom{m+j}{m-j} t^j \mathcal{U}^{2j}(t).$$

For a fixed $m \in \mathbb{N}$, let $\mu_j = \binom{m+j}{m-j}$ for $j \geq 0$. Note that $\mu_j = 0$ for $j \geq m+1$.

Theorem 5. *If u_n is the coefficient of t^n in $\mathcal{U}(t)$, then*

$$(7) \quad u_n = \sum_{k=1}^n \binom{2n}{k-1} \frac{(k-1)!}{n!} B_{n,k}(1!\mu_1, 2!\mu_2, \dots).$$

Proof. With $X(t) = 1 + \sum_{j=1}^m \mu_j t^j$, identity (6) means $\mathcal{U}(t) = X(t\mathcal{U}^2(t))$. Denote the right-hand side of (7) by v_n and let $\mathcal{V}(t) = 1 + \sum_{n=1}^{\infty} v_n t^n$.

We will prove that $\mathcal{V}(t)$ also satisfies $\mathcal{V}(t) = X(t\mathcal{V}^2(t))$, hence $\mathcal{U}(t) = \mathcal{V}(t)$ and so $u_n = v_n$ for every n .

First, using the identity (cf. [7, Sec. 11.2, Eqn. (11.11)])

$$B_{n,k+1}(z_1, z_2, \dots) = \sum_{\ell=k}^{n-1} \binom{n-1}{\ell} z_{n-\ell} B_{\ell,k}(z_1, z_2, \dots),$$

we have

$$\begin{aligned} n!v_n &= \sum_{k=1}^n \binom{2n}{k-1} (k-1)! B_{n,k}(1!\mu_1, 2!\mu_2, \dots) \\ &= \sum_{k=1}^n \binom{2n}{k-1} (k-1)! \sum_{\ell=k-1}^{n-1} \binom{n-1}{\ell} (n-\ell)! \mu_{n-\ell} B_{\ell,k-1}(1!\mu_1, \dots) \\ &= \sum_{k=0}^{n-1} \binom{2n}{k} k! \sum_{\ell=k}^{n-1} \binom{n-1}{\ell} (n-\ell)! \mu_{n-\ell} B_{\ell,k}(1!\mu_1, \dots) \\ &= \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} (n-\ell)! \mu_{n-\ell} \sum_{k=0}^{\ell} \binom{2n}{k} k! B_{\ell,k}(1!\mu_1, \dots) \\ &= n!\mu_n + \sum_{\ell=1}^{n-1} \frac{n!}{\ell!} \mu_{n-\ell} S_{n,\ell}(1!\mu_1, 2!\mu_2, \dots), \end{aligned}$$

where $S_{n,\ell}(1!\mu_1, 2!\mu_2, \dots) = 2(n-\ell) \sum_{k=1}^{\ell} \binom{2n-1}{k-1} (k-1)! B_{\ell,k}(1!\mu_1, 2!\mu_2, \dots)$.

On the other hand, if (w_n) is defined by $t\mathcal{V}^2(t) = \sum_{n=1}^{\infty} w_n t^n$, then Faà di Bruno's formula (cf. [8, Sec. 3.4, Theorem A]) implies

$$n![t]^n X(t\mathcal{V}^2(t)) = \sum_{\ell=1}^n \ell! \mu_{\ell} B_{n,\ell}(1!w_1, 2!w_2, \dots).$$

Moreover, by Equation (3\ell) in [8, Sec. 3.3], and since $w_1 = 1$, we get

$$\begin{aligned} B_{n,\ell}(1!w_1, 2!w_2, \dots) &= \sum_{k=2\ell-n}^{\ell} \frac{n!}{(n-\ell)!k!} B_{n-\ell,\ell-k}(1!w_2, 2!w_3, \dots) \\ &= \sum_{k=0}^{n-\ell} \frac{n!}{(n-\ell)!(\ell-k)!} B_{n-\ell,k}(1!w_2, 2!w_3, \dots). \end{aligned}$$

Therefore,

$$\begin{aligned} n![t]^n X(t\mathcal{V}^2(t)) &= \sum_{\ell=1}^n \ell! \mu_{\ell} \sum_{k=0}^{n-\ell} \frac{n!}{(n-\ell)!(\ell-k)!} B_{n-\ell,k}(1!w_2, 2!w_3, \dots) \\ &= \sum_{\ell=0}^{n-1} (n-\ell)! \mu_{n-\ell} \sum_{k=0}^{\ell} \frac{n!}{\ell!(n-\ell-k)!} B_{\ell,k}(1!w_2, 2!w_3, \dots) \\ &= n!\mu_n + \sum_{\ell=1}^{n-1} \frac{n!}{\ell!} \mu_{n-\ell} T_{n,\ell}(1!w_2, 2!w_3, \dots), \end{aligned}$$

where $T_{n,\ell}(1!w_2, 2!w_3, \dots) = \sum_{k=1}^{\ell} \binom{n-\ell}{k} k! B_{\ell,k}(1!w_2, 2!w_3, \dots)$.

Now, using [8, Sec. 3.5, Thm. B] we can write $(\mathcal{V}^2(t))^{n-\ell}$ and $\mathcal{V}^{2n-2\ell}(t)$ in terms of partial Bell polynomials as follows:

$$\begin{aligned} (1 + w_2t + w_3t^2 + \dots)^{n-\ell} &= 1 + \sum_{\ell=1}^{\infty} \sum_{k=1}^{\ell} \binom{n-\ell}{k} \frac{k!}{\ell!} B_{\ell,k}(1!w_2, 2!w_3, \dots) t^{\ell}, \\ (1 + v_1t + v_2t^2 + \dots)^{2n-2\ell} &= 1 + \sum_{\ell=1}^{\infty} \sum_{k=1}^{\ell} \binom{2n-2\ell}{k} \frac{k!}{\ell!} B_{\ell,k}(1!v_1, 2!v_2, \dots) t^{\ell}. \end{aligned}$$

This implies $T_{n,\ell}(1!w_2, 2!w_3, \dots) = \ell! [t^{\ell}] (\mathcal{V}^2(t))^{n-\ell} = \ell! [t^{\ell}] \mathcal{V}^{2n-2\ell}(t)$, and so

$$\begin{aligned} T_{n,\ell}(1!w_2, 2!w_3, \dots) &= \sum_{k=1}^{\ell} \binom{2n-2\ell}{k} k! B_{\ell,k}(1!v_1, 2!v_2, \dots) \\ &= 2(n-\ell) \sum_{k=1}^{\ell} \binom{2n-2\ell-1}{k-1} (k-1)! B_{\ell,k}(1!v_1, 2!v_2, \dots). \end{aligned}$$

Finally, applying [4, Theorem 15] with $\lambda = 2n - 2\ell - 1$ on the expression

$$\ell! v_{\ell} = \sum_{k=1}^{\ell} \binom{2\ell}{k-1} (k-1)! B_{\ell,k}(1!\mu_1, 2!\mu_2, \dots),$$

we get

$$\sum_{k=1}^{\ell} \binom{2n-2\ell-1}{k-1} (k-1)! B_{\ell,k}(1!v_1, \dots) = \sum_{k=1}^{\ell} \binom{2n-1}{k-1} (k-1)! B_{\ell,k}(1!\mu_1, \dots).$$

Multiplying both sides by $2(n-\ell)$, we finally obtain

$$T_{n,\ell}(1!w_2, 2!w_3, \dots) = S_{n,\ell}(1!\mu_1, 2!\mu_2, \dots),$$

which implies $\mathcal{V}(t) = X(t\mathcal{V}^2(t))$. □

Let us now address the enumeration of the set D of factor-free Dyck words with slope $\frac{2m+1}{2}$. Let \mathcal{D} denote the generating function of D . As discussed in Section 2, we have

$$D = \varepsilon + L_1^2 b + L_2 b.$$

Thus Proposition 4 together with the factorization $\mathcal{L}_1(\tau) = \tau^{m+1}\mathcal{U}(\tau)$ give

$$\begin{aligned} \mathcal{D} &= 1 + \tau(\tau^{m+1}\mathcal{U}(\tau))^2 + \sum_{j=0}^{m-1} \binom{m+j}{m-j-1} \tau^{j+m+2} (\tau^{m+1}\mathcal{U}(\tau))^{2j+1}(\tau) \\ &= 1 + \tau^{2m+3}\mathcal{U}^2(\tau) + \sum_{j=0}^{m-1} \binom{m+j}{m-j-1} \tau^{(2m+3)(j+1)} \mathcal{U}^{2j+1}(\tau) \\ &= 1 + \tau^{2m+3}\mathcal{U}^2(\tau) + \sum_{j=1}^m \binom{m+j-1}{m-j} \tau^{(2m+3)j} \mathcal{U}^{2j-1}(\tau). \end{aligned}$$

Setting again $t = \tau^{2m+3}$ and denoting the generating functions with the same letters (\mathcal{U} and \mathcal{D} , respectively) but as functions of t , we arrive at

$$(8) \quad \mathcal{D}(t) = 1 + t\mathcal{U}^2(t) + \sum_{j=1}^m \binom{m+j-1}{m-j} t^j \mathcal{U}^{2j-1}(t).$$

As a consequence of (8) and Theorem 5, we obtain:

Corollary 6. *If we write $\mathcal{D}(t) = 1 + \sum_{n \geq 1} \theta_n t^n$, then*

$$\theta_n = \sum_{\ell=0}^{\min(m,n-1)} \binom{m+\ell+1}{m-\ell} \Delta_{n-\ell-1,\ell},$$

where

$$\Delta_{\nu,\ell} = \frac{2\ell+1}{2\nu+2\ell+1} \sum_{k=0}^{\nu} \binom{2\nu+2\ell+1}{k} \frac{k!}{\nu!} B_{\nu,k}(1!\mu_1, 2!\mu_2, \dots).$$

Proof. We use a similar strategy as for the proof of Theorem 5. Here we write the powers of $\mathcal{U}(t)$ in terms of partial Bell polynomials, and then apply [4, Theorem 15] to rewrite all expressions in terms of the sequence μ_1, μ_2, \dots . For convenience, let $!\mu$ denote the sequence $(1!\mu_1, 2!\mu_2, \dots)$.

We will work with $\mathcal{D}(t)$ as given in (8). First, let us consider $t\mathcal{U}^2(t)$. Using (7) together with [8, Sec. 3.5, Theorem B] and [4, Theorem 15] with $\lambda = 1$, we obtain

$$\begin{aligned} [t^n]t\mathcal{U}^2(t) &= \sum_{k=1}^{n-1} 2 \binom{2n-1}{k-1} \frac{(k-1)!}{(n-1)!} B_{n-1,k}(!\mu) \\ &= \sum_{k=1}^{n-1} \left(\binom{2n-2}{k-1} + \frac{2n+k-2}{2n-1} \binom{2n-1}{k-1} \right) \frac{(k-1)!}{(n-1)!} B_{n-1,k}(!\mu) \\ &= \sum_{k=1}^{n-1} \frac{1}{2n-1} \binom{2n-1}{k} \frac{k!}{(n-1)!} B_{n-1,k}(!\mu) \\ &\quad + \sum_{k=0}^{n-2} \frac{1}{2n-1} \binom{2n-1}{k} \frac{k!}{(n-1)!} (2(n-1) + k + 1) B_{n-1,k+1}(!\mu). \end{aligned}$$

Now, using identities [7, Sec. 11.2, Eqns. (11.11) & (11.12)] we get

$$(2(n-1) + k + 1)B_{n-1, k+1}(!\mu) = \sum_{\ell=1}^{n-k-1} (2\ell+1) \binom{n-1}{\ell} \ell! \mu_{\ell} B_{n-1-\ell, k}(!\mu),$$

and thus $[t^n]t\mathcal{U}^2(t)$ can be written as

$$[t^n]t\mathcal{U}^2(t) = \sum_{\ell=0}^{n-1} \frac{2\ell+1}{2n-1} \binom{m+\ell}{m-\ell} \sum_{k=0}^{n-\ell-1} \binom{2n-1}{k} \frac{k!}{(n-1-\ell)!} B_{n-\ell-1, k}(!\mu).$$

Similarly, but now applying [4, Theorem 15] with $\lambda = 2\ell - 2$, we obtain

$$[t^n]t^{\ell}\mathcal{U}^{2\ell-1}(t) = \frac{2\ell-1}{2n-1} \sum_{k=1}^{n-\ell} \binom{2n-1}{k} \frac{k!}{(n-\ell)!} B_{n-\ell, k}(!\mu).$$

This implies

$$[t^n] \sum_{\ell=1}^m \binom{m+\ell-1}{m-\ell} t^{\ell} \mathcal{U}^{2\ell-1} = \sum_{\ell=0}^{m-1} \frac{2\ell+1}{2n-1} \binom{m+\ell}{m-\ell-1} \sum_{k=1}^{n-\ell-1} \binom{2n-1}{k} \frac{k!}{(n-\ell-1)!} B_{n-\ell-1, k}(!\mu).$$

Finally, the corollary follows by combining the above formulas. \square

4. EXAMPLES: FACTOR-FREE WORDS WITH SLOPE $3/2$ AND $5/2$

The purpose of this section is to illustrate Theorem 5 for factor-free words with slope $\frac{3}{2}$ and $\frac{5}{2}$. In the case of slope $\frac{3}{2}$, we recover the results obtained by Duchon [9, Sec. 6.3] who observed that (u_n) is the sequence of Catalan numbers (C_n) . Indeed, by Theorem 5, the number of words of length $5n$ in $U_{\frac{3}{2}}$ is given by

$$u_n = \sum_{k=1}^n \binom{2n}{k-1} \frac{(k-1)!}{n!} B_{n, k}(1, 0, \dots) = \binom{2n}{n-1} \frac{(n-1)!}{n!} = C_n.$$

Moreover, since $\mathcal{U}(t) = 1 + t\mathcal{U}^2(t)$ by (6), identity (8) implies

$$\mathcal{D}(t) = 1 + t\mathcal{U}^2(t) + t\mathcal{U}(t) = \mathcal{U}(t) + t\mathcal{U}(t).$$

Thus the number of factor-free Dyck words with slope $\frac{3}{2}$ and length $5n$ is given by

$$\theta_n = u_n + u_{n-1} = C_n + C_{n-1},$$

as already established in [9, Proposition 5].

Let us now look at the set U for the case $m = 2$ (slope $\frac{5}{2}$).

By Theorem 5, the number of words of length $7n$ in $U_{\frac{5}{2}}$ is given by

$$\begin{aligned} u_n &= \sum_{k=1}^n \binom{2n}{k-1} \frac{(k-1)!}{n!} B_{n,k}(3, 2, 0, \dots) \\ &= \sum_{k=\lceil n/2 \rceil}^n \frac{1}{k} \binom{2n}{k-1} \binom{k}{n-k} 3^{2k-n} \\ &= \frac{1}{2n+1} \sum_{k=\lceil n/2 \rceil}^n \binom{2n+1}{k} \binom{k}{n-k} 3^{2k-n}, \end{aligned}$$

which gives the numbers 3, 19, 153, 1390, 13581, 139315, 1479855, ...

Moreover, by Corollary 6, the number of factor-free Dyck words with slope $\frac{5}{2}$ and length $7n$ is given by

$$\theta_n = \sum_{\ell=0}^{\min(2, n-1)} \binom{\ell+3}{2-\ell} \Delta_{n-\ell-1, \ell},$$

where

$$\Delta_{\nu, \ell} = \frac{2\ell+1}{2\nu+2\ell+1} \sum_{k=0}^{\nu} \binom{2\nu+2\ell+1}{k} \binom{k}{\nu-k} 3^{2k-\nu}.$$

Hence $\theta_1 = 3$, $\theta_2 = 13$, and for $n \geq 3$ we have

$$\theta_n = \sum_{\ell=0}^2 \binom{\ell+3}{2-\ell} \Delta_{n-\ell-1, \ell} = 3\Delta_{n-1, 0} + 4\Delta_{n-2, 1} + \Delta_{n-3, 2}$$

This gives the sequence 3, 13, 94, 810, 7667, 76998, 805560, ..., [11, A274052].

Combinatorial interpretation of $U_{\frac{5}{2}}$. As discussed in Example 3, the language $U = U_{\frac{5}{2}}$ has the grammar

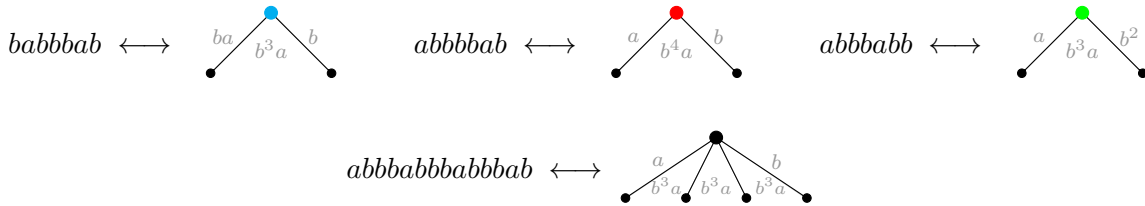
$$U = \varepsilon + aUbbbaUbbbaUbbbaUb + baUbbbaUb + aUbbbbaUb + aUbbbaUbb.$$

This suggests a natural bijection to rooted planar trees. Specifically:

The words of length $7n$ in $U_{\frac{5}{2}}$ are in one-to-one correspondence with rooted trees with $2n$ edges having nonleaf nodes of outdegrees 2 or 4, where nodes of outdegree 2 may be colored in three different ways.

We finish this section with a discussion of this bijection.

Note that $babbbab$, $abbbbab$, $abbabb$, and $abbabbabbab$, are the basic words needed to build all other words in U . We identify these building blocks with colored trees as follows:



The grammar of U implies that any word of length $7n$ can be formed by inserting (right after an a):

- (i) a word of length $7(n - 1)$ into one of the three words of length 7,
- (ii) or a word of length $7(n - 2)$ into the word $abbabbbabbbab$.

Therefore, an ‘ a ’ or a ‘ ba ’ that is not part of a $bbba$ string is always followed by one element of the set of subwords $\{a, ba, bbba, bbbba\}$.

For any word $u \in U$ of length $7n$, we will construct a tree with the properties stated in the bijection. To this end, traverse the word from left to right and do the following (labeling the left edges accordingly):

- o For every a or ba that is not part of a $bbba$ string, draw a left edge and move to the leaf just created. If the edge is labeled with ‘ ba ’, color the parent node blue.
- o For every $bbba$ that is not part of a $bbbba$ string, draw a right edge from the parent node and move to the leaf just created.
- o For every $bbbba$, there are two possible steps:
 - If the edge created last was a left edge, draw a right edge from the parent node, move to the leaf just created, and color the parent node red.
 - Otherwise, draw a right edge from the *grandparent* node and move to the leaf just created.
- o For every bb that is not part of a $bbba$ or a $bbbba$ string, move to the grandparent node unless the current node is a leaf of a binary subtree whose left edge is labeled with an ‘ a ’. In the latter case, just move to the parent node and color it green.
- o For every b that doesn’t fall into any of the previous cases, move to the parent node.

Since every appearance of a is responsible for the creation of an edge, a word in U of length $7n$ corresponds to a tree with $2n$ edges.

The reverse algorithm is clear. Given a colored tree with $2n$ edges, label each of the four building subtrees according to the identification given above. Then traverse the tree counterclockwise, starting at the root, and record the labels writing the letters from left to right. Depending on the color of the nodes, write a or ba when traveling down along a left edge, write $bbba$ or $bbbba$ when traveling up and down between adjacent edges, and write b or bb when traveling up along a right edge. The resulting word has length $7n$ and belongs to $U_{\frac{5}{2}}$.

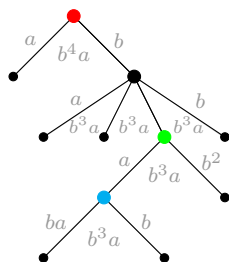


FIGURE 3. Tree representation of $abbbaabbabbbbaababbabbbabbbbabbb$.

5. FURTHER REMARKS AND APPLICATIONS

5.1. **Slope $\frac{7}{2}$.** If $m = 3$, we get $\mu_1 = \binom{4}{2} = 6$, $\mu_2 = \binom{5}{1} = 5$, and $\mu_3 = \binom{6}{0} = 1$. Thus there are 12 basic words that may be used as building blocks to create all words in $U_{\frac{7}{2}}$:

<i>abbbbabbb</i>	<i>abbbbabbbbabbbbabb</i>
<i>abbbbabb</i>	<i>abbbbabbbbabbbbab</i>
<i>abbbbbab</i>	<i>abbbbabbbbabbbbab</i>
<i>babbbbabb</i>	<i>abbbbabbbbabbbbab</i>
<i>babbbbab</i>	<i>babbbbabbbbabbbbab</i>
<i>bbabbbbab</i>	<i>abbbbabbbbabbbbabbbbab</i>

5.2. **Connection to colored Dyck paths.** Combining equation (7) with [3, Theorem 3.5], we conclude that there is a one-to-one correspondence between the words in $U_{\frac{2m+1}{2}}$ of length $(2m+3)n$ and the set of Dyck words of semilength $2n$ created from strings of the form ‘ d ’ and ‘ $u^{2j}d$ ’ for $j = 1, \dots, \min(m, n)$, such that each maximal $2j$ -ascent may be colored in μ_j different ways. For example, for $m = 2$ this means that there is a bijection between the words in $U_{\frac{5}{2}}$ of length $7n$ and the set of Dyck words of semilength $2n$ with ascents of length 2 or 4 and such that each ascent of length 2 may be colored in three different ways.

5.3. **Factor-free words.** The importance of factor-free Dyck words relies on the fact that, as shown in [9, Section 4], words in a generalized Dyck language can be obtained uniquely by inserting words of the language into factor-free words of the same language. In particular, in the case of a two-letter alphabet, for which words correspond to rational Dyck paths, Duchon [9, Theorem 9] established a direct generating function connection between rational Dyck paths and their subset of factor-free elements.

As discussed in this paper, when the slope of the Dyck words is $\frac{2m+1}{2}$, understanding the auxiliary language U introduced in Section 2 suffices to generate and enumerate the set of corresponding factor-free Dyck words with same slope. An explicit enumeration formula is given in (7). The language U turned out to have interesting combinatorial properties, which we have illustrated for the cases of slope $\frac{3}{2}$ and $\frac{5}{2}$.

5.4. **Cross-bifix-free codes.** By definition, if w is a factor-free Dyck word in $\mathcal{D}_{A,h}$, then for any representation $w = w_1w_2$ with nonempty subwords w_1 and w_2 , we must have $h(w_1) > 0$ and $h(w_2) < 0$. Therefore, no prefix of any length of any factor-free word is the suffix of any other factor-free word. This means that any set of factor-free words in $\mathcal{D}_{A,h}$ is a cross-bifix-free set, and in the case of slope $\frac{2m+1}{2}$, we can generate these sets using the auxiliary language U and the derivation rules (2).

For example, the set of factor-free Dyck words over the alphabet $\{0, 1\}$ with $h(0) = 3$ and $h(1) = -2$ (slope $\frac{3}{2}$) gives a cross-bifix-free set (non-overlapping code) of binary words with variable length $\equiv 0$ modulo 5. Here are the words in $D_{\frac{3}{2}}$ of length 5, 10, and 15:

00111, 01011,
 0100110111, 0001101111, 0011011011,
 010011001101111, 010001101110111, 001101100110111, 000110011011111,
 001100110111011, 000011011101111, 000110111011011,

and here are the nineteen words in $D_{\frac{3}{2}}$ of length 20:

01001100110011011111, 01001100011011101111, 01000110111001101111,
 01000110011011110111, 01000011011101110111, 00110110011001101111,
 00110110001101110111, 00110011011100110111, 00011011101100110111,
 00011001100110111111, 00110011001101111011, 00011000110111011111,
 00110001101110111011, 00001101110011011111, 00011011100110111011,
 00001100110111101111, 00011001101111011011, 00000110111011101111,
 00001101110111011011.

Of course, similar sets can be constructed by means of U for binary Dyck words with slope $\frac{2m+1}{2}$. However, if we require words in the set to have a fixed length (as it is customary in the literature, see e.g. [2, 6]), the codes obtained with these slopes may be too restrictive. Nonetheless, rational lattice paths of other slopes may be used to produce larger codes along the lines of those given by Bilotta et al. [2]. For instance, there is a straightforward bijection between the set $CBFS_2(2m+1)$ constructed in op. cit. and the set of rational Dyck paths from $(0, 0)$ to $(m, m+1)$.

REFERENCES

- [1] C. Banderier and M. Wallner, Lattice paths of slope $2/5$, *2015 Proceedings of the Twelfth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, 105–113, SIAM, Philadelphia, PA, 2015.
- [2] S. Bilotta, E. Pergola, and R. Pinzani, A new approach to cross-bifix-free sets, *IEEE Trans. Inform. Theory* **58** (2012), 4058–4063.
- [3] D. Birmajer, J. Gil, P. McNamara, M. Weiner, Enumeration of colored Dyck paths via partial Bell polynomials, preprint arXiv:1602.03550, 2016.
- [4] D. Birmajer, J. Gil, and M. Weiner, Some convolution identities and an inverse relation involving partial Bell polynomials, *Electron. J. Combin.* **19** (2012), no. 4, Paper 34, 14 pp.
- [5] D. Birmajer, J. Gil, and M. Weiner, *Linear recurrence sequences and their convolutions via Bell polynomials*, *J. Integer Seq.* **18** (2015), no. 1, Article 15.1.2, 14 pp.
- [6] S. Blackburn, Non-overlapping codes, *IEEE Trans. Inform. Theory*, **61** (2015), no. 9, 4890–4894.
- [7] C. A. Charalambides, *Enumerative Combinatorics*, Chapman and Hall/CRC, Boca Raton, 2002.
- [8] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions*, D. Reidel Publishing Co., Dordrecht, 1974.
- [9] P. Duchon, On the enumeration and generation of generalized Dyck words, *Discrete Math.* **225** (2000), no. 1-3, 121–135.
- [10] J. Labelle and Y.-N. Yeh, Generalized Dyck paths, *Discrete Math.* **82** (1990), 1–6.
- [11] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, <http://oeis.org>.

NAZARETH COLLEGE, 4245 EAST AVE., ROCHESTER, NY 14618

PENN STATE ALTOONA, 3000 IVYSIDE PARK, ALTOONA, PA 16601