#A69 INTEGERS 21 (2021)

ENUMERATING RESTRICTED DYCK PATHS WITH CONTEXT FREE GRAMMARS

AJ Bu
Department of Mathematics, Rutgers University, New Brunswick, New Jersey
ab1854@rutgers.edu

Robert Dougherty-Bliss
Department of Mathematics, Rutgers University, New Brunswick, New Jersey
robert.w.bliss@gmail.com

Received: 11/30/20, Accepted: 6/16/21, Published: 7/9/21

Abstract

The number of Dyck paths of semilength n is famously C_n, the nth Catalan number. This fact follows after noticing that every Dyck path can be uniquely *parsed* according to a context-free grammar. In a recent paper, Zeilberger showed that many *restricted* sets of Dyck paths satisfy different, more complicated grammars, and from this derived various generating function identities. We take this further, highlighting some combinatorial results about Dyck paths obtained via grammatical proof and generalizing some of Zeilberger's grammars to infinite families.

1. Introduction

As Flajolet and Sedgewick masterfully demonstrate in their seminal text, *Analytic Combinatorics* [4], mathematicians have occasionally borrowed the study of formal languages from computer science and linguistics for combinatorial reasons. Many combinatorial classes can be reinterpreted as languages generated by certain grammars, and these grammars often make writing down generating functions, another favorite combinatorial tool, routine. Such grammars are sometimes called "combinatorial specifications."

For example, consider the well-known *Dyck paths*. A Dyck path is a finite list of $+1$'s and -1's whose partial sums are nonnegative, and whose sum is 0. We will write U (up) for $+1$ and D (down) for -1. Thus, the following are all Dyck paths:

$$UUDD$$
$$UDUD$$
$$UUUDUDDD$$

A Dyck path must have even length, and for this reason we often refer to Dyck paths of *semilength* n (length $2n$).

The number of Dyck paths of semilength n equals the nth *Catalan number*,

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

There are many proofs of this fact, but here is a *grammatical* proof.

Let $\mathcal{P}$ denote the set of all Dyck paths. Then, $\mathcal{P}$ is generated by the unambiguous, context-free grammar

$$\mathcal{P} = \epsilon \ \cup \ U\mathcal{P}D\mathcal{P}, \tag{1}$$

where ϵ denotes the empty string. In words, a path is either empty or begins with a U, is followed by a Dyck path (shifted to height 1), a D, then another Dyck path.[1] This is a unique parsing of all Dyck paths.

Given a set of objects E each with a nonnegative integer size, let $GF(E) = \sum_{k\geq 0} |E(k)|z^k$ be a formal generating function, where $|E(k)|$ is the number of objects of size k in E. The main result about formal grammars is that, in an unambiguous context free grammar,

$$GF(A \cup B) = GF(A) + GF(B),$$

for disjoint clauses A and B, and

$$GF(AB) = GF(A)GF(B),$$

where $A \cup B$ is the union of the words of A and the words of B, and AB stands for "concatenation of words of A with words in B." The "sizes" of a grammar are the lengths of the words it generates.

In our case, if $P(z)$ is the generating function for the number of Dyck paths of semilength n, then the grammar (1) implies

$$\begin{aligned} P(z) &= GF(\epsilon) + GF(U\mathcal{P}D\mathcal{P}) \\ &= 1 + zP(z)^2. \end{aligned}$$

The generating function $C(z)$ for the Catalan numbers *also* satisfies

$$C(z) = 1 + zC^2(z),$$

and since there are only two possible solutions, it is not hard to see that $P(z) = C(z)$.

The grammatical technique offers a unifying framework: Devise a grammar and you get an equation. Sometimes the equations turn out to be well-known or simple.

[1]Note that D denotes the first time the path returns to height 0.

Other times they are new and messy. The enumeration of all Dyck paths is one application of this framework, and here we want to demonstrate others. In particular, we will give grammatical proofs of several combinatorial facts about *restricted* Dyck paths, and also establish several infinite families of grammars in closed form.

First, let us define the restrictions we shall consider.

Definition 1. Given a Dyck path, the *height* of the path at position k is the partial sum of its first k terms. A *peak* of a Dyck path at height h (or simply "at h") is the bigram UD where the height of the path after the U is h. Similarly, a *valley* occurs at the bigram DU, and its height is analogously defined. The empty path has, by convention, a peak at 0 but no valley.

Definition 2. Given a sequence of steps L, define L^n to be the repetition of L n times. (For example, $U^2 = UU$ and $(UD)^3 = UDUDUD$.) A Dyck path has an *up-run of length* n provided that it contains at least one U^n that is not preceded nor followed by U. Similarly, it contains a *down-run of length* n provided that it contains at least one D^n that is neither preceded nor followed by D.

We will study Dyck paths whose peak heights, valley heights, up-run lengths, and down-run lengths avoid certain sets. We will, for example, discuss the set of all Dyck paths whose peak heights avoid $\{2, 4, 6, \dots\}$ and have no up-run of length greater than 2.

Definition 3. For arbitrary sets of positive integers A, B, C, and D, let $\mathcal{P}(A, B, C, D)$ be the set of Dyck paths whose peak heights avoid A, whose valley heights avoid B, whose up-run lengths avoid C, and whose down-run lengths avoid D. Let $P_{A,B,C,D}(z)$ be be the generating function for the number of Dyck paths of semilength n in $\mathcal{P}(A, B, C, D)$.

Some of these sets have been studied. In [8], Peart and Woan provide a continued-fraction recurrence for the generating functions $P_{\{k\},\emptyset,\emptyset,\emptyset}(z)$. In [3], Eu, Liu, and Yeh take this idea further and express $P_{A,\emptyset,\emptyset,\emptyset}(z)$ as a finite continued fraction whenever A is finite or an arithmetic progression. In [5], Hein and Huang enumerate the number of Dyck paths which avoid up-runs of length k after a down step. In [2], Zeilberger presents a rigorous experimental method to derive equations for $P_{A,B,C,D}(z)$ when the sets involved are finite or arithmetic progressions. Proving "by hand" some of Zeilberger's interesting discoveries *ex post facto* was a motivation for the present work. We generalize some of Zeilberger's results to infinite families which are likely out of reach for symbolic methods.

Our results include several explicit grammars (and therefore generating function equations) for infinite families of the sets A and B, and also grammatical proofs of several interesting special cases suggested in [2]. Many of these—any grammars referencing restrictions on up- or down-runs—are not in [3]. Some of our results are suggested in the OEIS [6]; see, for example, A1006 (Motzkin numbers) and A004148 (generalized Catalan numbers).

The remainder of the paper is organized as follows. Section 2 presents some results discovered by experimentation with software from [2] and proven with grammatical methods. Section 3 presents some infinite families of explicit grammars. Section 4 offers some concluding remarks about the limitations of grammars.

2. Combinatorial Results

In this section we will present a number of results with grammatical proofs.

Proposition 1. *The number of Dyck paths of semilength n whose peak heights avoid $\{2r+3 \mid r \geq 0\}$ and whose up-runs are no longer than 2 is 1 when $n = 0$, and 2^{n-1} when $n \geq 1$.*

Proof. Let $\mathcal{P}$ be the set of all such Dyck paths, and $\mathcal{Q}$ the set of all Dyck paths which avoid peaks in $\{2r+2\}$ and up-runs longer than 2. Note that $\mathcal{P}$ and $\mathcal{Q}$ satisfy the following grammar:

$$\begin{aligned}\mathcal{P} &= \epsilon \ \cup \ UD\mathcal{P} \ \cup \ UUD\mathcal{Q}D\mathcal{P} \\ Q &= \epsilon \ \cup \ UDQ.\end{aligned}$$

This implies the following system of equations:

$$\begin{aligned}P &= 1 + zP + z^2QP \\ Q &= 1 + zQ.\end{aligned}$$

Thus, $Q(z) = (1-z)^{-1}$ (the only path in Q of semilength n is $(UD)^n$), and

$$P(z) = \frac{1-z}{1-2z}.$$

Therefore, $[z^0]P(z) = 1$, and $[z^n]P(z) = 2^{n-1}$. $\square$

The following proposition concerns *generalized Catalan numbers* (see A4148 in the OEIS and [9]). These numbers are defined by the recurrence

$$\begin{aligned}G_0 &= 1 \\ G_1 &= 1 \\ G_{n+2} &= G_{n+1} + \sum_{1 \leq k < n+1} G_k G_{n-k}.\end{aligned}$$

Proposition 2. *The number of Dyck paths of semilength n whose peak heights avoid $\{2r+3 \mid r \geq 0\}$ and whose up-runs are no longer than 3 equals the $(n+1)$th generalized Catalan number.*

Proof. Let $\mathcal{P}$, $\mathcal{O}$, and $\mathcal{E}$ be the set of all Dyck paths with up-runs no longer than 3, and whose peak heights avoid $\{2r+3 \mid r \geq 0\}$, $\{2r+2 \mid r \geq 0\}$, and $\{2r+1 \mid r \geq 0\}$, respectively. Observe that $\mathcal{P}$, $\mathcal{O}$, and $\mathcal{E}$ satisfy the following grammar:

$$\begin{aligned}
\mathcal{P} &= \epsilon \ \cup \ UD\mathcal{P} \ \cup \ UUD\mathcal{O}D\mathcal{P} \\
\mathcal{O} &= \epsilon \ \cup \ UD\mathcal{O} \ \cup \ UUUD\mathcal{O}D\mathcal{E}D\mathcal{O} \\
\mathcal{E} &= \epsilon \ \cup \ UUD\mathcal{O}D\mathcal{E}.
\end{aligned}$$

This grammar implies the following equations:

$$\begin{aligned}
P &= 1 + zP + z^2OP \\
O &= 1 + zO + z^3EO^2 \\
E &= 1 + z^2OE.
\end{aligned}$$

This system has two possible solutions for P, but only one is holomorphic near the origin, namely

$$P(z) = \frac{2}{1 - z - z^2 + (z^4 - 2z^3 - z^2 - 2z + 1)^{1/2}}.$$

The generating function $G(z)$ for the generalized Catalan numbers is (see A4148 in the OEIS)

$$G(z) = \frac{1 - z + z^2 - (1 - 2z - z^2 - 2z^3 + z^4)^{1/2}}{2z^2},$$

and it is routine to verify that $G(z) = zP(z) + 1$. Therefore $G_{n+1} = [z^n]P(z)$ for $n \geq 0$. □

The following proposition is concerned with *Motzkin numbers* (see A1006 in the OEIS and [1]). A *Motzkin path* is like a Dyck path, but includes a “sideways” step S which does not change the height. The nth Motzkin number M_n is the number of Motzkin paths of length n. The generating function $M = M(z)$ for M_n satisfies the quadratic equation

$$M = 1 + zM + z^2M^2.$$

There are numerous bijections between Motzkin paths and various restricted classes of Dyck paths. Such bijections are often variations of the “folding” map

$$\begin{aligned}
UD &\mapsto S \\
DU &\mapsto S \\
UU &\mapsto U \\
DD &\mapsto D,
\end{aligned}$$

which in general is not injective, but many restrictions on Dyck paths *make* it injective. For example, this idea shows that the Dyck paths of semilength n with no up-runs longer than 2 are in bijection with the Motzkin paths of length n. We offer a grammatical proof of this fact.

Proposition 3. *The number of Dyck paths of semilength n which avoid up-runs of length 3 or more equals the nth Motzkin number M_n.*

Proof. Let $\mathcal{P}$ be the set of such paths. A grammar for $\mathcal{P}$ is

$$\mathcal{P} = \epsilon \ \cup \ UUD\mathcal{P}D\mathcal{P} \ \cup \ UD\mathcal{P}.$$

Our grammar implies that

$$P = 1 + zP + z^2P^2.$$

This is the same equation satisfied by the Motzkin generating function, and it is easy to check that $P(z) = M(z)$. □

Proposition 4. *Consider the set of Dyck paths such that no peak or valley has positive, even height. The numbers of such paths of semilength $2n$ and $2n+1$ are $\binom{2n-1}{n}$ and $\binom{2n}{n}$, respectively.*

Proof. Let $\mathcal{P}$ denote the set of such paths, and let $\mathcal{O}$ denote the set of all Dyck paths whose peaks and valleys avoid odd heights. These sets satisfy the following grammars

$$\begin{aligned}\mathcal{P} &= \epsilon \cup U\mathcal{O}D\mathcal{P},\\ \mathcal{O} &= \epsilon \cup UU\mathcal{O}DD\mathcal{O}.\end{aligned}$$

This grammar can be translated into the following equations:

$$\begin{aligned}P &= 1 + zOP, \text{ and}\\ O &= 1 + z^2O^2\end{aligned}$$

Solving this system for O, we get two solutions for O, but only the following is holomorphic near the origin

$$O = \frac{1-\sqrt{1-4z^2}}{2z^2}.$$

Thus,

$$P = \frac{2z-1-\sqrt{1-4z^2}}{2(2z-1)},$$

and it is easy to check that

$$\begin{aligned}[z^{2n}]P(z) &= \binom{2n-1}{n}, \text{ and}\\ [z^{2n+1}]P(z) &= \binom{2n}{n}.\end{aligned}$$

□

Now, let us define a mapping which allows us to translate restrictions on up-run (respectively, down-run) lengths into restrictions on down-run (respectively, up-run) lengths. Let $\mathcal{P}$ denote the set of all Dyck paths. Define the mapping

$$\phi : \mathcal{P} \to \mathcal{P}, \qquad P \mapsto Q, \tag{2}$$

where applying ϕ reverses the order and direction of the steps in P. For example,

$$\phi(UUUDUUDDDDUD) = UDUUUUDDUDDD.$$

It is obvious that $\phi(P)$ must be a Dyck path. Moreover, it is easy to check that ϕ is an involution. Note that the up-runs (respectively, down-runs) in P become down-runs (respectively, up-runs) in $\phi(P)$ of the same length.

Proposition 5. *Let A and B be arbitrary sets of positive integers. The number of Dyck paths of semi-length n which avoid up-runs and down-runs with lengths in A and B, respectively, equals the number of Dyck paths of semi-length n which avoid down-runs with lengths in A and up-runs with lengths in B.*

Proof. Let $\mathcal{P}(A, B)$ be the set of Dyck paths such that no up-run has length in A and no down-run has length in B, and $\mathcal{P}(B, A)$ be the set of Dyck paths such that no up-run has length in B and no down-run has length in A. Then ϕ – defined in equation 2 – gives a one-to-one correspondence between the Dyck paths of semi-length n in $\mathcal{P}(A, B)$ and the Dyck paths of semi-length n in $\mathcal{P}(B, A)$. □

Note that ϕ also allows us to translate the grammar of $\mathcal{P}(A, B)$ into the grammar of $\mathcal{P}(B, A)$, as seen in the following section.

3. Grammatical Families

In this section we provide some explicit grammars for infinite families of restricted Dyck paths. In many cases, such grammars are guaranteed to exist. The reasoning in [2] shows that, for every set of Dyck paths whose peaks, valleys, and up- and down-runs avoid specific arithmetic progressions, we may construct a finite, context-free grammar which generates them. The method implied in [2] to compute these grammars gives no hint as to their *form*, and this is what we try to provide here.

Our first two results are about Dyck paths whose up-run lengths avoid a fixed arithmetic progression $\{Ar+B \mid r \geq 0\}$; each of these is accompanied by a corollary on Dyck paths that avoid down-run lengths in $\{Ar + B \mid r \geq 0\}$. It turns out that when $B < A$, there is a simple context-free grammar for such paths. When $B \geq A$ the situation is more complicated, but we can derive a “grammatical equation” which again leads to a generating function.

Proposition 6. *Let $B < A$ be non-negative integers. The set $\mathcal{P}$ of Dyck paths whose up-run lengths avoid $\{Ar + B \mid r \geq 0\}$ has the unambiguous grammar*

$$\mathcal{P} = \left(\bigcup_{\substack{0 \leq k < A \\ k \neq B}} U^k (D\mathcal{P})^k \right) \cup U^A (\mathcal{P}D)^A \mathcal{P},$$

and therefore

$$P(z) = \left(\sum_{\substack{0 \leq k < A \\ k \neq B}} z^k P^k(z) \right) + z^A P^{A+1}(z),$$

where $P(z)$ is the weight-enumerator of $\mathcal{P}$.

Proof. The grammar clearly uniquely parses the empty path, so suppose that $P \in \mathcal{P}$ has length $n > 0$. Then P starts with a up-run of length $k > 0$ for some $k \not\equiv B \mod A$. If $k < A$, then write $P = U^k D W$, where W is a walk from height $k-1$ to height 0 with the same restrictions on up-runs as P. For $0 \leq i < k-1$, let D_i indicate the down-step in W which hits the height i for the first time. Then

$$W = P_{k-1} D_{k-2} P_{k-2} D_{k-3} \ldots P_1 D_0 P_0,$$

where P_i is a Dyck path shifted to height i with the same restrictions on up-runs as P. This uniquely parses P into the case $U^k(D\mathcal{P})^k$ in the grammar.

If the initial up-run has length $k \geq A$, then write $P = U^A W$, where W is a walk from height A to height 0 whose up-run lengths avoid $\{Ar + B \mid r \geq 0\}$. By argument analogous to the previous paragraph, we can decompose W as

$$W = P_A D_{A-1} P_{A-1} D_{A-2} \ldots P_1 D_0 P_0,$$

where $P_i \in \mathcal{P}$. Thus W is of the form $(\mathcal{P}D)^A \mathcal{P}$, and this uniquely parses P into the final case of the grammar.

We have shown that $\mathcal{P}$ is contained in the language generated by this grammar, and it is easy to see that the first k cases of the grammar are contained in $\mathcal{P}$. The final case, $U^A(\mathcal{P}D)^A\mathcal{P}$, is also contained in the grammar, because concatenating U^A to the beginning of a path does not change the length any of the up-runs modulo A. The different cases are clearly disjoint, so the grammar is also unambiguous. $\square$

Corollary 1. *Let $A, B \in \mathbb{Z}_{\geq 0}$ such that $B < A$. The set $\mathcal{P}$ of Dyck paths avoiding down-run lengths in $\{Ar + B | r \in \mathbb{Z}_{\geq} 0\}$ has the unambiguous grammar*

$$\mathcal{P} = \left(\bigcup_{\substack{0 \leq k < A \\ k \neq B}} (\mathcal{P}U)^k D^k \right) \cup \mathcal{P}(U\mathcal{P})^A D^A,$$

and therefore

$$P(z) = \left(\sum_{\substack{0 \le k < A \\ k \ne B}} z^k P^k(z) \right) + z^A P^{A+1}(z),$$

where $P(z)$ is the weight-enumerator of $\mathcal{P}$.

Proof. Let ϕ be the involution defined in equation 2, and let $\mathcal{Q}$ be the set of Dyck paths avoiding up-run lengths in $\{Ar + B | r \in \mathbb{Z}_{\ge} 0\}$. By proposition 6,

$$\mathcal{Q} = \bigcup_{\substack{0 \le k < A \\ k \ne B}} U^k(D\mathcal{Q})^k \cup U^A(\mathcal{Q}D)^A \mathcal{Q}.$$

Since

$$\begin{aligned} \phi(\mathcal{Q}) &= \mathcal{P}, \\ \phi(U^k(D\mathcal{Q})^k) &= (\mathcal{P}U)^k D^k, \text{ for all } 0 \le k < A, \text{ and} \\ \phi(U^A(\mathcal{Q}D)^A \mathcal{Q}) &= \mathcal{P}(U\mathcal{P})^A U^A, \end{aligned}$$

ϕ translates the grammar of $\mathcal{Q}$ into the desired grammar for $\mathcal{P}$. $\square$

Proposition 7. *Let $A \le B$ be nonnegative integers. The set $\mathcal{P}$ of Dyck paths avoiding up-run lengths in $\{Ar + B \mid r \ge 0\}$ satisfies the "grammatical equation"*

$$\mathcal{P} \cup U^B(D\mathcal{P})^B = \left(\bigcup_{0 \le k < A} U^k(D\mathcal{P})^k \right) \cup \; U^A(\mathcal{P}D)^A \mathcal{P},$$

and therefore

$$P(z) + z^B P(z)^B = \left(\sum_{0 \le k < A} z^k P^k(z) \right) + z^A P^{A+1}(z),$$

where $P(z)$ is the weight-enumerator of $\mathcal{P}$.

Note that the right-hand side is nearly identical to proposition 6; the difference being that we can get paths in $U^B(D\mathcal{P})^B$, which we will show below.

Proof. If P is a path in $\mathcal{P}$, then we can uniquely parse P into a case of the right-hand side by the same argument given in the previous proposition. Note that

$$\begin{aligned} U^B(D\mathcal{P})^B &= U^A U^{B-A}(D\mathcal{P})^B \\ &= U^A\{U^{B-A}(D\mathcal{P})^{B-A}\}(D\mathcal{P})^A \\ &= U^A[\{U^{B-A}(D\mathcal{P})^{B-A}\}D(\mathcal{P}D)^{A-1}]\mathcal{P}. \end{aligned}$$

The expression in brackets, $U^{B-A}(D\mathcal{P})^{B-A}$, is in $\mathcal{P}$, which shows that $U^B(D\mathcal{P})^B$ is contained in $U^A(\mathcal{P}D)^A\mathcal{P}$.

Conversely, it remains to show that the left-hand side is *all* that the right-hand side can generate. $\bigcup_{0\le k<A} U^k(D\mathcal{P})^k$ is contained in $\mathcal{P}$ as in the previous proposition. For $W \in U^A(\mathcal{P}D)^A\mathcal{P}$, write

$$W = U^A P_1 D \dots P_A D P_{A+1}.$$

Let ℓ be the length of the initial up-run in P_1. If $\ell \not\equiv B \pmod{A}$, then W contains no up-runs of lengths in $\{Ar + B \mid r \ge 0\}$ and is a path in $\mathcal{P}$. If $\ell \equiv B \pmod{A}$, then $\ell \le B - A$. If $\ell < B - A$ then the initial run of W has length less than B. Thus, W contains no up-runs of lengths in $\{Ar + B \mid r \ge 0\}$. For $\ell = B - A$, let D_i denote the first time W steps down to height i for $A < i < B$ and write

$$\begin{aligned} W &= U^A P_1 D \dots P_A D P_{A+1} \\ &= U^A(U^{B-A} D_{B-1} W_{B-1} \dots D_A W_A) D P_2 D \dots P_A D P_{A+1} \\ &= U^B D_{B-1} W_{B-1} \dots D_A W_A D P_2 D \dots P_A D P_{A+1}. \end{aligned}$$

W_i is Dyck path shifted to height i by the definition of D_i. Hence, $W \in U^B(D\mathcal{P})^B$. □

Corollary 2. *Let $A, B \in \mathbb{Z}_{\ge 0}$ such that $B \ge A$. The set $\mathcal{P}$ of Dyck paths avoiding down-run lengths in $\{Ar + B | r \in \mathbb{Z}_{\ge} 0\}$ satisfies the grammatical equation*

$$\mathcal{P} \cup (\mathcal{P}U)^B D^B = \left(\bigcup_{0\le k<A} (\mathcal{P}U)^k D^k\right) \cup \mathcal{P}(U\mathcal{P})^A D^A.$$

and therefore

$$P(z) + z^B P^B(z) = \left(\sum_{0\le k<A} z^k P^k(z)\right) + z^A P^{A+1}(z).$$

where $P(z)$ is the weight-enumerator of $\mathcal{P}$.

Proof. Let ϕ be the involution defined in equation 2, and let $\mathcal{Q}$ be the set of Dyck paths avoiding up-run lengths in $\{Ar + B | r \in \mathbb{Z}_{\ge} 0\}$. Applying ϕ to each clause of the grammar of $\mathcal{Q}$ given in proposition 7, we get

$$\mathcal{P} \cup (\mathcal{P}U)^B D^B = \left(\bigcup_{0\le k<A} (\mathcal{P}U)^k D^k\right) \cup \mathcal{P}(U\mathcal{P})^A D^A,$$

as desired. □

Proposition 8. *Let $r \in \mathbb{Z}^+$. The set $\mathcal{P}$ of Dyck paths avoiding ascending and descending runs of lengths in $\{1, ..., r\}$ satisfies the grammatical equation*

$$\mathcal{P} \cup UD\mathcal{P} = \epsilon \cup U^{r+1}D^{r+1}\mathcal{P} \cup U\mathcal{P}D\mathcal{P}.$$

and therefore

$$P(z) + zP(z) = 1 + z^{r+1}P(z) + zP^2(z),$$

where $P(z)$ is the weight-enumerator of $\mathcal{P}$.

Proof. If $P \in \mathcal{P}$ is the empty path, then the grammar uniquely parses P. Otherwise, $P \in \mathcal{P}$ must begin with an ascending run of length $\ell > r$. If $\ell = r+1$, then clearly U^{r+1} must be immediately followed by the descending run D^{r+1}, and P is uniquely parsed into the case $U^{r+1}D^{r+1}\mathcal{P}$.

If $\ell > r+1$, then let D_0 denote the step where P returns to height 0 for the first time and write

$$P = UP_1D_0P_2.$$

It is obvious that $P_2 \in \mathcal{P}$ and P_1 is a Dyck path shifted to height 1. By restrictions on P, the final descending run in P_1 must have length $L \geq r$. If $L = r$ then the preceding ascending run ends at height $r+1$. But the ascending runs in P must have length of at least $r+1$, and hence P_1 hits height 0, contradicting the definition of D_0. From here, it is clear that P_1 has the same restrictions on ascending and descending runs as P. Thus, P is uniquely parsed into the case $U\mathcal{P}D\mathcal{P}$.

Since it is trivial that $UD\mathcal{P}$ is contained in $U\mathcal{P}D\mathcal{P}$, we have shown that the left-hand side of the given equation is generated by the right-hand side. It is also obvious that the cases defined on the right-hand side are disjoint and that $\epsilon \cup U^{r+1}D^{r+1}\mathcal{P}$ is contained in $\mathcal{P}$. A path $UP_1DP_2 \in U\mathcal{P}D\mathcal{P}$ is contained in $UD\mathcal{P}$ if P_1 is the empty path and $\mathcal{P}$ otherwise. Thus, $\mathcal{P}$ satisfies the given grammatical equation.

□

Proposition 9. *Let $m, n \in \mathbb{Z}^+$. The set $\mathcal{P}$ of Dyck paths avoiding ascending runs of lengths in $\{1, ..., m\}$ and descending runs of lengths in $\{1, ..., n\}$ satisfies the grammatical equation*

$$\mathcal{P} \cup UD\mathcal{P} = \epsilon \cup U\mathcal{P}D\mathcal{P} \cup U^{m+1}D^{n+1}(\mathcal{P}D)^{m-n}\mathcal{P}, \ \textit{if } m \geq n \tag{3}$$

$$\mathcal{P} \cup \mathcal{P}UD = \epsilon \cup \mathcal{P}U\mathcal{P}D \cup \mathcal{P}(U\mathcal{P})^{n-m}U^{m+1}D^{n+1}, \ \textit{if } m \leq n. \tag{4}$$

Proof. We have already shown that this statement is true for $m = n$. Suppose $m > n$. If $P \in \mathcal{P}$ is the empty path, then the grammar uniquely parses P. Otherwise, P must begin with an ascending run of length $\ell > m$. If $\ell = m+1$ then U^{m+1} is followed by a descending chain of length of at least $n+1$. Let D_i denote the first time P returns to height i for $0 \leq i \leq m-n-1$, and write

$$P = U^{m+1}D^{n+1}P_{m-n}D_{m-n-1}...P_1D_0P_0.$$

It is obvious that P_i is a Dyck path, shifted to height i, that has the same restrictions on ascending runs and descending runs (with the exception of the final descending run) as P. Since P_i is a Dyck path, its final descending run must be at least as long as the ascending run preceding it. Thus, P_i is either the empty path or ends with a descending run of length $L > m > n$. Thus, P is uniquely parsed into the case $U^{m+1}D^{n+1}(\mathcal{P}D)^{m-n}\mathcal{P}$.

If $\ell > m+1$ then, letting D_0 denote the first time P returns to height 0, write

$$P = UP_1D_0P_0.$$

Clearly, $P_0 \in \mathcal{P}$, and P_1 is a Dyck path shifted to height 1 and has the same restrictions on ascending runs as P. Using the same argument as for P_i in the previous case, the descending runs in P_1 also have the same restrictions as P. This uniquely parses P into the case $U\mathcal{P}D\mathcal{P}$. Finally, it is obvious that $UD\mathcal{P}$ is contained in $U\mathcal{P}D\mathcal{P}$, so the left-hand side of (1) is generated by the right-hand side.

It is clear that the cases on the right-hand side are disjoint, and the empty path is an element of $\mathcal{P}$. Also, $UP_1DP_2 \in U\mathcal{P}D\mathcal{P}$ is contained in $\mathcal{P}$ if P_1 is not the empty path, and is contained in $UD\mathcal{P}$ otherwise. $U^{m+1}D^{n+1}(\mathcal{P}D)^{m-n}\mathcal{P}$ is contained in $\mathcal{P}$, since all ascending runs clearly avoid restrictions on $\mathcal{P}$ and the descending runs are formed by concatenating down-steps to descending runs of length of at least $n-1$. Thus, we have proved the grammar for the case $m \geq n$.

Now assume that $n \geq m$. Applying the involution ϕ from equation 2, we can directly translate the grammar 3 into the desired grammar 4. □

Proposition 10. *Let $r, k \in \mathbb{Z}^+$ and let $\mathcal{P}$ be the set of Dyck paths avoiding ascending runs of length $\{1, ..., r\}$ and descending runs of length $\{k+1, ..., r\}$. Then the 'grammar' of $\mathcal{P}$ is*

$$\mathcal{P} \cup UD\mathcal{P} \cup U^{r+1}D^k(D\mathcal{P})^{r+1-k} = \epsilon \cup U\mathcal{P}D\mathcal{P} \cup U^{r+1}D^{r+1}\mathcal{P} \cup U^{r+1}(D\mathcal{P})^{r+1}$$

Proof. If $P \in \mathcal{P}$ is the empty path, then the grammar uniquely parses P. Otherwise, P begins an ascending run of length $\ell > r$, and we can deduce that it also ends with a descending run of length $L > r$. If $\ell > r+1$, then let D_0 denote the first time that P returns to the $x-$axis and write

$$P = UP_1D_0P_0.$$

It is easy to see that P_0 is a path in $\mathcal{P}$ and P_1 is a Dyck path shifted to height 1. The initial ascending run in P_1 has length $\ell - 1 > r$. Thus, all ascending runs in P_1 have length of at least $r+1$ and, since P_1 is a shifted Dyck path, the final descending run in P_1 must also have length of at least $r+1$. From here, it is easy to see that P_1 has the same restrictions on ascending and descending runs as P. P is therefore uniquely parsed into the case $U\mathcal{P}D\mathcal{P}$.

Suppose $\ell = r+1$. Let D_i be the step where P returns to height i for the first time and write

$$P = U^{r+1}D_rP_r...D_0P_0.$$

P_i is a Dyck path for all i and, if P_i is not the empty path, it must end with a descending run of length $r+1$ by restrictions on ascending runs. Thus P_i is a path in $\mathcal{P}$, and P is parsed into the case $U^{r+1}(D\mathcal{P})^{r+1}$.

It is trivial that $UD\mathcal{P}$ is contained in $U\mathcal{P}D\mathcal{P}$ and $U^{r+1}D^k(D\mathcal{P})^{r+1-k}$ is contained in $U^{r+1}(D\mathcal{P})^{r+1}$. Thus, the left-hand side is generated by the right-hand side. Note that, on the left-hand side,

$$UD\mathcal{P} \cap \mathcal{P} = UD\mathcal{P} \cap U^{r+1}D^k(D\mathcal{P})^{r+1-k} = \emptyset,$$

however

$$\mathcal{P} \cap U^{r+1}D^k(D\mathcal{P})^{r+1-k} = U^{r+1}D^{r+1}\mathcal{P}.$$

Looking at the right-hand side, it is clear that $\epsilon, U\mathcal{P}D\mathcal{P}$, and $U^{r+1}(D\mathcal{P})^{r+1}$ are disjoint, and $U^{r+1}D^{r+1}\mathcal{P}$ is contained in $U^{r+1}(D\mathcal{P})^{r+1}$. Note that this resolves the issue of double counting paths in $U^{r+1}D^{r+1}\mathcal{P}$ on the left-hand side. Thus, all that remains to show is that all the paths generated by the right-hand side are contained in the left-hand side.

The path UP_1DP_0 in $U\mathcal{P}D\mathcal{P}$ is clearly in $\mathcal{P}$ if P_1 is not the empty path and in $UD\mathcal{P}$ otherwise. For W in $U^{r+1}(D\mathcal{P})^{r+1}$, write

$$W = U^{r+1}D_rP_r...D_1P_1D_0P_0.$$

Choose the smallest i such that P_{r-i} is not the empty path or, if no such i exists, set $i=r$. Then the first descending run in W has length $i+1$. If $i \geq k$ then W is an element of $U^{r+1}D^k(D\mathcal{P})^{r+1-k}$. Otherwise, we claim that W is a path in $\mathcal{P}$. It is clear that W is a Dyck path and we have seen that nonempty $P_j \in \mathcal{P}$ must end in a descending run of length of at least $r+1$. Thus, we only need to show that the first descending run in W follows the restrictions in $\mathcal{P}$. This is clearly true since $i < k$. Hence $W \in \mathcal{P}$, and $\mathcal{P}$ satisfies the grammatical equation as desired. □

4. Conclusion

We have given several grammatical proofs of various combinatorial results about restricted Dyck paths and established some infinite families of grammars. Our methods work because we are able to derive *context-free grammars* describing certain restricted classes Dyck paths, namely when our restrictions involved sets of arithmetic progressions.

It is natural to ask if context-free grammars exist for other types of restrictions. Parikh's theorem [7] states that the set of lengths of any context-free language is the union of finitely-many arithmetic progressions, so it seems likely that restrictions involving arithmetic progressions are *essentially* all that can be done. However, addressing this question in full is beyond our current scope.

References

[1] R. Donaghey and L.W. Shapiro, Motzkin numbers, *J. Combin. Theory Ser. A* **23(3)** (1977), 291-301.

[2] S.B. Ekhad and D. Zeilberger, Automatic counting of restricted Dyck paths via (numeric and symbolic) dynamic programming, *arXiv* (2020), 2006.01961.

[3] S. Eu, S. Liu, Y. Yeh, Dyck paths with peaks avoiding or restricted to a given set, *Stud. Appl. Math.* **111(4)** (2003), 453-465.

[4] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*, Cambridge University Press, 2009.

[5] N. Hein and J. Huang, Modular Catalan numbers, *European J. Combin.* **61** (2017) 197-218.

[6] OEIS Foundation Inc. (2020), The On-Line Encyclopedia of Integer Sequences, http://oeis.org.

[7] R.J. Parikh, On context-free languages, *J. ACM* **13(4)** (1966), 570-581.

[8] P. Peart and W.J. Woan, Dyck paths with no peaks at height k, *J. Integer Seq.* **4(1)** (2001).

[9] P.R. Stein and M.S. Waterman, On some new sequences generalizing the Catalan and Motzkin numbers, *Discrete Math.* **26(3)** (1979), 261-272.