A Bijection on Dyck Paths and Its Cycle Structure

DAVID CALLAN
Department of Statistics
University of Wisconsin-Madison
1300 University Ave
Madison, WI 53706-1532
callan@stat.wisc.edu

November 21, 2006

Abstract

The known bijections on Dyck paths are either involutions or have notoriously intractable cycle structure. Here we present a size-preserving bijection on Dyck paths whose cycle structure is amenable to complete analysis. In particular, each cycle has length a power of 2. A new manifestation of the Catalan numbers as labeled forests crops up enroute as does the Pascal matrix mod 2. We use the bijection to show the equivalence of two known manifestations of the Motzkin numbers. Finally, we consider some statistics on the new Catalan manifestation.

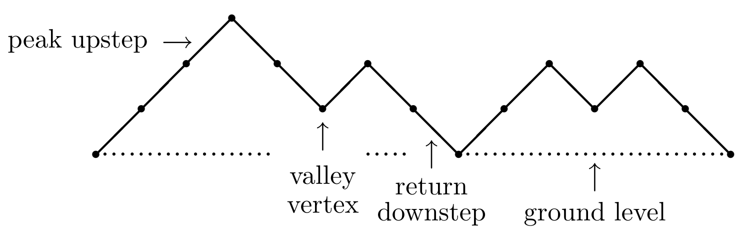
1 Introduction

There are several bijections on Dyck paths in the literature [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], usually introduced to show the equidistribution of statistics: if a bijection sends statistic A to statistic B, then clearly both have the same distribution. Another aspect of such a bijection is its cycle structure considered as a permutation on Dyck paths. Apart from involutions, this question is usually intractable. For example, Donaghey [7] introduces a bijection, gets some results on a restriction version, and notes its apparently chaotic behavior in general. In similar vein, Knuth [8] defines a conjugate (R) and transpose (T), both involutions, on ordered forests, equivalently on Dyck paths, and asks when they commute [8, Ex. 17, 7.2.1.6], equivalently, what are the fixed points of $(RT)^2$? This question is still open. (Donaghey's bijection is equivalent to the composition RT.)

In this paper, after reviewing Dyck path terminology (§2), we recursively define a new bijection F on Dyck paths (§3) and analyze its cycle structure (§4, §5). §4 treats the restriction of F to paths that avoid the subpath DUU, and involves an encounter with the Pascal matrix mod 2. §5 generalizes to arbitrary paths. This entails an explicit description of F involving a new manifestation of the Catalan numbers as certain colored forests in which each vertex is labeled with an integer composition. We show that each orbit has length a power of 2, find generating functions for orbit size, and characterize paths with given orbit size in terms of subpath avoidance. In particular, the fixed points of F are those Dyck paths that avoid $DUDD$ and $UUP^{+}DD$ where P^{+} denotes a nonempty Dyck path. §6 uses the bijection F to show the equivalence of two known manifestations of the Motzkin numbers. §7 considers some statistics on the new Catalan manifestation.

2 Dyck Path Terminology

A Dyck path, as usual, is a lattice path of upsteps $U = (1, 1)$ and downsteps $D = (1, -1)$, the same number of each, that stays weakly above the horizontal line joining its initial and terminal points (vertices). A peak is an occurrence of UD, a valley is a DU.



A Dyck 7-path with 2 components, 2DUDs, and height 3

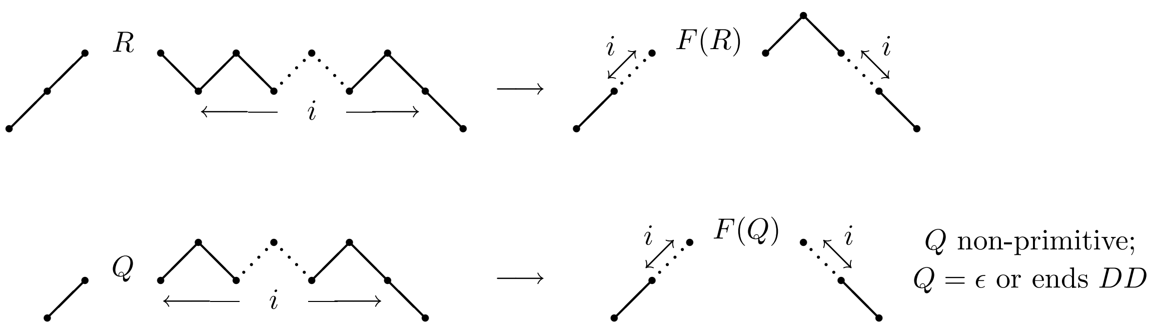
The size (or semilength) of a Dyck path is its number of upsteps and a Dyck path of size n is a Dyck n-path. The empty Dyck path (of size 0) is denoted ϵ. The number of Dyck n-paths is the Catalan number C_n, sequence A000108 in OEIS . The height of a vertex in a Dyck path is its vertical height above ground level and the height of the path is the maximum height of its vertices. A return downstep is one that returns the path to ground level. A *primitive* Dyck path is one with exactly one return (necessarily at the end). Note that the empty Dyck path ϵ is not primitive. Its returns split a nonempty Dyck path into one or more primitive Dyck paths, called its *components*. Upsteps and downsteps come in matching pairs: travel due east from an upstep to the first downstep encountered. More precisely, D_0 is the matching downstep for upstep U_0 if D_0 terminates the shortest Dyck subpath that starts with U_0. We use $\mathcal{P}$ to denote the set of primitive Dyck paths, $\mathcal{P}_n$ for n-paths, $\mathcal{P}(DUU)$ for those that avoid DUU as a subpath, and $\mathcal{P}[DUU]$ for those that contain at least one DUU. A path $UUUDUDDD$, for example, is abbreviated U^3DUD^3.

3 The Bijection

Define a size-preserving bijection F on Dyck paths recursively as follows. First, $F(\epsilon) = \epsilon$ and for a non-primitive Dyck path P with components $P_1, P_2, \ldots, P_r$ $(r \geq 2)$, $F(P) = F(P_1)F(P_2)\ldots F(P_r)$ (concatenation). This reduces matters to primitive paths. From a consideration of the last vertex at height 3 (if any), every primitive Dyck path P has the form $UQ(UD)^iD$ with $i \geq 0$ and Q a Dyck path that is either empty (in case no vertex is at height 3) or ends DD; define $F(P)$ by

$$F(P) = \begin{cases} U^i F(R) UDD^i & \text{if } Q \text{ is primitive, say } Q = URD, \text{ and} \\ U^{i+1} F(Q) D^{i+1} & \text{if } Q \text{ is not primitive.} \end{cases}$$

Schematically,



definition of F on primitive Dyck paths

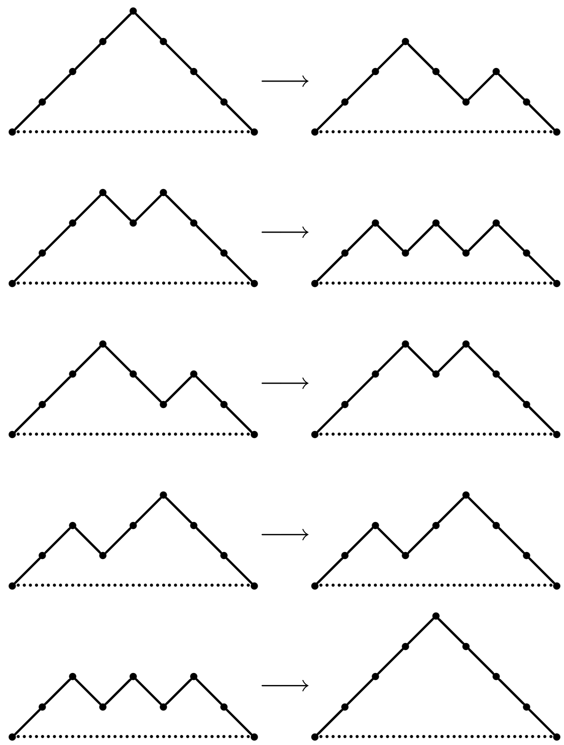
Note that $R = \epsilon$ in the top left path duplicates a case of the bottom left path but no matter: both formulas give the same result.

The map G, defined as follows, serves as an inverse of F and hence F is indeed a bijection. Again, $G(\epsilon) = \epsilon$ and for a non-primitive Dyck path P with components $P_1, P_2, \ldots, P_r$ $(r \geq 2)$, $G(P) = G(P_1)G(P_2)\ldots G(P_r)$. By considering the lowest valley vertex, every primitive Dyck path has the form $U^{i+1}QD^{i+1}$ with $i \geq 0$ and Q a non-primitive Dyck path ($Q = \epsilon$ in case valley vertices are absent); define $G(P)$ by

$$G(P) = \begin{cases} UUG(R)D(UD)^iD & \text{if } Q \text{ ends } UD, \text{ say } Q = RUD, \text{ and} \\ UG(Q)(UD)^iD & \text{otherwise.} \end{cases}$$

The bijection F is the identity on Dyck paths of size ≤ 3, except that it interchanges U^3D^3 and U^2DUD^2. Its action on primitive Dyck 4-paths is given in the Figure below.

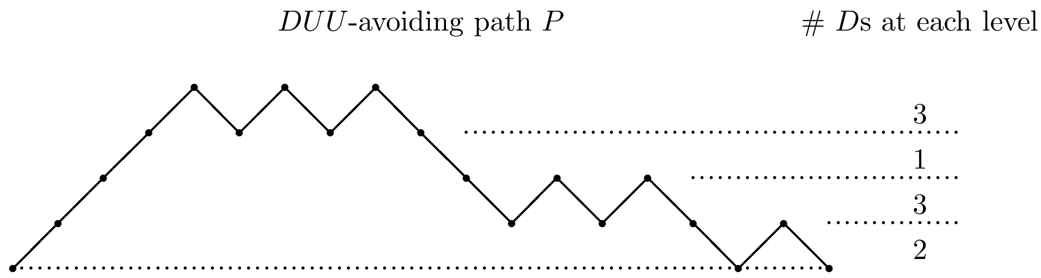
Dyck path P $\longrightarrow$ image $F(P)$



action of F on primitive Dyck 4-paths

4 Restriction to DUU-avoiding Paths

To analyze the structure of F a key property, clear by induction, is that it preserves $\#\,DUU$s, in particular, it preserves the property "path avoids DUU". A DUU-avoiding Dyck n-path corresponds to a composition $\mathbf{c} = (c_1, c_2, \ldots, c_h)$ of n via c_i = number of Ds ending at height $h-i,\ i = 1, 2, \ldots, h$ where h is the height of the path:



DUU-avoiding path P $\leftrightarrow$ composition $(3, 1, 3, 2)$

Under this correspondence, F acts on compositions of n: F is the identity on compositions of length 1, and for $\mathbf{c} = (c_i)_{i=1}^r$ with $r \geq 2$, $F(\mathbf{c})$ is the concatenation of $IncrementLast\big(F(c_1, \ldots, c_{r-2})\big)$, $1^{c_{r-1}-1}$, c_r where $IncrementLast$ means "add 1 to the last entry" and the superscript refers to repetition. In fact, F can be described explicitly on compositions of n:

Proposition 1. *For a composition* $\mathbf{c}$ *of* n, $F(\mathbf{c})$ *is given by the following algorithm. For each entry* c *in even position measured from the end* (*so the last entry is in position* 1), *replace it by* $c-1$ 1*s and increment its left neighbor.*

For example, $4\,2\,1\,5\,2\,3 = \overset{6}{4}\overset{5}{2}\overset{4}{1}\overset{3}{5}\overset{2}{2}\overset{1}{3} \to 1\ 1^3\ 3\ 1^0\ 6\ 1^1\ 3 = 1^4\,3\,6\,1\,3$. ☐

Primitive DUU-avoiding Dyck n-paths correspond to compositions of n that end with a 1. Let $\mathcal{C}_n$ denote the set of such compositions. Thus $|\mathcal{C}_1| = 1$ and for $n \geq 2$, $|\mathcal{C}_n| = 2^{n-2}$ since there are 2^{n-2} compositions of $n-1$.

Denote the length of a composition $\mathbf{c}$ by $\#\mathbf{c}$. The *size* of $\mathbf{c}$ is the sum of its entries. The *parity* of $\mathbf{c}$ is the parity (even/odd) of $\#\mathbf{c}$. There are two operations on nonempty compositions that increment (that is, increase by 1) the size: P = prepend 1, and I = increment first entry. For example, for $\mathbf{c} = (4, 1, 1)$ we have size$(\mathbf{c}) = 6$, $\#\mathbf{c} = 3$, the parity of $\mathbf{c}$ is odd, $P(\mathbf{c}) = (1, 4, 1, 1)$, $I(\mathbf{c}) = (5, 1, 1)$.

Lemma 2. *P changes the parity of a composition while I preserves it.* ☐

We'll call P and I *augmentation operators* on $\mathcal{C}_n$ and for A an augmentation operator, A' denotes the other one.

Lemma 3. *Let A be an augmentation operator. On a composition* $\mathbf{c}$ *with* $\#\mathbf{c} \geq 2$, $A \circ F = F \circ A$ *if* $\#\mathbf{c}$ *is odd and* $A \circ F = F \circ A'$ *if* $\#\mathbf{c}$ *is even.*

This follows from Proposition 1. □

Using Lemma 3, an F-orbit $(\mathbf{c}_1, \ldots, \mathbf{c}_m)$ in $\mathcal{C}_n$ together with an augmentation operator $A_1 \in \{P, I\}$ yields part of an F-orbit in $\mathcal{C}_{n+1}$ via a "commutative diagram" as shown:

$$\begin{array}{ccccccccccccccc}
\mathbf{c}_1 & \xrightarrow{F} & \mathbf{c}_2 & \xrightarrow{F} & \ldots & \xrightarrow{F} & \mathbf{c}_i & \xrightarrow{F} & \mathbf{c}_{i+1} & \xrightarrow{F} & \ldots & \xrightarrow{F} & \mathbf{c}_m & \xrightarrow{F} & \mathbf{c}_1 \\
\downarrow{\scriptstyle A_1} & & \downarrow{\scriptstyle A_2} & & & & \downarrow{\scriptstyle A_i} & & \downarrow{\scriptstyle A_{i+1}} & & & & \downarrow{\scriptstyle A_m} & & \downarrow{\scriptstyle A_{m+1}} \\
\mathbf{d}_1 & \xrightarrow{F} & \mathbf{d}_2 & \xrightarrow{F} & \ldots & \xrightarrow{F} & \mathbf{d}_i & \xrightarrow{F} & \mathbf{d}_{i+1} & \xrightarrow{F} & \ldots & \xrightarrow{F} & \mathbf{d}_m & \xrightarrow{F} & \mathbf{d}_{m+1}
\end{array}$$

Let $B(\mathbf{c}_1, A_1)$ denote the sequence of compositions $(\mathbf{d}_1, \ldots, \mathbf{d}_m)$ thus produced. By Lemma 3, $A_{i+1} = A_i$ or A_i' according as $\#\mathbf{c}_i$ is odd or even $(1 \le i \le m)$. Hence, if the orbit of $\mathbf{c}_1$ contains an even number of compositions of even parity, then $A_{m+1} = A_1$ and so $\mathbf{d}_{m+1} = \mathbf{d}_1$ and $B(\mathbf{c}_1, A_1)$ is a complete F-orbit in $\mathcal{C}_{n+1}$ for each of $A_1 = P$ and $A_1 = I$. On the other hand, if the orbit of $\mathbf{c}_1$ contains an odd number of compositions of even parity, then $A_{m+1} = A_1'$ and the commutative diagram will extend for another m squares before completing an orbit in $\mathcal{C}_{n+1}$, consisting of the concatenation of $B(\mathbf{c}_1, P)$ and $B(\mathbf{c}_1, I)$, denoted $B(\mathbf{c}_1, P, I)$. In the former case orbit size is preserved; in the latter it is doubled.

Our goal here is to generate F-orbits recursively and to get induction going, we now need to investigate the parities of the compositions comprising these "bumped-up" orbits $B(\mathbf{c}, A)$ and $B(\mathbf{c}, P, I)$. A bit sequence is a sequence of 0s and 1s. **In the sequel all operations on bit sequences are modulo 2**. Let $\mathbf{S}$ denote the partial sum operator on bit sequences: $\mathbf{S}\big((\epsilon_1, \epsilon_2, \ldots, \epsilon_m)\big) = (\epsilon_1, \epsilon_1 + \epsilon_2, \ldots, \epsilon_1 + \epsilon_2 + \ldots + \epsilon_m)$. Let $\mathbf{e}_m$ denote the all 1s bit sequence of length m and let $\mathbf{e}$ denote the infinite sequences of 1s. Thus $\mathbf{Se} = (1, 0, 1, 0, 1, \ldots)$. Let P denote the infinite matrix whose ith row $(i \ge 0)$ is $\mathbf{S}^i\mathbf{e}$ ($\mathbf{S}^i$ denotes the i-fold composition of $\mathbf{S}$). The (i, j) entry p_{ij} of P satisfies $p_{ij} = p_{i-1,j} + p_{i,j-1}$ and hence P is the symmetric Pascal matrix mod 2 with (i, j) entry $= \binom{i+j}{i}$ mod 2. The following lemma will be crucial.

Lemma 4. *Fix $k \ge 1$ and let P_k denote the $2^k \times 2^k$ upper left submatrix of P. Then the sum modulo 2 of row i in P_k is 0 for $0 \le i \le 2^k - 1$ and is 1 for $i = 2^k - 1$.*

Proof The sum of row i in P_k is, modulo 2,

$$\sum_{j=0}^{2^k-1} p_{ij} = \sum_{j=0}^{2^k-1} \binom{i+j}{i} = \binom{i+2^k}{i+1} = \binom{i+2^k}{i+1,\, 2^k-1}$$

and for $i < 2^k - 1$ there is clearly at least one carry in the addition of $i+1$ and $2^k - 1$ in base 2 so that, by Kummer's well known criterion, $2 \mid \binom{i+2^k}{i+1,2^k-1}$ and the sum of row i is 0 (mod 2). On the other hand, for $i = 2^k - 1$ there are no carries, so $2 \nmid \binom{i+2^k}{i+1,2^k-1}$ and the sum of row i is 1 (mod 2). □

Now let $p(\mathbf{c})$ denote the mod-2 parity of a composition $\mathbf{c}$: $p(\mathbf{c}) = 1$ if $\#\mathbf{c}$ is odd, $= 0$ if $\#\mathbf{c}$ is even. For purposes of addition mod 2, represent the augmentation operators P and I by 0 and 1 respectively so that, for example, $p(A(\mathbf{c})) = p(\mathbf{c}) + A + 1$ for $A = P$ or I by Lemma 2. Then the parity of $\mathbf{d}_{i+1}$ above can be obtained from the following commutative diagram (all addition modulo 2)

$$\begin{array}{ccc} p(\mathbf{c}_i) & \longrightarrow & p(\mathbf{c}_{i+1}) \\ \Big\downarrow{\scriptstyle A} & & \Big\downarrow{\scriptstyle p(\mathbf{c}_i)+A+1} \\ \cdots & \longrightarrow & p(\mathbf{c}_{i+1}) + p(\mathbf{c}_i) + A \end{array}$$

This leads to

Lemma 5. *Let p_i denote the parity of $\mathbf{c}_i$ so that $\mathbf{p} = (p_i)_{i=1}^m$ is the parity vector for the F-orbit $(\mathbf{c}_i)_{i=1}^m$ of the composition $\mathbf{c}_1$. Then the parity vector for $B(\mathbf{c}, A)$ is*

$$\mathbf{Sp} + \mathbf{Se}_m + (A+1)\mathbf{e}_m.$$

□

Now we are ready to prove the main result of this section concerning the orbits of F on primitive DUU-avoiding Dyck n-paths identified with the set $\mathcal{C}_n$ of compositions of n that end with a 1. The parity of an orbit is the sum mod 2 of the parities of the compositions comprising the orbit, in other words, the parity of the total number of entries in all the compositions.

Theorem 6. *For each $n \geq 1$,*

(*i*) *all F-orbits on $\mathcal{C}_n$ have the same length and this length is a power of 2.*

(*ii*) *all F-orbits on $\mathcal{C}_n$ have the same parity.*

(*iii*) *the powers in (i) and the parities in (ii) are given as follows:*

For $n = 1$, the power (i.e. the exponent) is 0 and the parity is 1. For $n = 2$, the power and parity are both 0. As n increases from 2, the powers remain unchanged *and the parity stays* 0 *except that when* n *hits a number of the form* 2^k+1*, the parity becomes* 1*, and at the next number,* 2^k+2*, the power increases by* 1 *and the parity reverts to* 0*.*

Proof We consider orbits generated by the augmentation operators P and I. No orbits are missed because all compositions, in particular those ending 1, can be generated from the unique composition of 1 by successive application of P and I. The base cases $n = 1, 2, 3$ are clear from the orbits $(1) \to (1),\ (1,1) \to (1,1),\ (2,1) \to (1,1,1) \to (2,1)$. To establish the induction step, suppose given an orbit, $\mathrm{orb}(\mathbf{c})$, in $\mathcal{C}_{2^k+1}$ $(k \geq 1)$ with parity vector $\mathbf{p} = (a_i)_{i=1}^{2^k}$ and (total) parity 1. Then the next orbit $B(\mathbf{c}, P, I)$ has parity vector

$$\mathbf{p}_1 = (\mathbf{S}\,\mathbf{p}, \mathbf{S}\,\mathbf{p} + \mathbf{e}_{2^k}) + \mathbf{S}\,\mathbf{e}_{2^{k+1}}$$

with parity ($\mathbf{S}\,\mathbf{p}$'s cancel out) $\underbrace{1+1+\ldots+1}_{2^k} + \underbrace{1+0+1+0+\ldots+1+0}_{2^{k+1}} = 0$ for $k \geq 1$. Successively "bump up" this orbit using $A = \epsilon_1, \epsilon_2, \ldots,$ in turn until the parity hits 1 again. With $\mathrm{Sum}(\mathbf{v})$ denoting the sum of the entries in $\mathbf{v}$, the successive parity vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots$ are given by

$$\mathbf{p}_i = \big(\mathbf{S}^i\mathbf{p}, \mathbf{S}^i\mathbf{p} + \sum_{j=1}^{i-2} \mathrm{Sum}(\mathbf{S}^j\mathbf{p})\mathbf{S}^{i-1-j}\mathbf{e}_{2^k} + \mathbf{S}^{i-1}\mathbf{e}_{2^k}\big) +$$
$$\mathbf{S}^i\mathbf{e}_{2^{k+1}} + \mathbf{S}^{i-1}\mathbf{e}_{2^{k+1}} + \sum_{j=1}^{i-2} \epsilon_j \mathbf{S}^{i-1-j}\mathbf{e}_{2^{k+1}} + (\epsilon_{i-1}+1)\mathbf{e}_{2^{k+1}}.$$

Applying Lemma 4 we see that, independent of the ϵ_i's, $\mathbf{p}_i$ has sum 0 for $i < 2^k - 1$ and sum 1 for $i = 2^k - 1$. This establishes the induction step in the theorem. □

Corollary 7. *For* $n \geq 2$*, the length of each* F*-orbit in* $\mathcal{P}_n(DUU)$ *is* 2^k *where* k *is the number of bits in the base-*2 *expansion of* $n-2$*.*

Proof This is just a restatement of part of the preceding Theorem. □

5 The Orbits of F

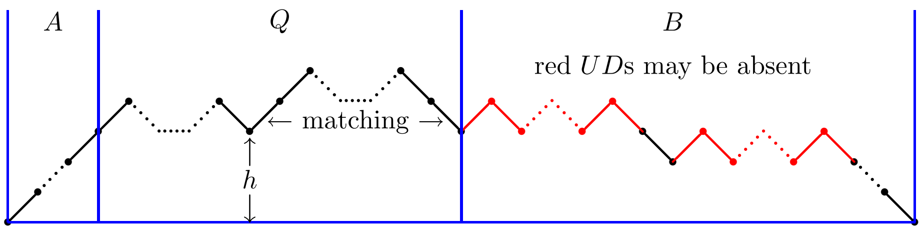
The preceding section analyzed F on $\mathcal{P}(DUU)$, paths avoiding DUU. Now we consider F on $\mathcal{P}[DUU]$, the primitive Dyck paths containing a DUU. Every $P \in \mathcal{P}[DUU]$ has the form AQB where

(i) A consists of one or more Us

(ii) $C := AB \in \mathcal{P}(DUU)$

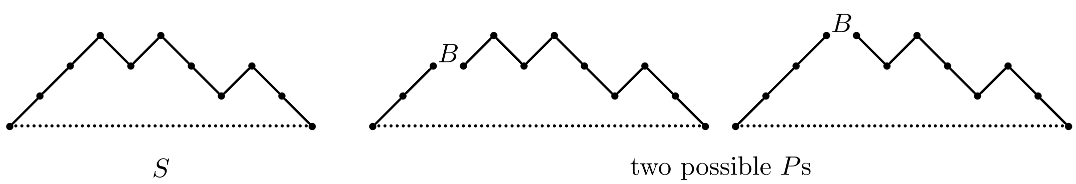
(iii) $Q \notin \mathcal{P}$ and Q ends DD (and hence Q contains a DUU at ground level).

To see this, locate the rightmost of the lowest DUUs in P, say at height h. Then $A = U^h$, Q starts at step number $h + 1$ and extends through the matching downstep of the middle U in this rightmost lowest DUU, and B consists of the rest of the path.



The AQB decomposition of a path containing a DUU

Call the path AB the (DUU-avoiding) *skeleton* of P and Q the (DUU-containing) *body* of P. In case $P \in \mathcal{P}(DUU)$, its skeleton is itself and its body is empty. If the skeleton of P is UD, then P is uniquely determined by its skeleton and body. On the other hand, a skeleton of size ≥ 2 and a nonempty body determine precisely two paths P in $\mathcal{P}[DUU]$, obtained by inserting the body at either the top or the bottom of the first peak upstep in the skeleton, as illustrated.



Recapturing a path $P \in \mathcal{P}[DUU]$ from a skeleton S and body B

Thus paths in $\mathcal{P}[DUU]$ correspond bijectively to triples (S, B, pos) where $S \in \mathcal{P}(DUU)$ is the skeleton, $B \neq \epsilon$ is the body, and $pos = top$ or bot according as B is positioned at the top or bottom of the first peak upstep in S, with the proviso that $pos = top$ if $S = UD$.

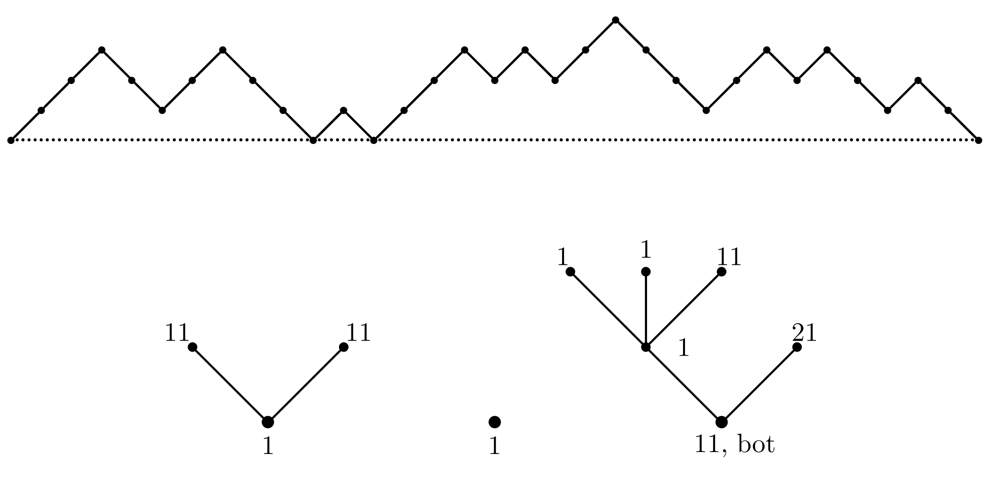
In these terms, F can be specified on $\mathcal{P}[DUU]$ as follows.

Proposition 8.

$$F\big((S,B,pos)\big) = \begin{cases} (F(S), F(B),\ pos\ )\ \textit{if height}(S)\ \textit{is odd, and} \\ (F(S), F(B),\ pos'\ )\ \textit{if height}(S)\ \textit{is even.} \end{cases}$$

Proof Let $h(P)$ denote the height of the terminal point of the lowest DUU in $P \in \mathcal{P}[DUU]$. The result clearly holds for $h(P) = 1$. If $h(P) \geq 2$, then P has the form $U^2Q(UD)^aD(UD)^bD$ with $a, b \geq 0$ and Q a Dyck path that ends DD. So $F(P) = U^{b+1}F(Q)(UD)^{a+1}D^{b+1}$ and $h(Q) = h(P) - 2$. These two facts are the basis for a proof by induction that begins as follows. If $h(Q) = 0$, then the body of $F(P)$ has position = bottom, while the body of P has position bottom or top according as $a \geq 1$ or $a = 0$. In the former case, the skeleton of P has height 3 and position has been preserved, in the latter height 2 and position has been reversed. $\square$

Iterating the skeleton-body-position decomposition on each component, a Dyck path has a forest representation as illustrated below. Each vertex represents a skeleton and is labeled with the corresponding composition. When needed, a color (*top* or *bot*) is also applied to a vertex to capture the position of that skeleton's body.



A Dyck path and corresponding LCO forest

The 3 trees in the forest correspond to the 3 components of the Dyck path. The skeleton of the first component is UD and its body has 2 identical components, each consisting of a skeleton alone, yielding the leftmost tree. The skeleton of the third component is $UUDD$ and its body is positioned at the bottom of its first peak upstep, and so on. Call this forest the LCO (labeled, colored, ordered) forest corresponding to the Dyck path. Here is the precise definition.

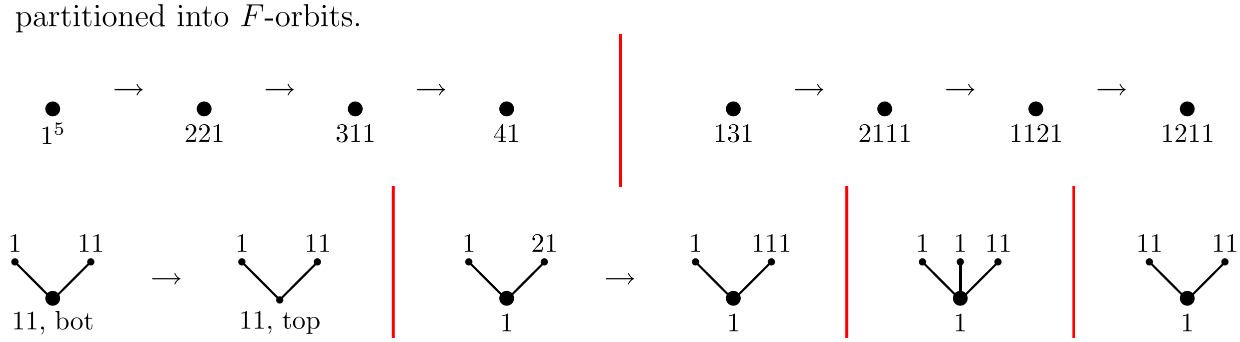
Definition. *An LCO forest is a labeled, colored, ordered forest such that*

- *the underlying forest consists of a list of ordered trees (a tree may consist of a root only)*
- *no vertex has outdegree* 1 *(i.e., exactly one child)*
- *each vertex is labeled with a composition that ends* 1
- *each vertex possessing children and labeled with a composition of size* ≥ 2 *is also colored top or bot*
- *For each leaf (i.e. vertex with a parent but no child) that is the rightmost child of its parent, its label composition has size* ≥ 2.

The *size* of an LCO forest is the sum of the sizes of its label compositions. The correspondence Dyck path ↔ LCO forest preserves size, and primitive Dyck paths correspond to one-tree forests. Thus we have

Proposition 9. *The number of LCO forests of size* n *is the Catalan number* C_n*, as is the number of one-tree LCO forests of size* $n+1$*.* □

The $C_4 = 14$ one-tree LCO forests corresponding to primitive Dyck 5-paths are shown, partitioned into F-orbits.



The LCO one-tree forests of size 5, partitioned into F-orbits

We can now give an explicit description of F on Dyck paths identified with LCO forests. On an LCO forest, F acts as follows:

- the underlying list of ordered trees is preserved
- each label $\mathbf{c}$ becomes $F(\mathbf{c})$ as defined in Prop. 1
- each color (*top*/*bot*) is preserved or switched according as the associated label $\mathbf{c}$ has odd or even length.

From this description and Cor. 7, the size of the F-orbit of a Dyck path P can be determined as follows. In the LCO forest for P, let ℓ denote the maximum size of a leaf label and i the maximum size of an internal (i.e., non-leaf) label (note that an isolated root is an internal vertex). Let k denote the number of bits in the base-2 expansion of $\max\{\ell - 2, i - 1\}$. Then the F-orbit of P has size 2^k.

It is also possible to specify orbit sizes in terms of subpath avoidance. For Dyck paths Q and R, let Q *top* R (resp. Q *bot* R) denote the Dyck path obtained by inserting R at the top (resp. bottom) of the first peak upstep in Q. Then the F-orbit of a Dyck path P has size $\le 2^k$ iff P avoids subpaths in the set $\{Q\ top\ R,\ Q\ bot\ R : R \ne \epsilon,\ Q \in \mathcal{P}_i(DUU),\ 2^{k-1} + 1 < i \le 2^k + 1\}$. For $k \ge 1$, listing these Qs explicitly would give $2^{2^k} - 2^{2^{k-1}}$ proscribed patterns of the form $Q\ top\ R,\ R \ne \epsilon$ (and the same number of the form $Q\ bot\ R$). For $k = 0$, that is, for fixed points of F, the proscribed patterns are UP^+UDD and UUP^+DD with P^+ a nonempty Dyck path, and avoiding the first of these amounts to avoiding the subpath $DUDD$.

The generating function for the number of F-orbits of size $\le 2^k$ can be found using the "symbolic" method [15]. With $F_k(x)$, $G_k(x)$, $H_k(x)$ denoting the respective generating functions for general Dyck paths, primitive Dyck paths, and primitive Dyck paths that end DD (x always marking size), we find

$$\begin{aligned} F_k(x) &= 1 + G_k(x)F_k(x) \\ G_k(x) &= x + \frac{x(1-(2x)^{2^k}}{1-2x}\big(x + (F_k(x) - 1)H_k(x)\big) \\ H_k(x) &= G_k(x) - x \end{aligned}$$

leading to

$$F_k(x) = \frac{1 - a_k - \sqrt{1 - 4x - \frac{a_k(2-a_k)x}{1-x}}}{2x - a_k},$$

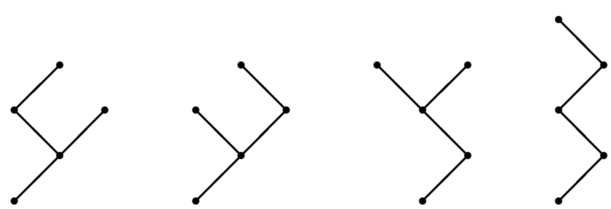
where $a_k = (2x)^{2^k+1}$. In this formulation it is clear, as expected, that $\lim_{k\to\infty} F_k(x) = \frac{1-\sqrt{1-4x}}{2x}$, the generating function for the Catalan numbers. The counting sequence for fixed points of F, with generating function $F_0(x)$, is sequence A086625 in OEIS .

6 An Application

Ordered trees and binary trees are manifestations of the Catalan numbers A000108 . Donaghey [12, 13] lists several types of restricted tree counted by the Motzkin numbers A001006 . In particular, the following result is implicit in item III C of [13].

Proposition 10. *The Motzkin number M_n counts right-planted binary trees on $n+1$ edges with no erasable vertices.*

Here, planted means the root has only one child, and erasable refers to a vertex incident with precisely 2 edges *both of the same slope*—the vertex could then be erased, preserving the slope, to produce a smaller binary tree. The $M_3 = 4$ such trees on 4 edges are shown.



The right-planted binary 4-trees with no erasable vertices

Translated to Dyck paths, Prop. 10 is equivalent to

Proposition 11. *M_n counts Dyck $(n+1)$-paths that end DD and avoid subpaths $DUDU$ and UUP^+DD with P^+ denoting a nonempty Dyck subpath.*

We will use F to give a bijective proof of Prop. 11 based on the fact [14] that M_n also counts DUD-avoiding Dyck $(n+1)$-paths. (Of course, path reversal shows that $\#\,UDU$s and $\#\,DUD$s are equidistributed on Dyck paths.) Define statistics X and Y on Dyck paths by $X = \#\,DUD$s and $Y = \#\,DUDU$s $+ \#\,UUP^+DD$s $+$ [paths ends with UD] (Iverson notation) so that the paths in Prop. 11 are those with $Y = 0$. Prop. 11 then follows from

Proposition 12. *On Dyck n-paths with $n \geq 2$, F sends the statistic X to the statistic Y.*

Proof Routine by induction from the recursive definition of F. However, using the explicit form of F, it is also possible to specify precisely which DUDs correspond to each of the three summands in Y. For this purpose, given a DUD in a Dyck path P, say $D_1U_2D_3$ (subscripts used simply to identify the individual steps), let $S(D_1U_2D_3)$ denote the longest Dyck subpath of P containing $D_1U_2D_3$ in its skeleton and let h denote the height at which $D_1U_2D_3$ terminates in $S(D_1U_2D_3)$. If h is odd, $D_1U_2D_3$ is immediately followed in P by D_4 or by UD_4 (it cannot be followed by UU). In either case, let U_4 be the matching upstep for D_4. Then the steps D_1, U_2, D_3, U_4 show up in $F(P)$ as part of a subpath $U_4U_2P^+D_3D_4$ with P^+ a Dyck path that ends D_1. On the other hand, if h is even, $D_1U_2D_3$ either (i) ends the path (here $S(D_1U_2D_3) = P$ and $h = 0$) or is immediately followed by (ii) U_4 or (iii) D. In case (iii), let U_4 be the matching upstep. Then D_1, U_2, D_3, U_4 show up in $F(P)$ as a subpath in that order (cases (ii) and (iii)) or $F(P)$ ends U_2D_3 (case (i)). The details are left to the reader.

7 Statistics Suggested by LCO Forests

There are various natural statistics on LCO forests, some of which give interesting counting results. Here we present two such. First let us count one-tree LCO forests by size of root label. This is equivalent to counting primitive Dyck paths by skeleton size. Recall that the generalized Catalan number sequence $\left(C_n^{(j)}\right)_{n\geq 0}$ with $C_n^{(j)} := \frac{j}{2n+j}\binom{2n+j}{n}$ is the j-fold convolution of the ordinary Catalan number sequence A000108. (See [16] for a nice bijective proof.) And, as noted above, in the skeleton-body-position decomposition of a primitive Dyck path, if the body is nonempty it contains a DUU at (its own) ground level and ends DD.

Lemma 13. *The number of Dyck n-paths that contain a DUU at ground level and end DD is $C_{n-3}^{(4)}$.*

Proof In such a path, let U_0 denote the middle U of the *last* DUU at ground level. The path then has the form AU_0BD where A and B are arbitrary *nonempty* Dyck paths, counted by $C_{n-1}^{(2)}$. So the desired counting sequence is the convolution of $\left(C_{n-1}^{(2)}\right)$ with itself and, taking the U_0D into account, the lemma follows. □

The number of primitive DUU-avoiding Dyck k-paths is 1 if $k = 1$, and 2^{k-2} if $k \geq 2$. But if $k \geq 2$, there are two choices (top/bottom) to insert the body. So the number of primitive Dyck $(n+1)$-paths with skeleton size k is $2^{k-1}C_{n-k-2}^{(4)}$ for $1 \leq k \leq n-2$ and is 2^{n-1} for $k = n+1$. Since there are C_n primitive Dyck $(n+1)$-paths altogether, we have established the following identity.

Proposition 14.

$$C_n = 2^{n-1} + \sum_{k=1}^{n-2} \frac{2^k}{n-k} \binom{2n-2k}{n-2-k}.$$

□

Lastly, turn an LCO forest into an LCO tree by joining all roots to a new root. The purpose of doing this is so that isolated roots in the forest will qualify as leaves in the tree. The symbolic method then yields

Proposition 15. *The generating function for LCO trees by number of leaves* (x *marks size,* y *marks number of leaves*) *is*

$$\frac{1 - \sqrt{1 - 4x\,\frac{1-x}{1-xy}}}{2x}.$$

The first few values are given in the following table.

$n \backslash k$	1	2	3	4	5	6	7	8
1	1							
2	1	1						
3	2	2	1					
4	4	6	3	1				
5	8	17	12	4	1			
6	16	46	44	20	5	1		
7	32	120	150	90	30	6	1	
8	64	304	482	370	160	42	7	1

number of LCO trees of size n with k leaves

References

[1] David Callan, Two bijections for Dyck path parameters, math.CO/0406381, 2004, 4pp.

[2] David Callan, Some bijections and identities for the Catalan and Fine numbers, Sém. Lothar. Combin. **53** (2004/06), Art. B53e, 16 pp.

[3] Emeric Deutsch, An involution on Dyck paths and its consequences. *Discrete Math.* **204** (1999), no. 1-3, 163–166.

[4] Emeric Deutsch, A bijection on Dyck paths and its consequences, *Discrete Math.* **179** (1998), no. 1-3, 253–256.

[5] Emeric Deutsch, A bijection on ordered trees and its consequences, *J. Combin. Theory Ser. A* **90** (2000), no. 1, 210–215.

[6] Emeric Deutsch and Sergi Elizalde, A simple and unusual bijection for Dyck paths and its consequences, *Ann. Comb.* **7** (2003), no. 3, 281–297.

[7] Robert Donaghey, Automorphisms on Catalan trees and bracketings. *J. Combinatorial Theory Ser. B* **29** (1980), no. 1, 75–90. MR0584162

[8] Donald Knuth, *Art of Computer Programming, Vol.4, Fascicle 4: Generating all Trees – History of Combinatorial Generation*, Addison-Wesley, 2006, vi+120pp, draft available from http://www-cs-faculty.stanford.edu/~knuth/fasc4a.ps.gz

[9] J.-C. Lalanne, Une involution sur les chemins de Dyck, *European J. Combin.* **13** (1992), no. 6, 477–487.

[10] J.-C. Lalanne, Sur une involution sur les chemins de Dyck, Conference on Formal Power Series and Algebraic Combinatorics *Theoret. Comput. Sci.* **117** (1993), no. 1-2, 203–215.

[11] J. Vaillé, Une bijection explicative de plusieurs propriétés remarquables des ponts, *European J. Combin.* **18** (1997), no. 1, 117–124.

[12] Robert Donaghey and Louis Shapiro, Motzkin numbers, *J. Combinatorial Theory Ser. A* **23**, 291–301, 1977. MR0505544

[13] Robert Donaghey, Restricted plane tree representations of four Motzkin-Catalan equations, *J. Combinatorial Theory Ser. B* **22**, (1977), no. 2, 114–121, 1977. MR0432532

[14] Y. Sun, The statistic "number of udu's" in Dyck paths, *Disc. Math.*, **287** (2004), Issue 1-3 (October 2004), 177-186.

[15] Robert Sedgewick and Philippe Flajolet, An Introduction to the Analysis of Algorithms, Addison-Wesley, 1996.

[16] Wen-jin Woan, Uniform partitions of lattice paths and Chung-Feller generalizations. *Amer. Math. Monthly* 108 (2001), no. 6, 556–559.