A bijective proof of
Shapiro's Catalan convolution

Péter Hajnal   Gábor V. Nagy

Bolyai Institute
University of Szeged
Szeged, Hungary

{hajnal,ngaba}@math.u-szeged.hu

Abstract

We present a bijective proof of Shapiro's convolution formula involving Catalan numbers of even index. As a corollary, we give a new interpretation of the Catalan numbers.

Keywords: Catalan numbers; Shapiro convolution; bijective proof

1   Introduction

In 2002, L. Shapiro found the following elegant convolution formula involving Catalan numbers of even index [2, p. 123]:

Theorem 1.
$$\sum_{k=0}^{n} C_{2k}C_{2n-2k} = 4^n C_n, \tag{1}$$

where C_n denotes the nth Catalan number, i.e. $C_n = \frac{1}{n+1}\binom{2n}{n}$.

This identity is listed in Stanley's Bijective Proof Problems [5, p. 46] as unsolved. It can be easily proved using generating functions, but no bijective proof has been known (see [1] and [3] for two different combinatorial arguments). In this paper we improve the combinatorial proof of our previous paper [3] by replacing the only non-bijective part with a bijective one.

So let us recall the required definitions and lemmas (for more details see [3]). A (conventional) *path of length l* is an *l*-element sequence of up-steps ($\nearrow$) and down-steps ($\searrow$). We interpret paths in the 2-dimensional coordinate system in the usual way: Unless

otherwise stated, they start from the origin, $\nearrow$ is a step $(1,1)$ and $\searrow$ is a step $(1,-1)$; see the figures below. A path is *balanced*, if it ends on the x-axis, and it is *non-negative*, if it never falls below the x-axis. Non-negative balanced (conventional) paths are also called *Dyck-paths*. It is well known that C_n counts the number of Dyck-paths of length $2n$. We call a (balanced or non-balanced) path *even-zeroed*, if its x-intercepts are all divisible by 4.

The key lemma for the combinatorial interpretation of the left-hand side of (1) has been proved bijectively (Lemma 3 of [3]):

Lemma 2. C_{2n} *counts the number of even-zeroed balanced paths of length* $4n$.

As a corrigendum, we note that, after publishing our previous paper, it turned out that Lemma 2 has been already known earlier: It has been posed as an American Mathematical Monthly problem and it was first solved bijectively by W. Nichols [4], see [6, p. 266, 6.22] for the full history.

The following fact (Lemma 6.b of [3]) is an easy corollary of Lemma 2:

Lemma 3. $\sum_{k=0}^{n} C_{2k} C_{2n-2k}$ *counts the number of even-zeroed paths from the origin to* $(4n+1, 1)$.

To obtain a bijective proof of Theorem 1, we enumerate these paths bijectively in the next section:

Lemma 4. *The number of even-zeroed paths from the origin to* $(4n+1,1)$ *is* $4^n C_n$.

We will end our paper with an open problem that generalizes Lemma 4.

2   Proof of Lemma 4

Fix an arbitrary n. For technical reasons, we remove the first step from the paths in Lemma 4 and, in addition, if the first step was a down-step, then we reflect the obtained path across the x-axis, see the left side of Figure 1. So we will prove the following equivalent form of the lemma:

Lemma 5. *The number of even-zeroed paths starting from* $(1,1)$ *and ending at* $(4n+1,1)$ *or* $(4n+1,-1)$ *is* $4^n C_n$.

Proof. Let $\mathcal{E}$ denote the set of even-zeroed paths in the lemma. (We may write $\mathcal{E}(n)$, if we want to indicate n in the notation, and we follow the same convention throughout the paper.) We call a Dyck-path 4-labeled, if each step in even position is labeled with a number from $\{0,1,2,3\}$. (We will write labeled steps in the form $\nearrow_{\mathbf{0}}$, $\searrow_{\mathbf{3}}$, and so on.) Let $\mathcal{D}_4$ denote the set of 4-labeled Dyck-paths of length $2n$. Clearly, $|\mathcal{D}_4| = 4^n C_n$, so it is enough to give a bijection between $\mathcal{E}$ and $\mathcal{D}_4$.

Pick an arbitrary path $P \in \mathcal{E}$. We view P as a sequence $o_1 \boldsymbol{e_1} o_2 \boldsymbol{e_2} \cdots o_n \boldsymbol{e_n}$, where each o_i is a single step and each $\boldsymbol{e_i}$ is a sequence of three steps (that is, $o_i \in \{\nearrow, \searrow\}$ and $\boldsymbol{e_i} \in \{\nearrow, \searrow\}^3$). The reason for this somewhat unintuitive grouping of the steps of P

is that the steps o_i are exactly those that could hit the x-axis in a forbidden point. We are going to replace each $\boldsymbol{e_i}$ with a 4-labeled step in a way to obtain a path of $\mathcal{D}_4$. As a first phase, for each $\boldsymbol{e_i}$, we just encode $\searrow\!\nearrow\!\nearrow$, $\nearrow\!\searrow\!\nearrow$ and $\nearrow\!\nearrow\!\searrow$ as labeled "short" up-steps $\nearrow_{\boldsymbol{1}}$, $\nearrow_{\boldsymbol{2}}$ and $\nearrow_{\boldsymbol{3}}$, respectively; encode $\nearrow\!\nearrow\!\nearrow$ as a "long" up-step $(1,3)$; and make similar conversions for the 4 remaining "downward" types, see Figure 1. It is quite obvious that the path P is in $\mathcal{E}$, if and only if the obtained "compressed" $2n$-length path P^* never hits the x-axis (but it is allowed to jump over the x-axis at the long steps). After a translation of 1 unit to the left, we assume that P^* starts from $(0,1)$. We denote by $\mathcal{E}_3^*$ the set $\{P^* : P \in \mathcal{E}\}$, i.e. the set of such paths starting from $(0,1)$ and ending at $(2n, \pm1)$ that contain either a long step $(1, \pm3)$ or a 3-labeled short step $(1, \pm1)$ in each even position and never hit the x-axis. $\mathcal{E}^*$ denotes the same set but with the labels omitted. In sum, the first phase establishes a natural bijection between $\mathcal{E}$ and $\mathcal{E}_3^*$.
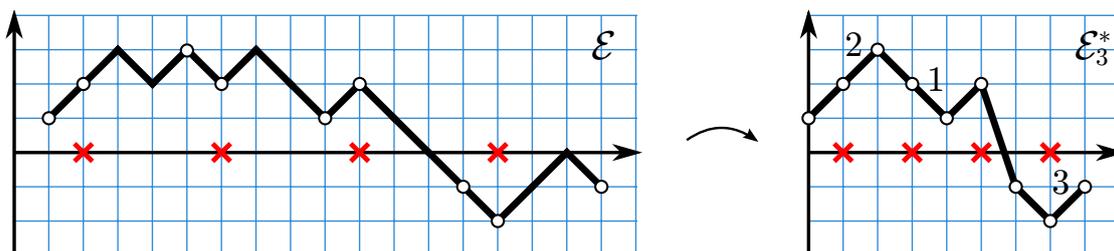
Figure 1: Illustration of the first phase

The bijection $\mathcal{E}_3^* \to \mathcal{D}_4$ is the difficult part of our construction. (It is not easy to "replace" the long steps with steps $\nearrow_{\boldsymbol{0}}$ and $\searrow_{\boldsymbol{0}}$, because such a change translates the rest of the path vertically which is difficult to control. In fact, "moving" these steps is also necessary, see the remark at the end of this section.) The essential part is provided by the following lemma:

Lemma 6. *Let* $\mathcal{D}_2$ *denote the set of such Dyck-paths of length* $2n$ *in which each step in even position is* 2*-labeled (either marked or unmarked). Then there exists a bijection* $\phi\colon \mathcal{E}^* \to \mathcal{D}_2$ *such that, for all* $P \in \mathcal{E}^*$, *the number of marked steps in* $\phi(P)$ *is equal to the number of long steps in* P.

Before proving this, we note that it quickly implies Lemma 5, because ϕ can be adopted to 3-labeled paths to obtain a required bijection $\psi\colon \mathcal{E}_3^* \to \mathcal{D}_4$. To see this, pick an arbitrary path $Q \in \mathcal{E}_3^*$ and let Q' denote the same path but with the labels removed ($Q' \in \mathcal{E}^*$). Let us define $\psi(Q)$ to be the path that is obtained from $\phi(Q')$ by labeling its marked steps with 0 and copying the labels 1, 2, 3 from Q to the remaining unlabeled even positions of $\phi(Q')$, from left to right. The cardinality condition on ϕ guarantees that this can be done. (Any other "bijective" way of copying is appropriate, we will suggest a more natural one later, based on the proof of Lemma 6.) We leave the reader to check that ψ is indeed a bijection. $\qquad\square$

Proof of Lemma 6. Pick an arbitrary path $P_0 \in \mathcal{E}^*$. There may be some long steps in P_0 where the path jumps over the x-axis; these steps will be replaced with marked (short) down-steps (from height 2 to height 1), see Figure 2: The first such long step is a down-step (from height 2 to height -1, by the parity condition on long steps), which we replace with a marked down-step and reflect the rest of the path across the x-axis, and do the same procedure for the second "jumping" long step (which also becomes a down-step after the first reflection), and so on, all the jumping long steps are replaced in this way from left to right. It is clear that the obtained path P is positive (i.e. never hits the x-axis and never goes below it) and ends at $(2n, 1)$.
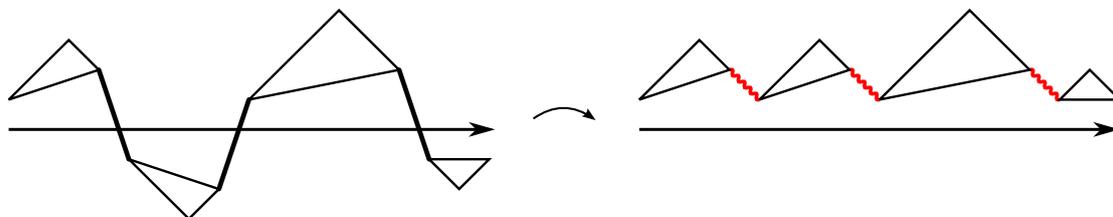
Figure 2: Obtaining a positive path

Now we are going to replace the remaining long steps in P. Pick an arbitrary long up-step s in P, say, from height h to height $h + 3$. Let $\hat{s}$ be the first down-step after s where P goes below the height $h + 1.5$ (such an $\hat{s}$ exists, because P ends at height 1), and let Q denote the subpath of P between s and $\hat{s}$. To introduce some ideas and ingredients of our bijection, let us start with a *"first-try algorithm"* for the simplest case (see Figure 3) that will be refined in the general case: If $\hat{s}$ is a short (down-)step, we replace s with an unmarked short up-step and replace $\hat{s}$ with a marked short down-step. Q is a 1-descending path, by which we mean that its final height is one less lower than its initial height and never goes below it. We convert Q into a 1-ascending path, because after doing that P can be left unmodified before s and after $\hat{s}$. (1-ascending paths are defined analogously, their initial height is one less lower than their final height and they never go below it.) The conversion is easy: Find the first down-step c where Q reaches its final height, and change it to a short up-step. Here we used the fact that c is in an odd position of P (since s is in even position), so it must be a short step, by the definition of $\mathcal{E}^*$. We say that c is the *critical step* of Q, and we will keep using this terminology for 1-ascending and 1-descending paths. In this proof, critical steps are always in odd position. (We will not detail this anymore.)

We note that the original P can be reconstructed from the modified one at this point: $\hat{s}$ became the marked down-step, and s became the last up-step before $\hat{s}$ that ends at the same height as $\hat{s}$ does, and c can also be easily found. And the reconstructability still remains true if we make more of these changes.

If s were a long down-step from height $h + 3$ to height h, then the first-try algorithm would search $\hat{s}$ in the direction to the left: Then $\hat{s}$ would be the last up-step before s that jumps over the horizontal line $y = h + 1.5$. The simplest case is the same here: If $\hat{s}$ is a short (up-)step, then s is replaced with an unmarked short down-step, $\hat{s}$ is replaced with
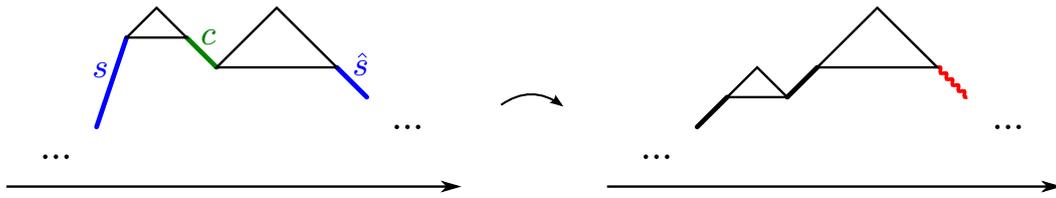
Figure 3: The first-try algorithm

a marked short up-step, and so on, the whole "procedure", including the reconstruction of P, can be reflected across a vertical line. So we will always define our construction for one "direction", and only sketch or skip the analogous one. (An up-step always induces a modification in the right direction, whereas a down-step does it in the left direction.)

However, if there is a long step s for which $\hat{s}$ is a long step, then the first-try algorithm gets stuck, so we improve it now. In the general case, we consider the components of the following directed graph G_P: The edges of G_P are formed by connecting s to $\hat{s}$ for all long steps s of P (so the vertex set of G_P is $\{s, \hat{s} : s$ is a long step in $P\}$). Using the fact that the long steps are in even positions (i.e. their initial height is always even), it is straightforward to check that each (undirected) component of G_P is either a directed path that ends in a short step (component of type 1) or a directed path with one extra edge added from the last vertex to the preceding one (component of type 2), see Figure 4: These paths contain alternating long up-steps and down-steps (except possibly for the last vertex) in a tower-like arrangement in P, i.e. if we take them in their order in P then roughly the first half of them are long up-steps and roughly the last half of them are long down-steps (and there may be a short step in the middle). We note that there are no 1-vertex components in G_P, in some sense those are the x-axis-jumping long steps that have been already replaced in the initial phase.
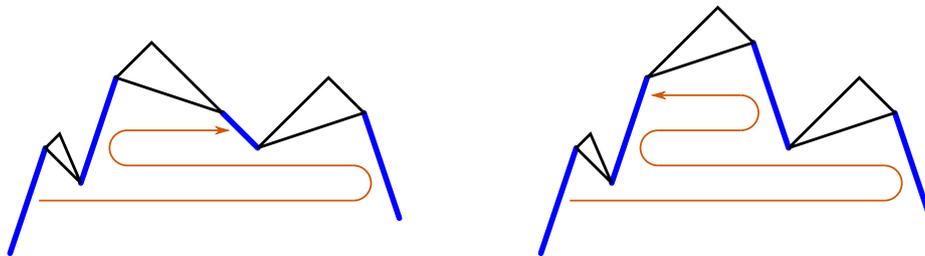
Figure 4: Type 1 and type 2 components of G_P

A component C with vertices $s_1, \ldots, s_k$ cuts P into segments, i.e. P can be viewed as $P = Q_0 s_1 Q_1 s_2 Q_2 \cdots s_k Q_k$, where the segments $Q_1, \ldots, Q_{k-1}$ are 1-ascending or 1-descending paths. We will use the notation $P - \{s_1, \ldots, s_k\}$ for $\{Q_0, \ldots, Q_k\}$. There is one more important property of C beyond its type: The subpath $s_1 Q_1 s_2 Q_2 \cdots Q_{k-1} s_k$ is either 1-ascending or 1-descending, and we will refer to this when using these attributes on components. We note that the parity of k does not really matter here. For $i = 1, \ldots, k-1$,

each Q_i can be divided further into $Q_i = L_i c_i R_i$, where c_i is the critical step of Q_i (i.e. L_i and R_i are non-negative balanced paths). And if C is 1-ascending, then Q_k can be also written in the form $L_k c_k R_k$ where c_k is a "critical" down-step, by which we mean here that L_k is a non-negative balanced path. We say that $(s_1, c_1, \ldots, s_k, c_k)$ is a *block* in P, associated with C. (In the 1-descending case, Q_0 is divided further with a "critical" up-step c_0, instead of Q_k.) See the upper part of Figure 6 for a visualization of blocks. The blocks are nested in the following way: If we consider two arbitrary blocks B_1 and B_2, then they are disjoint and either B_1 is completely contained in one segment of $P - B_2$, or B_2 is completely contained in one segment of $P - B_1$. Recall that P already contains some marked down-steps $d_1, \ldots, d_m$ as replacements for the x-axis-jumping long steps of P_0. It is also true that every block is completely contained in $P - \{d_1, \ldots, d_m\}$. These structural properties can be verified easily.

We define a conversion procedure χ on the blocks first. For a 1-ascending block $B = \{s_1, c_1, \ldots, s_k, c_k\}$, the step s_1 is replaced with a short up-step, and all the other s_i's are replaced with marked short down-steps; the step c_k is replaced with a (short) down-step, and all the other c_i's are replaced with (short) up-steps; where the type of B is indicated by the s_1-replacement up-step: it is unmarked if B is of type 1, and it is marked if B is of type 2. See Figure 5 for a visualization. Briefly, we replace s_i's, the vertices of a component of G_P, with the pattern $\nearrow \searrow \searrow \cdots \searrow \searrow$, and make the required 1-ascending $\leftrightarrow$ 1-descending conversions. In case of 1-descending B, we would use the pattern $\nearrow \nearrow \cdots \nearrow \nearrow \searrow$. With a slight abuse of notation, we will denote by $\chi(B)$ the set of steps in B, after the conversion applied to them. We note that the path P never changes after s_k, so c_k (or c_0) could be completely dropped from the definition of the blocks, but our blocks can be thought as the generalizations of the usual "trapezoid" building blocks of Dyck-paths (where the pair of an up-step u is the first down-step after u that returns to the initial height of u), and in addition, c_k marks the end of the component (it justifies that s_1 is the first step of the component).
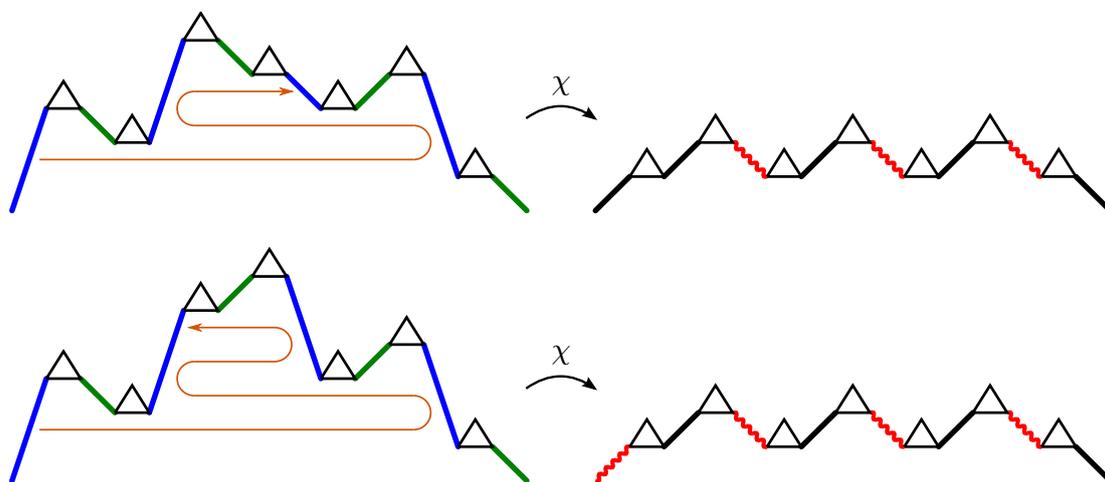
Figure 5: The conversion of blocks

After these preliminaries, we can give the definition of $\phi(P_0)$: It is obtained from P by applying the above conversion χ for all blocks of P, as illustrated in Figure 6. We leave the reader to verify that this definition makes sense: The conversion of a block B does not change the block structure of the actual state of P, i.e. if we recompute blocks in the obtained path, then we find that they are same (except that B is not a block anymore). This follows from the nested property of blocks and the fact that the modified parts (changed or vertically translated steps) of the path do not go below the minimum height of B. It is clear that $\phi(P_0) \in \mathcal{D}_2$, because the 2-labeled steps are placed in even positions and it is obvious that the other required properties also hold. ϕ satisfies the required cardinality condition because, for each block B, the number of marked steps in $\chi(B)$ is equal to the number of long steps in B.
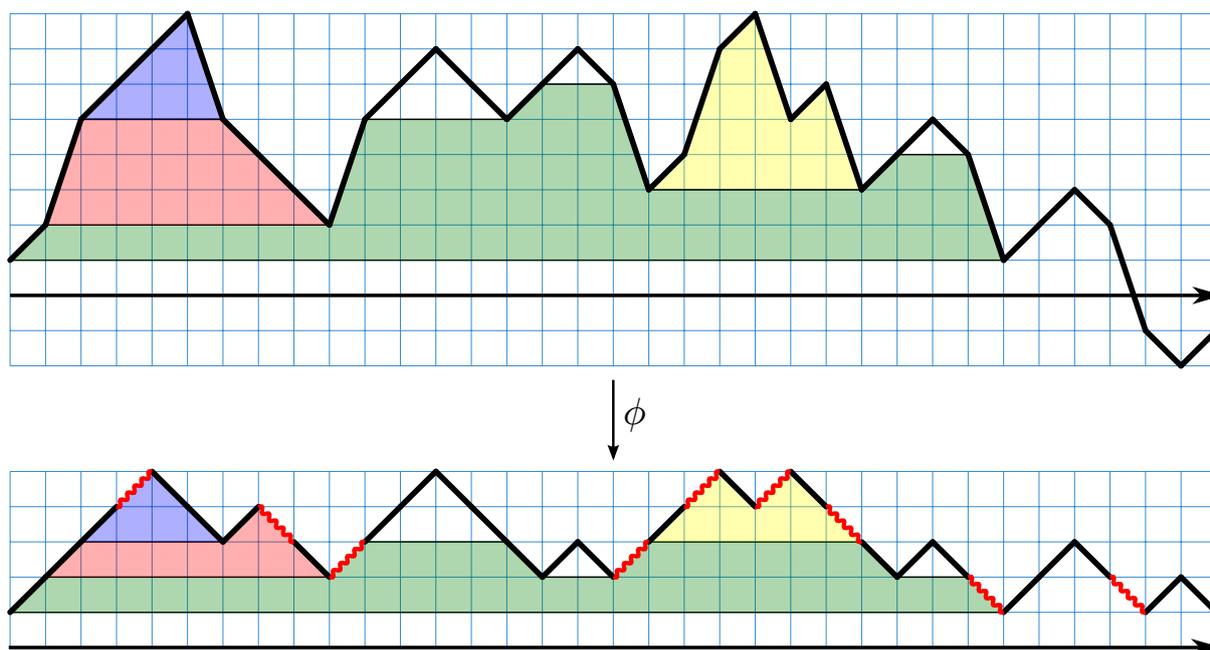
Figure 6: Illustration of ϕ

Now we sketch why ϕ is a bijection. Pick an arbitrary 2-labeled Dyck-path $D \in \mathcal{D}_2$ and try to find an inverse image P. After a translation of 1 unit up, we assume that D starts from the point $(0, 1)$. The components of G_P are encoded with the patterns $\nearrow \searrow \searrow \cdots \searrow \searrow$ or $\nearrow \nearrow \cdots \nearrow \nearrow \searrow$ in $\phi(P)$, with 1-ascending or 1-descending paths between these steps, and all steps on the "upper level" are marked. Since all the information is contained in a pattern to reconstruct the corresponding component of G_P (and block of P), we only have to identify these patterns correctly to obtain P (for a marked step of D, we have to figure out whether it belongs to the upper or lower level of a pattern). This can be done, proceeding from up to down: Take the marked steps of D that have the maximal height h; these steps must belong to the upper level of some patterns. For each such steps s, we can find the step $\tilde{s}$ in D that belongs to the lower level of the pattern containing s. (If s

is an up-step, then $\tilde{s}$ is the first down-step after s that goes below the initial height of s.) In this way we can identify all patterns in D whose upper level has height h. From these patterns, we reconstruct the steps of the corresponding blocks of (the hypothetical) P. Then we repeat the process for the new maximal height h' of marked steps ($h' < h$), and iterate it until we reach the lowest possible height for marked steps. The remaining steps are down-steps from height 2 to height 1 that encode the x-axis-jumping long steps of P, so we replace them with x-axis-jumping long steps and make the required reflections across the x-axis. It is not hard to see that this procedure finds the unique path $P \in \mathcal{E}^*$ for which $\phi(P) = D$, proving the bijectivity of ϕ. $\qquad\square$

Remark 7. We note that a very similar bijection can be constructed to prove Lemma 7 of [3], which states that the number of all even-zeroed paths of length $4n$ is $4^n \binom{2n}{n}$. We omit the proof here because that statement is weaker, it is just Lemma 5, multiplied by n (that can be seen bijectively).

One may conjecture that the condition on ϕ can be strengthened in Lemma 6 by requiring that the positions of the marked steps in $\phi(P)$ agree with the positions of the long steps in P, but small-sized examples show that this is not possible, because the corresponding cardinalities can differ.

When extending ϕ to a bijection $\psi \colon \mathcal{E}_3^* \to \mathcal{D}_4$ at the end of the proof of Lemma 5, we could copy the labels $1, 2, 3$ from Q to $\phi(Q')$ in a way that follows the proof of Lemma 6: If a labeled step s does not belong to any block of Q', then ϕ does not change this step, so we can keep its label in $\phi(Q')$; if s is in a block B, then its label can be moved to the unique unmarked step of $\chi(B)$ (in even position).

Corollary 8. a) *The number of all paths in $\mathcal{E}^*(n)$ is $2^n C_n$. The number of paths in $\mathcal{E}^*(n)$ with k long steps is $\binom{n}{k} C_n$.*
  b) *Consequently, C_n counts the number of paths in $\mathcal{E}^*(n)$ with n long steps (i.e. with alternating short and long steps).*
  c) *If $n \geqslant 1$, C_n counts the number of such paths of length n starting from the origin and ending at $(n, 1)$ in which every step is either $(1, \pm 1)$ or $(1, \pm 2)$ and that never hit the x-axis after the starting point.*

Proof. Claims a) and b) follow directly from Lemma 6. Claim c) just rephrases b): If we remove the first (up-)step from the paths in b), extend them with a final step onto the x-axis, and replace every long-short pair of steps with their "sum" (which is either $(2, \pm 4)$, if the two directions agree, or $(2, \pm 2)$, if not), we obtain the paths in c), scaled by 2 and reflected across a vertical axis, in a bijective way. $\qquad\square$

3   Further problems

We suspect that there exists a simpler bijective proof for Lemma 4. The following generalization of the problem also suggests this.

For a 0-1-2-sequence $b_0, b_1, \ldots, b_n$, let $\mathcal{P}[b_0 b_1 \cdots b_n]$ denote the set of such (conventional) paths of length $2n$ that start from the origin and avoid the points $\{(2i, 0) : b_i = 0\}$

on the x-axis but visit at least one point of $\{(2i, 0) : b_i = 2\}$. In this notation, the set of even-zeroed paths of length $4n$ is denoted by $\mathcal{P}[10101 \cdots 101]$ or $\mathcal{P}[(10)^n 1]$, using an abbreviation for the repetitive part. Similarly, Lemma 4 counts (the half of) the paths in $\mathcal{P}[(10)^n 12]$.

The following conjecture generalizes Lemma 7 of [3] and Lemma 4 of this paper, using the notation $B_n = \binom{2n}{n}$:

Conjecture 9. a) $\left|\mathcal{P}[(1^k 0^k)^n]\right| = \left|\mathcal{P}[10^n 1^{2kn-n-1}]\right| = 4^{2kn-n-1} B_n$,
  b) $\left|\mathcal{P}[(1^k 0^k)^{n-1} 1^k 2^k]\right| = \left|\mathcal{P}[10^{n-1} 21^{2kn-n-1}]\right| = 4^{2kn-n-1} 2C_{n-1}$.

Note that the exponent $2kn-n-1$ ensures that all paths have the same length $4kn-2$ here. The second equalities are obvious in both cases: Clearly, $\left|\mathcal{P}[10^{n-1}2]\right| = 2C_{n-1}$ and it is also well known that $\left|\mathcal{P}[10^n]\right| = B_n$ (i.e. B_n counts the number of non-zero paths of length $2n$), and adding a suffix 1 to the defining sequence of $\mathcal{P}$ is just extending its paths by two arbitrary steps, i.e. multiplying its cardinality by 4.

We do not know any (combinatorial or non-combinatorial) way to determine the first cardinalities, but a) has been verified by computer up to paths of length 40. And it is not hard to see that the two statements are equivalent: If we subtract the number of "wrong" paths from the number of all paths of the required length, we get

$$\left|\mathcal{P}[(1^k 0^k)^n]\right| = 4^{2kn-1} - \sum_{i=1}^{n} \left|\mathcal{P}[(1^k 0^k)^{i-1} 1^k 2^k 1^{2k(n-i)}]\right|,$$

from which the required equivalence follows (non-bijectively), using the easy fact

$$\sum_{i=1}^{n} 2C_{i-1} 4^{n-i} = 4^n - B_n,$$

where both sides count those paths of length $2n$ that have a non-zero x-intercept (and the left-hand side groups these paths by their first non-zero x-intercept).

Acknowledgements

References

[1] G. E. Andrews. On Shapiro's Catalan convolution. *Adv. in Appl. Math.*, 46:15–24, 2011.

[2] T. Koshy. *Catalan Numbers with Applications.* Oxford University Press, New York, 2009.

[3] G. V. Nagy. A combinatorial proof of Shapiro's Catalan convolution. *Adv. in Appl. Math.*, 49:391–396, 2012.

[4] W. Nichols. A Path Bijection. *Amer. Math. Monthly*, 94:465–466, 1987.

[5] R. P. Stanley. Bijective proof problems. 2009.
http://www-math.mit.edu/~rstan/bij.pdf

[6] R. P. Stanley. *Enumerative Combinatorics, Vol. 2.* Cambridge University Press, Cambridge, 1999.