Math 222 Fall 2022, Lecture 29: Lattice paths

website: `https://www.cip.ifi.lmu.de/~grinberg/t/22fco`

5. Lattice paths and Catalan numbers

In this short chapter, we shall briefly survey the topics of **lattice paths** and of **Catalan numbers**. Both topics have books written about them ([Kratte17] for lattice paths, [Stanle15] or [Roman15] for Catalan numbers); we will only have time for the most elementary results.

5.1. Lattice paths

Lattice paths are paths in the integer lattice. We begin by defining the latter:

Definition 5.1.1. The **integer lattice** (or, for short, **lattice**) is the set $\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z}$ of all pairs of integers.

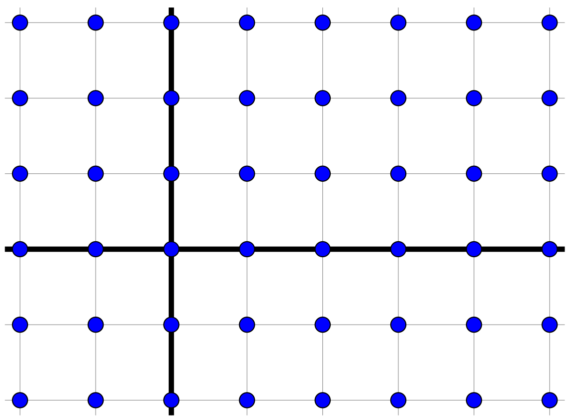
Its elements (i.e., the pairs of integers) are called **points**; indeed, every element $(a, b) \in \mathbb{Z}^2$ can be identified with the point with Cartesian coordinates a and b in the plane.

Points can be added and subtracted entrywise: i.e., we set

$$(a, b) + (c, d) = (a + c, \quad b + d) \qquad \text{and}$$
$$(a, b) - (c, d) = (a - c, \quad b - d)$$

for any $(a, b) \in \mathbb{Z}^2$ and $(c, d) \in \mathbb{Z}^2$.

The integer lattice is countably infinite, but here is a picture of a small piece of it:



(Here, the thick black lines are the x-axis and the y-axis; the other lines are parallels to the axes at integer levels (i.e., lines of the form $x = n$ or $y = n$ for $n \in \mathbb{Z}$), and the blue circles are the points of the lattice.)

Next, we define the notion of lattice paths. There are actually several different notions of lattice paths, but the following one is the simplest (and also the most useful)[1]:

Definition 5.1.2. Let $(a, b) \in \mathbb{Z}^2$ and $(c, d) \in \mathbb{Z}^2$ be two points. Then, a **lattice path** (for short: **LP**) from (a, b) to (c, d) is

- informally understood to be a path from (a, b) to (c, d) in the plane that uses only the following two kinds of steps:
  - "**up-steps**" (denoted "U") that go from a point (p, q) to $(p, q+1)$;
  - "**right-steps**" (denoted "R") that go from a point (p, q) to $(p+1, q)$.
- rigorously defined to be a tuple $(v_0, v_1, \ldots, v_n)$ of points $v_0, v_1, \ldots, v_n \in \mathbb{Z}^2$ such that

$$v_0 = (a, b) \qquad \text{and} \qquad v_n = (c, d) \qquad \text{and}$$
$$v_i - v_{i-1} \in \{(0, 1),\ (1, 0)\} \qquad \text{for each } i \in [n].$$

  If $(v_0, v_1, \ldots, v_n)$ is a LP from (a, b) to (c, d), then the differences $v_i - v_{i-1}$ (for $i \in [n]$) are called the **steps** of this LP. The pair $(0, 1)$ is called an **up-step** and is denoted by U; the pair $(1, 0)$ is called a **right-step** and is denoted by R. If $(v_0, v_1, \ldots, v_n)$ is an LP, then the sequence

$$(v_1 - v_0,\ \ v_2 - v_1,\ \ v_3 - v_2,\ \ \ldots,\ \ v_n - v_{n-1})$$

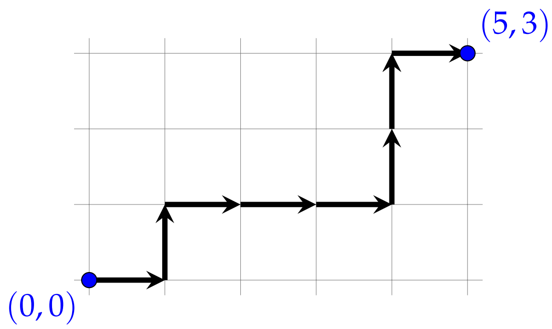
  of all its steps is called the **step sequence** of this LP.

An LP from (a, b) to (c, d) is also called an LP that **starts** at (a, b) and **ends** at (c, d).

Of course, any LP $(v_0, v_1, \ldots, v_n)$ can be drawn by marking the points $v_0, v_1, \ldots, v_n$ on the Cartesian plane and connecting each of them by a line segment to the next one.

[1]Various other notions are discussed in [Kratte17].

Example 5.1.3. The following picture shows a LP from $(0,0)$ to $(5,3)$:



Formally speaking, this LP is the 9-tuple

$$((0,0)\,,\ (1,0)\,,\ (1,1)\,,\ (2,1)\,,\ (3,1)\,,\ (4,1)\,,\ (4,2)\,,\ (4,3)\,,\ (5,3))\,.$$

Its step sequence is (R,U,R,R,R,U,U,R) (meaning that its first step is a right-step, its second step is an up-step, its third step is a right-step, and so on).

We agree to omit the commas and the parentheses when writing down the step sequence of a LP. Thus, the LP from Example 5.1.3 has step sequence $RURRRUUR$.

Note that any LP is uniquely determined by its starting point and its step sequence.

A counting problem immediately suggests itself: How many LPs are there from a given point to a given point? Here is the answer, in two equivalent forms:

Proposition 5.1.4. Let $(a,b) \in \mathbb{Z}^2$ and $(c,d) \in \mathbb{Z}^2$ be two points. Then,

$$\begin{aligned}
&(\text{\# of LPs from } (a,b) \text{ to } (c,d)) \\
&= \begin{cases} \dbinom{c-a+d-b}{c-a}, & \text{if } c \geq a \text{ and } d \geq b; \\ 0, & \text{otherwise} \end{cases} && (1) \\
&= \begin{cases} \dbinom{c-a+d-b}{c-a}, & \text{if } c+d \geq a+b; \\ 0, & \text{otherwise.} \end{cases} && (2)
\end{aligned}$$

Proof sketch. We shall first focus on proving (1).

We are in one of the following three cases:

Case 1: We have $c \geq a$ and $d \geq b$.

Case 2: We have $d < b$.

Case 3: We have $c < a$.

Let us first consider Case 1. In this case, we have $c \geq a$ and $d \geq b$.

If $\mathbf{v}$ is any LP from (a, b) to (c, d), then the step sequence of $\mathbf{v}$ must contain exactly $c - a$ many right-steps[2] and exactly $d - b$ many up-steps[3]. In other words, this step sequence must contain $c - a$ many R's and $d - b$ many U's. Hence, this step sequence must be an anagram[4] of the tuple

$$\left(\underbrace{R, R, \ldots, R}_{c-a \text{ times}}, \underbrace{U, U, \ldots, U}_{d-b \text{ times}}\right).$$

Thus, there is a map

$$\begin{aligned} &\text{from } \{\text{LPs from } (a,b) \text{ to } (c,d)\} \\ &\text{to } \left\{\text{anagrams of } \left(\underbrace{R, R, \ldots, R}_{c-a \text{ times}}, \underbrace{U, U, \ldots, U}_{d-b \text{ times}}\right)\right\} \end{aligned}$$

which sends each LP to its step sequence. This map is easily seen to be injective (since an LP is uniquely determined by its starting point and its step sequence) and surjective (since any sequence of $c - a$ right-steps and $d - b$ up-steps starting at the point (a, b) will lead to (c, d)). Thus, it is a bijection. Hence, the bijection principle yields

$$\begin{aligned} &(\text{\# of LPs from } (a,b) \text{ to } (c,d)) \\ &= \left(\text{\# of anagrams of } \left(\underbrace{R, R, \ldots, R}_{c-a \text{ times}}, \underbrace{U, U, \ldots, U}_{d-b \text{ times}}\right)\right) \\ &= \left(\text{\# of anagrams of } \left(\underbrace{1, 1, \ldots, 1}_{c-a \text{ times}}, \underbrace{2, 2, \ldots, 2}_{d-b \text{ times}}\right)\right) \\ &\qquad\qquad (\text{here, we have relabelled the } R\text{'s and the } U\text{'s as 1's and 2's}) \\ &= \binom{(c-a) + (d-b)}{c-a,\ d-b} \qquad (3) \end{aligned}$$

(by Theorem 2.10.11 in Lecture 23, applied to $n = (c - a) + (d - b)$ and $k = 2$ and $n_1 = c - a$ and $n_2 = d - b$ and $\alpha = \left(\underbrace{1, 1, \ldots, 1}_{c-a \text{ times}}, \underbrace{2, 2, \ldots, 2}_{d-b \text{ times}}\right)$).

[2]Indeed, the x-coordinate of a point increases by 1 when we move it along a right-step, and doesn't change when we move it along an up-step. Thus, in order to get from (a, b) to (c, d), we need to make $c - a$ right-steps (since the x-coordinate has to grow from a to c).

[3]for an analogous reason

[4]See Definition 2.10.9 in Lecture 23 for the concept of "anagram" that we are using here.

However, if k and ℓ are any two nonnegative integers, then $\dbinom{k+\ell}{k,\ell} = \dbinom{k+\ell}{k}$ (by Proposition 2.10.3 in Lecture 22, applied to $n = k+\ell$). Applying this to $k = c-a$ and $\ell = d-b$, we obtain $\dbinom{(c-a)+(d-b)}{c-a,\ d-b} = \dbinom{(c-a)+(d-b)}{c-a} = \dbinom{c-a+d-b}{c-a}$. Hence, we can rewrite (3) as

$$(\text{\# of LPs from } (a,b) \text{ to } (c,d)) = \binom{c-a+d-b}{c-a}.$$

This proves (1) in Case 1 (since $c \geq a$ and $d \geq b$).

Let us now consider Case 2. In this case, we have $d < b$. Thus, $d - b < 0$. Just as in Case 1, we can see that if $\mathbf{v}$ is any LP from (a,b) to (c,d), then the step sequence of $\mathbf{v}$ must contain exactly $d-b$ many up-steps. However, this is actually impossible, since $d - b < 0$. Thus, there is no LP from (a,b) to (c,d). In other words,

$$(\text{\# of LPs from } (a,b) \text{ to } (c,d)) = 0.$$

This proves (1) in Case 2 (since we don't have $c \geq a$ and $d \geq b$).

The proof of (1) in Case 3 is similar (but now we need to use $c - a < 0$ instead of $d - b < 0$).

Thus, we have proved (1) in all three Cases 1, 2 and 3. This completes the proof of (1).

It remains to prove (2). To this purpose, it suffices to show (since (1) has already been proven) that

$$\binom{c-a+d-b}{c-a} = 0 \tag{4}$$

whenever we have $c + d \geq a + b$ but not ($c \geq a$ and $d \geq b$). This is rather easy: Assume that we have $c + d \geq a + b$, but we don't have ($c \geq a$ and $d \geq b$). Hence, we have either $c < a$ or $d < b$. If $c < a$, then (4) is obvious (since $c < a$ entails $c - a \notin \mathbb{N}$). If $d < b$, then (4) is easily verified as well (by Proposition 1.3.5 in Lecture 5, since $c - a + d - b \in \mathbb{N}$ and $c - a > c - a + d - b$). Thus, (4) holds in either case, and this completes our proof of Proposition 5.1.4. □

We can ask a slightly subtler counting question:

Definition 5.1.5. Let $\mathbf{v} = (v_0, v_1, \ldots, v_n)$ be a LP from a point (a,b) to a point (c,d). Let $p \in \mathbb{Z}^2$ be a point. We say that $p \in \mathbf{v}$ (in words: p **lies on v**) if $p \in \{v_0, v_1, \ldots, v_n\}$.

For example, if $\mathbf{v}$ is the LP in Example 5.1.3, then $(4,2) \in \mathbf{v}$, but $(3,2)$ does not lie on $\mathbf{v}$.

Exercise 1. Find the # of LPs $\mathbf{v}$ from $(0,0)$ to $(6,6)$ such that $(2,2) \in \mathbf{v}$.

Solution sketch. Each such LP $\mathbf{v}$ consists of a LP from $(0,0)$ to $(2,2)$ and a LP from $(2,2)$ to $(6,6)$. Thus, the product rule yields

$$
\begin{aligned}
&(\#\text{ of LPs } \mathbf{v} \text{ from } (0,0) \text{ to } (6,6) \text{ such that } (2,2) \in \mathbf{v}) \\
&= \underbrace{(\#\text{ of LPs from } (0,0) \text{ to } (2,2))}_{\substack{=\binom{2-0+2-0}{2-0} \\ \text{(by Proposition 5.1.4)}}} \cdot \underbrace{(\#\text{ of LPs from } (2,2) \text{ to } (6,6))}_{\substack{=\binom{6-2+6-2}{6-2} \\ \text{(by Proposition 5.1.4)}}} \\
&= \binom{2-0+2-0}{2-0} \cdot \binom{6-2+6-2}{6-2} = \binom{4}{2} \cdot \binom{8}{4} = 420.
\end{aligned}
$$

□

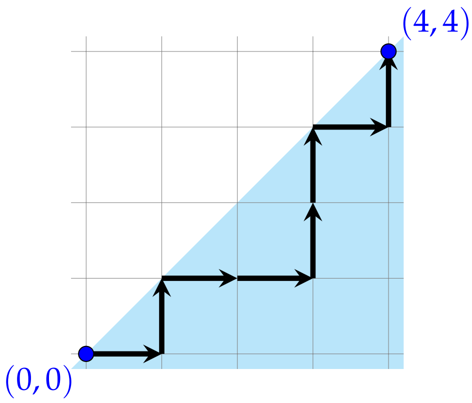
5.2. Catalan paths

More interesting counting problems concern lattice paths whose points are restricted by some requirements:

Definition 5.2.1. A LP $\mathbf{v}$ is said to be **Catalan** if each $(x,y) \in \mathbf{v}$ satisfies $x \geq y$.

In other words, a LP $\mathbf{v}$ is Catalan if and only if it lies entirely in the halfplane below the diagonal given by the equation $x = y$ (including the diagonal itself). In yet other words, a LP $\mathbf{v}$ is Catalan if it never strays above this diagonal.[5]

Example 5.2.2. The following picture shows a Catalan LP from $(0,0)$ to $(4,4)$:

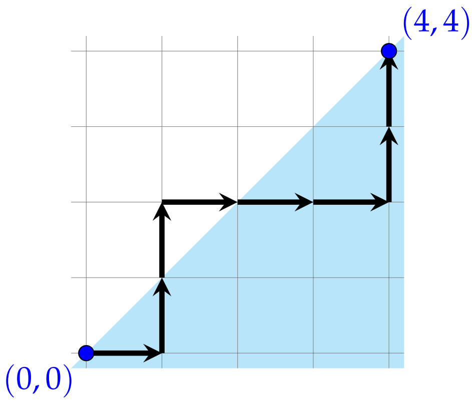


.

[5]The name “Catalan” refers not to the province of Catalonia, nor to its language, but to the Belgian mathematician Eugène Charles Catalan, who did not actually study LPs but rather counted an equivalent class of objects (the “legal words of parentheses” briefly discussed below) in 1838.

(Here, the region shaded cyan is the set of all $(x, y) \in \mathbb{R}^2$ satisfying $x \geq y$; this is the region in which a Catalan LP must stay.)

The following LP from $(0, 0)$ to $(4, 4)$ is **not** Catalan:



. (5)

For another example, the LP $\mathbf{v}$ in Example 5.1.3 is Catalan, whereas the LP with starting point $(0, 0)$ and step sequence $RUUR$ is not (since it contains the point $(1, 2)$).

Definition 5.2.3. If $n, m \in \mathbb{Z}$, then we set

$$L_{n,m} := (\text{\# of Catalan LPs from } (0,0) \text{ to } (n,m)).$$

Here is a table of some of the values of $L_{n,m}$:

$L_{n,m}$	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
$n = 0$	1	0	0	0	0	0
$n = 1$	1	1	0	0	0	0
$n = 2$	1	2	2	0	0	0
$n = 3$	1	3	5	5	0	0
$n = 4$	1	4	9	14	14	0
$n = 5$	1	5	14	28	42	42

.

How can $L_{n,m}$ be computed? Several answers are given in the following theorem:

Theorem 5.2.4. (a) We have $L_{n,m} = L_{n-1,m} + L_{n,m-1}$ for any $n \in \mathbb{Z}$ and $m \in \mathbb{Z}$ satisfying $n \geq m$ and $(n, m) \neq (0, 0)$.

(b) If $n \in \mathbb{Z}$ and $m \in \mathbb{Z}$ satisfy $n < m$, then $L_{n,m} = 0$.

(c) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq m-1$, then

$$L_{n,m} = \binom{n+m}{m} - \binom{n+m}{m-1}.$$

(d) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq m-1$, then

$$L_{n,m} = \frac{n+1-m}{n+1}\binom{n+m}{m}.$$

(e) If $n \in \mathbb{N}$, then

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

A straightforward inductive proof of Theorem 5.2.4 can be found in [18s-mt2s, Exercise 2] (where Catalan LPs are called "legal LPs").[6] Let us, however, prove a more general version of Theorem 5.2.4 **(c)**, counting the Catalan LPs between any two points (a,b) and (c,d):

Theorem 5.2.5. Let $a,b,c,d \in \mathbb{Z}$ satisfy $a \geq b$ and $c \geq d-1$ and $c+d \geq a+b$. Then,

$$\begin{aligned}&(\text{\# of Catalan LPs from } (a,b) \text{ to } (c,d)) \\ &= \binom{c-a+d-b}{c-a} - \binom{c-a+d-b}{d-a-1}.\end{aligned}$$

Note that if any of the conditions "$a \geq b$ and $c \geq d-1$ and $c+d \geq a+b$" is violated, then we simply have (# of Catalan LPs from (a,b) to (c,d)) $= 0$, since no Catalan LPs from (a,b) to (c,d) exist (check this!). Theorem 5.2.5 is [Kratte17, Theorem 10.3.1], and the below proof is similar to the one given in [Kratte17].

Proof of Theorem 5.2.5 (sketched). The difference rule yields

$$\begin{aligned}&(\text{\# of Catalan LPs from } (a,b) \text{ to } (c,d)) \\ &= (\text{\# of LPs from } (a,b) \text{ to } (c,d)) \\ &\qquad - (\text{\# of non-Catalan LPs from } (a,b) \text{ to } (c,d))\end{aligned}$$

(where "non-Catalan" means "not Catalan"). Since (2) yields that

$$(\text{\# of LPs from } (a,b) \text{ to } (c,d)) = \binom{c-a+d-b}{c-a},$$

[6] All parts of Theorem 5.2.4 except for part **(b)** are proved there. But part **(b)** is obvious, since a point (n,m) with $n < m$ cannot lie on a Catalan LP by definition.

it remains only to show that

$$(\text{\# of non-Catalan LPs from } (a,b) \text{ to } (c,d)) \overset{?}{=} \binom{c-a+d-b}{d-a-1}. \tag{6}$$

The very form of this equality suggests that we should be looking for a bijection. Specifically, a bijection

$$\begin{aligned} &\text{from } \{\text{non-Catalan LPs from } (a,b) \text{ to } (c,d)\} \\ &\text{to } \{\text{all LPs from } (a,b) \text{ to } (d-1,c+1)\} \end{aligned}$$

would yield (6), since (2) easily yields that

$$(\text{\# of all LPs from } (a,b) \text{ to } (d-1,c+1)) = \binom{c-a+d-b}{d-a-1}.$$

Such a bijection indeed exists, and its construction is rather peculiar:

The **transpose** of a pair (i,j) shall mean the pair (j,i). Thus, the transpose of an up-step is a right-step, and vice versa.

The entries $v_0, v_1, \ldots, v_k$ of a LP $(v_0, v_1, \ldots, v_k)$ shall be called its **vertices**. A point (x,y) will be called **illegal** if it satisfies $x < y$. Thus, a LP is non-Catalan if and only if it has an illegal vertex. (For instance, the LP shown in (5) has the illegal vertex $(1,2)$.)

Let $\mathbf{v} = (v_0, v_1, \ldots, v_k)$ be a non-Catalan LP starting at (a,b). Let $v_i = (x_i, y_i)$ be its **first** illegal vertex. Hence, we must have $x_i = y_i - 1$, since otherwise the previous vertex v_{i-1} of $\mathbf{v}$ would also be illegal[7]. In other words, v_i is "just

[7] Let us explain this argument in some more detail. First, we observe that v_i is illegal, but v_0 is not (since $v_0 = (a,b)$ satisfies $a \geq b$). Thus, $v_i \neq v_0$ and therefore $i \neq 0$. Hence, v_{i-1} is well-defined. Furthermore, v_{i-1} cannot be illegal (since v_i is the **first** illegal vertex of $\mathbf{v}$, but v_{i-1} comes before v_i).

Write v_{i-1} in the form $v_{i-1} = (x_{i-1}, y_{i-1})$. Then, $x_{i-1} \geq y_{i-1}$ (since v_{i-1} is not illegal). Note also that $x_i < y_i$ (since $(x_i, y_i) = v_i$ is illegal). Hence, $x_i \leq y_i - 1$ (since x_i and y_i are integers).

From $v_i = (x_i, y_i)$ and $v_{i-1} = (x_{i-1}, y_{i-1})$, we obtain $v_i - v_{i-1} = (x_i - x_{i-1},\ y_i - y_{i-1})$.

However, $v_i - v_{i-1}$ is either an up-step or a right-step (since v_i follows v_{i-1} on the LP $\mathbf{v}$). In other words, $v_i - v_{i-1}$ is either $(0,1)$ or $(1,0)$. In view of $v_i - v_{i-1} = (x_i - x_{i-1},\ y_i - y_{i-1})$, we can rewrite this as follows:

$$(x_i - x_{i-1},\ y_i - y_{i-1}) \text{ is either } (0,1) \text{ or } (1,0).$$

Hence, in particular, $x_i - x_{i-1}$ is either 0 or 1, whereas $y_i - y_{i-1}$ is either 1 or 0. Thus, in particular, $x_i - x_{i-1} \geq 0$ and $y_i - y_{i-1} \leq 1$. Now,

$$\begin{aligned} y_i - 1 &\leq y_{i-1} && (\text{since } y_i - y_{i-1} \leq 1) \\ &\leq x_{i-1} && (\text{since } x_{i-1} \geq y_{i-1}) \\ &\leq x_i && (\text{since } x_i - x_{i-1} \geq 0). \end{aligned}$$

Combining this with $x_i \leq y_i - 1$, we obtain $x_i = y_i - 1$, qed.

above" the diagonal line with equation $x = y$. From $x_i = y_i - 1$, we obtain $x_i - y_i = -1$.

Note that the LP $\mathbf{v}$ takes i steps to get from the point $v_0 = (a, b)$ to the point $v_i = (x_i, y_i)$. In its remaining $k - i$ steps, it then proceeds from v_i on to v_k.

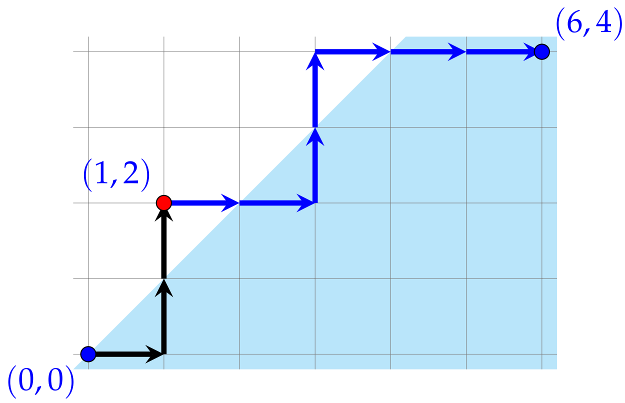
We now define a new LP $\mathbf{v}'$, which again starts at (a, b) and which again has k steps, as follows:

- Its first i steps are precisely the first i steps of $\mathbf{v}$. (These i steps take it to the point $v_i = (x_i, y_i)$.)
- Its remaining $k - i$ steps are the transposes of the last $k - i$ steps of $\mathbf{v}$.

In other words, the path $\mathbf{v}'$ makes the same first i steps as $\mathbf{v}$, but afterwards it makes a right-step whenever $\mathbf{v}$ makes an up-step and vice versa.[8]

Let us illustrate the definition of $\mathbf{v}'$ on an example:

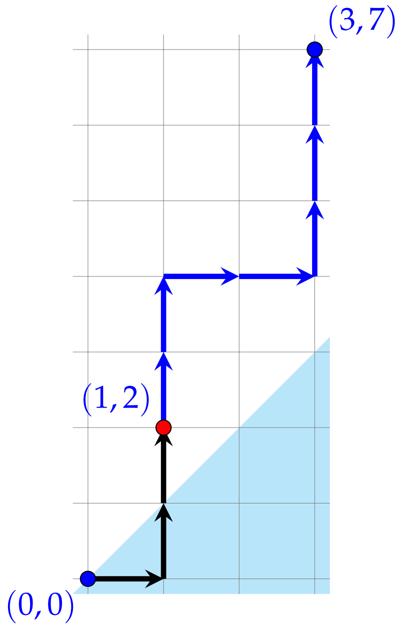
Example 5.2.6. Let $\mathbf{v}$ be the non-Catalan LP from $(0, 0)$ to $(6, 4)$ shown in the following picture:



(we will soon explain why some steps of $\mathbf{v}$ have been painted blue). Then, the first illegal vertex of $\mathbf{v}$ is the point $(1, 2)$ (circled in red in the above picture). The LP $\mathbf{v}$ makes 3 steps before reaching this point (these are the black steps), and subsequently makes 7 further steps (these are the blue steps). To construct $\mathbf{v}'$, we thus need to leave the first 3 steps of $\mathbf{v}$ (that is, the black steps) unchanged but replace the remaining 7 steps (that is, the blue steps)

[8] Geometrically, this can be viewed as reflecting the part of $\mathbf{v}$ that comes after the point v_i across the line through v_i with slope 1. For this reason, what we are doing right now is sometimes called the "reflection trick" or the "reflection principle".

by their transposes. Thus, $\mathbf{v}'$ is the following LP:



.

This is an LP from $(0,0)$ to $(3,7)$.

Back to the general case. We observe some properties of $\mathbf{v}'$:

1. If the LP $\mathbf{v}$ ends at (c,d), then the LP $\mathbf{v}'$ ends at $(d-1, c+1)$.

   Proof: Assume that the LP $\mathbf{v}$ ends at (c,d).

   Before reaching the point v_i, the LP $\mathbf{v}$ makes $x_i - a$ right-steps (since it goes from (a,b) to $v_i = (x_i, y_i)$) and $y_i - b$ up-steps (for the same reason). The LP $\mathbf{v}'$ contains all of these steps (by its definition).

   After reaching the point v_i, the LP $\mathbf{v}$ makes $c - x_i$ right-steps (since it goes from $v_i = (x_i, y_i)$ to (c,d)) and $d - y_i$ up-steps (for the same reason). The LP $\mathbf{v}'$ contains the transposes of all these steps (by its definition), so it contains $c - x_i$ up-steps and $d - y_i$ right-steps after reaching the point v_i.

   Altogether, the LP $\mathbf{v}'$ therefore contains $(x_i - a) + (d - y_i)$ right-steps and $(y_i - b) + (c - x_i)$ up-steps. Since it starts at (a,b), it therefore ends at

$$\left(\underbrace{a + (x_i - a) + (d - y_i)}_{\substack{=d+x_i-y_i \\ =d+(x_i-y_i)}}, \ \underbrace{b + (y_i - b) + (c - x_i)}_{\substack{=c-x_i+y_i \\ =c-(x_i-y_i)}} \right)$$
$$= \left(d + \underbrace{(x_i - y_i)}_{=-1}, \ c - \underbrace{(x_i - y_i)}_{=-1} \right) = (d + (-1), \ c - (-1)) = (d-1, c+1).$$

2. If the LP $\mathbf{v}$ ends at $(d-1, c+1)$, then the LP $\mathbf{v}'$ ends at (c, d).

   Proof: This is proved just like the previous observation, except that c and d are replaced by $d-1$ and $c+1$.

3. The LP $\mathbf{v}'$ is non-Catalan, and its first illegal vertex is precisely the first illegal vertex of $\mathbf{v}$ (that is, v_i).

   Proof: The first i steps of $\mathbf{v}'$ have been copied over from $\mathbf{v}$ unchanged, so it is clear that v_i still lies on $\mathbf{v}'$ and is furthermore the first illegal vertex of $\mathbf{v}'$.

Now, forget that we fixed $\mathbf{v}$. Thus, for each non-Catalan LP $\mathbf{v}$ starting at (a, b), we have constructed a non-Catalan LP $\mathbf{v}'$ starting at (a, b). Moreover, if the former LP $\mathbf{v}$ ends at (c, d), then the latter LP $\mathbf{v}'$ ends at $(d-1, c+1)$. This gives us a map

$$\begin{aligned} &\text{from } \{\text{non-Catalan LPs from } (a,b) \text{ to } (c,d)\} \\ &\text{to } \{\text{non-Catalan LPs from } (a,b) \text{ to } (d-1,c+1)\}, \end{aligned}$$

which sends each $\mathbf{v}$ to $\mathbf{v}'$.

I claim that this map is a bijection. Indeed, its inverse is given by the same rule: It, too, sends each $\mathbf{v}$ to $\mathbf{v}'$. Why? Consider any non-Catalan LP $\mathbf{v} = (v_0, v_1, \ldots, v_k)$ starting at (a, b). As we have seen, the first illegal vertex of $\mathbf{v}'$ is precisely the first illegal vertex of $\mathbf{v}$. Let v_i be this first illegal vertex. As we recall, $\mathbf{v}'$ is obtained from $\mathbf{v}$ by replacing the last $k-i$ steps by their transposes (while the first i steps stay unchanged). Since v_i is the first illegal vertex of $\mathbf{v}'$ as well, we see likewise that $\mathbf{v}''$ is obtained from $\mathbf{v}'$ by replacing the last $k-i$ steps by their transposes (while the first i steps stay unchanged). Thus, as we pass from $\mathbf{v}$ to $\mathbf{v}'$ and then further to $\mathbf{v}''$, the last $k-i$ steps get replaced by their transposes **twice** in succession. Since the transpose of the transpose of a step is the same step, this twofold replacement ends up returning them to their original values. Hence, $\mathbf{v}''$ has the same steps as $\mathbf{v}$. Therefore, $\mathbf{v}'' = \mathbf{v}$ (since $\mathbf{v}''$ and $\mathbf{v}$ both start at (a, b)).

Forget that we fixed $\mathbf{v}$. We thus have shown that $\mathbf{v}'' = \mathbf{v}$ for each non-Catalan LP $\mathbf{v} = (v_0, v_1, \ldots, v_k)$ starting at (a, b). Hence, the map

$$\begin{aligned} &\text{from } \{\text{non-Catalan LPs from } (a,b) \text{ to } (c,d)\} \\ &\text{to } \{\text{non-Catalan LPs from } (a,b) \text{ to } (d-1,c+1)\} \end{aligned}$$

that sends each $\mathbf{v}$ to $\mathbf{v}'$ has an inverse, and this inverse is given by the same rule (viz., sending $\mathbf{v}$ to $\mathbf{v}'$). Therefore, this map is a bijection. By the bijection principle, we thus obtain

$$\begin{aligned} &(\#\text{ of non-Catalan LPs from } (a,b) \text{ to } (c,d)) \\ &= (\#\text{ of non-Catalan LPs from } (a,b) \text{ to } (d-1,c+1)). \end{aligned}$$

However, **all** LPs from (a,b) to $(d-1,c+1)$ are non-Catalan, since their ending point $(d-1,c+1)$ does not satisfy $d-1 \geq c+1$ (because $c+1 > c \geq d-1$) and therefore cannot lie on any Catalan LP. Thus,

$$\begin{aligned}
&(\text{\# of non-Catalan LPs from } (a,b) \text{ to } (d-1,c+1)) \\
&= (\text{\# of LPs from } (a,b) \text{ to } (d-1,c+1)) \\
&= \binom{(d-1)-a+(c+1)-b}{(d-1)-a} \\
&\qquad\qquad \left(\begin{array}{c}\text{by (2) (applied to } d-1 \text{ and } c+1 \text{ instead of } c \text{ and } d\text{),} \\ \text{since } (d-1)+(c+1)=c+d \geq a+b\end{array}\right) \\
&= \binom{c-a+d-b}{d-a-1} \qquad (\text{by straightforward simplifications}).
\end{aligned}$$

Hence,

$$\begin{aligned}
&(\text{\# of non-Catalan LPs from } (a,b) \text{ to } (c,d)) \\
&= (\text{\# of non-Catalan LPs from } (a,b) \text{ to } (d-1,c+1)) = \binom{c-a+d-b}{d-a-1}.
\end{aligned}$$

This proves (6), thus completing our proof of Theorem 5.2.5. $\square$

From Theorem 5.2.5, we can easily derive parts **(c)**, **(d)** and **(e)** of Theorem 5.2.4:

*Proof of Theorem 5.2.4 parts **(c)**, **(d)** and **(e)** (sketched).* **(c)**. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq m-1$. Then, the definition of $L_{n,m}$ yields

$$\begin{aligned}
L_{n,m} &= (\text{\# of Catalan LPs from } (0,0) \text{ to } (n,m)) \\
&= \binom{n-0+m-0}{n-0} - \binom{n-0+m-0}{m-0-1} \\
&\qquad \left(\begin{array}{c}\text{by Theorem 5.2.5,} \\ \text{applied to } a=0,\ b=0,\ c=n \text{ and } d=m\end{array}\right) \\
&= \underbrace{\binom{n+m}{n}}_{\substack{=\binom{n+m}{(n+m)-n} \\ \text{(by the symmetry of BCs} \\ \text{(Lecture 6, Theorem 1.3.9))}}} - \binom{n+m}{m-1} = \binom{n+m}{(n+m)-n} - \binom{n+m}{m-1} \\
&= \binom{n+m}{m} - \binom{n+m}{m-1}.
\end{aligned}$$

This proves Theorem 5.2.4 **(c)**.

(d) Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq m - 1$. It is easy to see that

$$\binom{n+m}{m-1} = \frac{m}{n+1} \binom{n+m}{m}. \qquad (7)$$

(Indeed, this follows easily from the definition of binomial coefficients[9].) Now, Theorem 5.2.4 **(c)** yields

$$\begin{aligned} L_{n,m} &= \binom{n+m}{m} - \binom{n+m}{m-1} \\ &= \binom{n+m}{m} - \frac{m}{n+1}\binom{n+m}{m} \qquad \text{(by (7))} \\ &= \frac{n+1-m}{n+1}\binom{n+m}{m}. \end{aligned}$$

Thus, Theorem 5.2.4 **(d)** follows.

(e) Let $n \in \mathbb{N}$. Applying Theorem 5.2.4 **(d)** to $m = n$, we obtain

$$L_{n,n} = \frac{n+1-n}{n+1}\binom{n+n}{n} = \frac{1}{n+1}\binom{2n}{n}.$$

Thus Theorem 5.2.4 **(e)**. □

5.3. Catalan numbers

Using parts **(d)** and **(e)** of Theorem 5.2.4, we can make the following definition:

Definition 5.3.1. For any $n \in \mathbb{N}$, the number

$$L_{n,n} = \frac{1}{n+1}\binom{2n}{n} = \binom{2n}{n} - \binom{2n}{n-1}$$

is called the ***n*-th Catalan number**, and is denoted by C_n.

At least two books ([Roman15] and [Stanle15]) have been written about these numbers, and their history involves some of the most famous mathematicians of the last 4 centuries (including Euler, Goldbach and Cayley; see [Pak15] for a survey[10]). By now, they have found uses all over mathematics. We will soon

[9] The case $m = 0$ needs to be considered separately, but this case is trivial.

[10] Drew Armstrong's talk slides `https://www.math.miami.edu/~armstrong/Talks/Story_of_Catalan.pdf` give a quick overview.

Yes, Euler's work predates Catalan's by almost 90 years. But you wouldn't want to call them "Euler numbers", would you?

see some of their appearances[11]. First, let us give a table of the first 10 Catalan numbers:

n	0	1	2	3	4	5	6	7	8	9	$\cdots$
C_n	1	1	2	5	14	42	132	429	1430	4862	$\cdots$

.

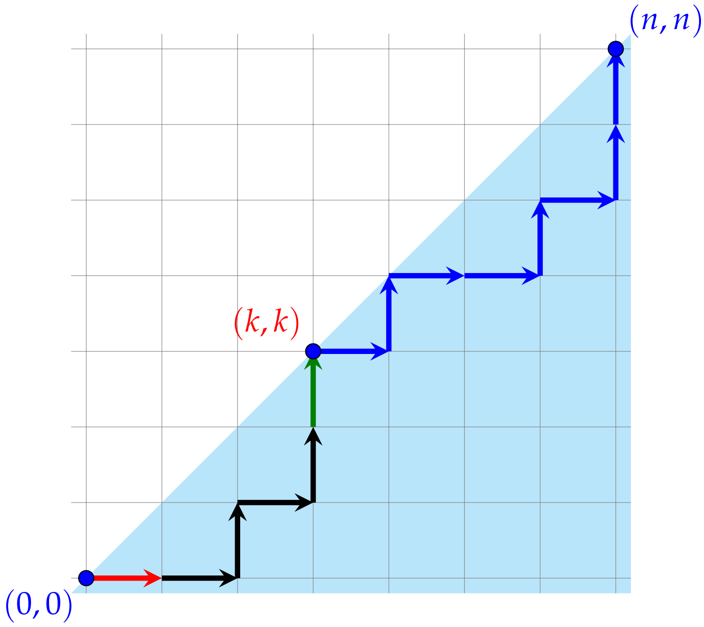
Apart from their explicit definition, the Catalan numbers satisfy a recursion:

Theorem 5.3.2 (Segner's recurrence)**.** For any integer $n > 0$, we have

$$C_n = \sum_{k=1}^{n} C_{k-1}C_{n-k}.$$

Proof sketch. Let $n > 0$ be an integer. Consider the Catalan LPs from $(0,0)$ to (n,n). If $\mathbf{v}$ is such an LP, then we define the **first return** of $\mathbf{v}$ to be the smallest $k \in [n]$ satisfying $(k,k) \in \mathbf{v}$. (This is always well-defined, since $(n,n) \in \mathbf{v}$ by definition. Note that k cannot be 0, since k should be positive. The word "first return" is short for "first return to the diagonal $x = y$".)

For example, if $n = 7$, then the Catalan LP shown in



(8)

has first return 3 (since it contains the point $(3,3)$ but neither $(1,1)$ nor $(2,2)$). (The colors of the steps will be explained further below.)

Now, we claim the following:

[11]See also Richard Stanley's talk slides `https://math.mit.edu/~rstan/transparencies/china.pdf` .

Claim 1: For each $k \in [n]$, we have

$$(\text{\# of Catalan LPs from } (0,0) \text{ to } (n,n) \text{ with first return } k) \\ = C_{k-1}C_{n-k}.$$

Once Claim 1 is proved, Theorem 5.3.2 will easily follow:

$$\begin{aligned}
C_n &= L_{n,n} \qquad \text{(by the definition of } C_n\text{)} \\
&= (\text{\# of Catalan LPs from } (0,0) \text{ to } (n,n)) \qquad \text{(by the definition of } L_{n,n}\text{)} \\
&= \sum_{k=1}^{n} \underbrace{(\text{\# of Catalan LPs from } (0,0) \text{ to } (n,n) \text{ with first return } k)}_{\substack{=C_{k-1}C_{n-k} \\ \text{(by Claim 1)}}} \\
&\qquad \text{(by the sum rule)} \\
&= \sum_{k=1}^{n} C_{k-1}C_{n-k}.
\end{aligned}$$

Thus, it remains to prove Claim 1.

Proof of Claim 1: Let $\mathbf{v}$ be a Catalan LP from $(0,0)$ to (n,n) with first return k. What can we say about it?

1. No point of the form (i,i) with $0 < i < k$ can lie on $\mathbf{v}$ (since k is the **first** return of $\mathbf{v}$).

2. The first step of $\mathbf{v}$ must be a right-step, leading to $(1,0)$. (Indeed, the only other option would be an up-step leading to $(0,1)$, but then $\mathbf{v}$ would not be Catalan, since $0 < 1$.) In the example (8), this right-step has been painted red.

3. The LP $\mathbf{v}$ must contain (k,k) (since k is its first return), and thus must reach (k,k) after $2k$ steps (namely, k up-steps and k right-steps).

4. The LP $\mathbf{v}$ must arrive at (k,k) either by an up-step or by a right-step. However, if $\mathbf{v}$ would arrive at (k,k) by a right-step, then the point on $\mathbf{v}$ preceding (k,k) would be $(k-1,k)$, which is impossible (since $\mathbf{v}$ is Catalan and thus cannot contain $(k-1,k)$). Hence, $\mathbf{v}$ must arrive at (k,k) by an up-step, coming from the point $(k,k-1)$. This up-step must be the $(2k)$-th step of $\mathbf{v}$ (since $\mathbf{v}$ reaches (k,k) after $2k$ steps). In the example (8), this up-step has been painted green.

5. So we know that the first step of $\mathbf{v}$ goes from $(0,0)$ to $(1,0)$, whereas the $(2k)$-th step of $\mathbf{v}$ goes from $(k,k-1)$ to (k,k). The $2k-2$ steps between these two steps must therefore form a LP from $(1,0)$ to $(k,k-1)$. Let us denote this LP by $\mathbf{v}'$. (In the example (8), this LP $\mathbf{v}'$ has been painted black.) Not only does this LP $\mathbf{v}'$ never stray above the diagonal with

equation $x = y$ (since it is a piece of the LP $\mathbf{v}$, which is Catalan), but it also never touches this diagonal (i.e., it contains no point of the form (i,i)), since no point of the form (i,i) with $0 < i < k$ can lie on $\mathbf{v}$. As a consequence, all points on this LP $\mathbf{v}'$ have the form (x,y) with $x > y$ (not only $x \geq y$), that is, with $x - 1 \geq y$. Hence, if we shift this LP $\mathbf{v}'$ to the left by 1 (that is, if we replace each point (x,y) on $\mathbf{v}'$ by $(x-1,y)$), then we obtain a Catalan LP from $(0,0)$ to $(k-1,k-1)$. Let us denote this latter Catalan LP by $\mathbf{v}''$.

6. Now, let $\mathbf{v}^*$ be the part of the LP $\mathbf{v}$ after the point (k,k). (In the example (8), this LP $\mathbf{v}^*$ has been painted blue.) This $\mathbf{v}^*$ is a Catalan LP from (k,k) to (n,n). If we shift this LP $\mathbf{v}^*$ to the left by k and downwards by k (that is, if we replace each point (x,y) on $\mathbf{v}^*$ by $(x-k,y-k)$), then we obtain a Catalan LP from $(0,0)$ to $(n-k,n-k)$. We denote this latter LP by $\mathbf{v}^{**}$.

Thus, we have decomposed our LP $\mathbf{v}$ into

- a single right-step that takes us from $(0,0)$ to $(1,0)$ (painted red in (8)); followed by
- an LP $\mathbf{v}'$ from $(1,0)$ to $(k,k-1)$ (painted black in (8)) which, if we shift it to the left by 1, becomes a Catalan LP $\mathbf{v}''$ from $(0,0)$ to $(k-1,k-1)$; followed by
- a single up-step that takes us from $(k,k-1)$ to (k,k) (painted green in (8)); followed by
- an LP $\mathbf{v}^*$ from (k,k) to (n,n) (painted blue in (8)) which, if we shift it to the left by k and downwards by k, becomes a Catalan LP $\mathbf{v}^{**}$ from $(0,0)$ to $(n-k,n-k)$.

This decomposition is defined for every $\mathbf{v}$. Hence, let us forget that we fixed $\mathbf{v}$. We thus obtained a map

$$\begin{aligned}
&\text{from } \{\text{Catalan LPs from } (0,0) \text{ to } (n,n) \text{ with first return } k\} \\
&\text{to } \{\text{Catalan LPs from } (0,0) \text{ to } (k-1,k-1)\} \\
&\qquad\qquad \times \{\text{Catalan LPs from } (0,0) \text{ to } (n-k,n-k)\},
\end{aligned}$$

which sends every Catalan LP $\mathbf{v}$ to the pair $(\mathbf{v}'', \mathbf{v}^{**})$. This map is injective (since we can reconstruct $\mathbf{v}$ from $\mathbf{v}''$ and $\mathbf{v}^{**}$) and surjective (check this!). Thus, it is a bijection. Hence, the bijection principle (together with the product rule) yields

$$\begin{aligned}
&(\text{\# of Catalan LPs from } (0,0) \text{ to } (n,n) \text{ with first return } k) \\
&= \underbrace{(\text{\# of Catalan LPs from } (0,0) \text{ to } (k-1,k-1))}_{\substack{=L_{k-1,k-1} \\ \text{(by the definition of } L_{k-1,k-1})}} \\
&\qquad \cdot \underbrace{(\text{\# of Catalan LPs from } (0,0) \text{ to } (n-k,n-k))}_{\substack{=L_{n-k,n-k} \\ \text{(by the definition of } L_{n-k,n-k})}} \\
&= \underbrace{L_{k-1,k-1}}_{=C_{k-1}} \cdot \underbrace{L_{n-k,n-k}}_{=C_{n-k}} = C_{k-1}C_{n-k}.
\end{aligned}$$

This proves Claim 1, thus completing the proof of Theorem 5.3.2. □

As mentioned above, the Catalan numbers C_n tend to appear in surprisingly many different places in mathematics. Stanley's book [Stanle15] lists 214 different counting problems whose answer is C_k for some k. Here are just a few of them:

- Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be **123-avoiding** if there exist no three elements $i < j < k$ of $[n]$ satisfying $\sigma(i) < \sigma(j) < \sigma(k)$ (that is, if the OLN[12] of σ has no increasing subsequence[13] of length 3).

  The # of all 123-avoiding permutations $\sigma \in S_n$ is the Catalan number C_n. (See [Bona22, Theorem 4.7] for a proof of this fact.)

- Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be **132-avoiding** if there exist no three elements $i < j < k$ of $[n]$ satisfying $\sigma(i) < \sigma(k) < \sigma(j)$. (The name "132-avoiding" is of course referring to the fact three such elements lead to a subsequence of the OLN of σ that "looks like" the triple $(1,3,2)$, in the sense that its first entry is its smallest entry, its second entry is its largest one, and its third entry is its mid-sized one.)

  The # of all 132-avoiding permutations $\sigma \in S_n$ is the Catalan number C_n. (Yes, this is the same answer as for 123-avoiding permutations! No really simple reason for this "coincidence" is known, although bijections between the 123-avoiding permutations and the 132-avoiding permutations have been found. See [Bona22, proof of Lemma 4.4] for such a bijection.)

- Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be **213-avoiding** if there exist no three elements $i < j < k$ of $[n]$ satisfying $\sigma(j) < \sigma(i) < \sigma(k)$. (At this point, the origin of this name should be clear.)

[12] Recall: "OLN" is shorthand for "one-line notation".

[13] Keep in mind that a subsequence doesn't have to be contiguous. For example, $(2,4,5)$ is a subsequence of $(2,1,6,3,4,5)$.

The # of all 213-avoiding permutations $\sigma \in S_n$ is, once again, C_n. (This fact, just like the previous two and the following three, is proved in [Bona22, §4.2].)

- Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be **231-avoiding** if there exist no three elements $i < j < k$ of $[n]$ satisfying $\sigma(k) < \sigma(i) < \sigma(j)$.

  The # of all 231-avoiding permutations $\sigma \in S_n$ is (you guessed it) C_n. (This is not hard to prove[14].)

[14] Here is the idea of the proof:

For any $n \in \mathbb{N}$, let q_n denote the # of all 231-avoiding permutations $\sigma \in S_n$.

Fix $n > 0$. Also, fix $k \in [n]$. Consider some 231-avoiding permutation $\sigma \in S_n$ with $\sigma(k) = n$. Thus, the entry n appears in the k-th position of the OLN of σ. The entries in the first $k-1$ positions of this OLN will thus be called the **pre-n entries** of σ, whereas the entries in the last $n-k$ positions will be called the **post-n entries** of σ. Since σ is 231-avoiding, each of the pre-n entries must be smaller than each of the post-n entries (why?). Hence, the pre-n entries must be $1, 2, \ldots, k-1$ in some order (why?), whereas the post-n entries must be $k, k+1, \ldots, n-1$ in some order (why?). Moreover, the orders must themselves be 231-avoiding. More precisely, the pre-n entries form the OLN of a 231-avoiding permutation of $[k-1]$, whereas the post-n entries (once $k-1$ has been subtracted from them) form the OLN of a 231-avoiding permutation of $[n-k]$. Let us denote these two permutations by σ' and σ''.

Forget that we fixed σ. Thus, for each 231-avoiding permutation $\sigma \in S_n$ with $\sigma(k) = n$, we have constructed a 231-avoiding permutation $\sigma' \in S_{k-1}$ (whose OLN consists of the pre-n entries of σ) and a 231-avoiding permutation $\sigma'' \in S_{n-k}$ (whose OLN consists of the post-n entries of σ, each decreased by $k-1$). This results in a map

$$\begin{aligned} &\text{from } \{\text{231-avoiding permutations } \sigma \in S_n \text{ with } \sigma(k) = n\} \\ &\text{to } \{\text{231-avoiding permutations in } S_{k-1}\} \times \{\text{231-avoiding permutations in } S_{n-k}\}, \end{aligned}$$

which sends each σ to the pair (σ', σ''). This map is easily seen to be a bijection (why? surjectivity needs a bit of thought), and thus the bijection principle (and the product rule) yields

$$\begin{aligned} &(\text{\# of 231-avoiding permutations } \sigma \in S_n \text{ with } \sigma(k) = n) \\ &= \underbrace{(\text{\# of 231-avoiding permutations in } S_{k-1})}_{=q_{k-1}} \cdot \underbrace{(\text{\# of 231-avoiding permutations in } S_{n-k})}_{=q_{n-k}} \\ &= q_{k-1} q_{n-k}. \end{aligned}$$

Now, forget that we fixed k. We have

$$\begin{aligned} q_n &= (\text{\# of 231-avoiding permutations } \sigma \in S_n) \\ &= \sum_{k=1}^{n} \underbrace{(\text{\# of 231-avoiding permutations } \sigma \in S_n \text{ with } \sigma(k) = n)}_{\substack{=q_{k-1}q_{n-k} \\ \text{(as we just proved)}}} \\ &\qquad \left(\begin{array}{c} \text{by the sum rule, since } \sigma^{-1}(n) \text{ is uniquely determined} \\ \text{for each } \sigma \in S_n \end{array}\right) \\ &= \sum_{k=1}^{n} q_{k-1} q_{n-k}. \end{aligned}$$

- Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be **312-avoiding** if there exist no three elements $i < j < k$ of $[n]$ satisfying $\sigma(j) < \sigma(k) < \sigma(i)$.

  The # of all 312-avoiding permutations $\sigma \in S_n$ is C_n (who could have known?).

- Let $n \in \mathbb{N}$. A permutation $\sigma \in S_n$ is said to be **321-avoiding** if there exist no three elements $i < j < k$ of $[n]$ satisfying $\sigma(k) < \sigma(j) < \sigma(i)$.

  The # of all 321-avoiding permutations $\sigma \in S_n$ is C_n. (And this time, you should really not be surprised: There is a simple bijection between the 123-avoiding permutations and the 321-avoiding ones. Do you see it?)

  At this point, it should be mentioned that there are also 1234-avoiding permutations, and 1324-avoiding ones, and 51243-avoiding ones, and so on; more generally, for any fixed permutation $\tau \in S_m$ (with arbitrary m), we can define the notion of a "τ-avoiding permutation". These so-called "pattern avoidance classes" (along with their intersections) are the subject of an active area of research that has a yearly conference devoted to it. An interested reader can find more on the Wikipedia page for "pattern avoidance" and explore Bridget Tenner's Database of Permutation Pattern Avoidance; much more can be learned from books like [Bona22, Chapters 4–5] and [Kitaev11].

  One disappointment, though: In general, the # of all 1234-avoiding permutations does not equal the # of 1324-avoiding ones[15] (and neither of them equals C_n).

- Fix $n \in \mathbb{N}$. Consider "words" consisting of n opening parentheses and n closing parentheses.

  For example, for $n = 5$, one such word is "$))()(()()($".

  Such a word is said to be **legal** if its parentheses can be matched (i.e., we can match each opening parenthesis to a closing parenthesis that lies somewhere to its right). For example, the word "$(()(()))$" is legal, whereas the word "$()))((()$" is not.

  How many legal words with n opening and n closing parentheses are there? The answer is C_n.

  This is not actually surprising if you think about it. Indeed, the step sequence of a Catalan LP from $(0,0)$ to (n,n) becomes a legal word if we replace each R by an opening parenthesis and each U by a closing parenthesis (why?). It is not hard to see that this is a bijection, and therefore the # of legal words equals the # of Catalan LPs from $(0,0)$ to (n,n), which

---

This is exactly the same recurrent equation as Theorem 5.3.2 for the Catalan numbers C_n. Since we furthermore have $q_0 = C_0$, we thus conclude that $q_n = C_n$ holds for each $n \in \mathbb{N}$, qed.

[15] Indeed, the former number is smaller when $n \geq 7$. This is proved in [Bona22, Theorem 4.29].

we know to be C_n. (See `https://math.stackexchange.com/questions/2991347` for details.)

- Consider a sum of 4 numbers:

$$a+b+c+d.$$

How many ways are there to **fully parenthesize** it – i.e., to put parentheses around some of the numbers so that we never have to add more than two numbers at the same time? A bit of experimentation shows that there are 5 ways:

$$((a+b)+c)+d, \qquad (a+(b+c))+d,$$
$$(a+b)+(c+d), \qquad a+((b+c)+d), \qquad a+(b+(c+d)).$$

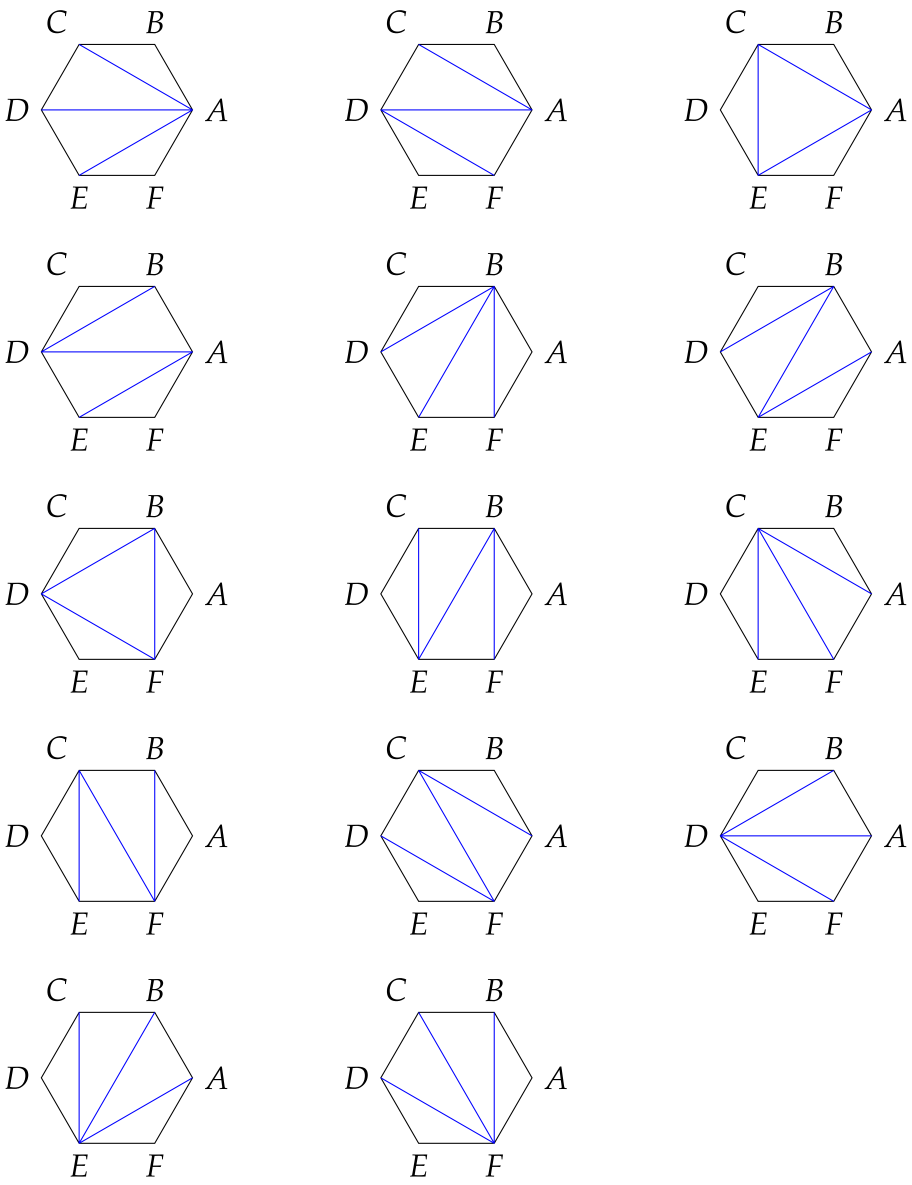
(We are not allowing the numbers to be permuted.) Similarly, for a sum of 3 numbers, there are 2 ways (namely, $(a+b)+c$ and $a+(b+c)$). Similarly, for a sum of 5 numbers, there are 14 ways.

How many ways are there for a sum of n numbers? The answer turns out to be the Catalan number C_{n-1}. (See [Reiner05] or [Singma78] or [Stanle15, Theorem 1.5.1 (v)] for various proofs, up to minor notational differences.)

- Fix an integer $n \geq 3$ and a convex n-gon G_n.

How many ways are there to triangulate G_n – i.e., to subdivide G_n into triangles? Of course, the answer is "infinitely many", since we can keep adding new internal vertices and thus obtain new triangulations each time.

However, we can ask a better question: How many ways are there to triangulate G_n **without** introducing new vertices? For example, for $n = 6$, there are 14 such triangulations:



For a general n, the answer (found by Euler) is C_{n-2}. (See [Stanle15, Theorem 1.5.1 (i)] for a proof.)

- Let $n \in \mathbb{N}$. A **standard Young tableau of shape** (n, n) (this is just a particular case of a much more general concept of "standard Young tableau") is defined to be a $2 \times n$-matrix $\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix}$ such that
  - its entries $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ are the $2n$ integers $1, 2, \ldots, 2n$ in some order;

– its entries strictly increase along each row (i.e., we have $a_1 < a_2 < \cdots < a_n$ and $b_1 < b_2 < \cdots < b_n$);

– its entries strictly increase down each column (i.e., we have $a_i < b_i$ for each $i \in [n]$).

Then, the # of standard Young tableaux[16] of shape (n, n) is the Catalan number C_n. This is due to a fairly simple bijection between these tableaux and the Catalan LPs from $(0, 0)$ to (n, n), which you might enjoy finding[17].

More interpretations of Catalan numbers can be found in [Davis16], [Stanle15] and [Roman15].

5.4. k-Catalan paths

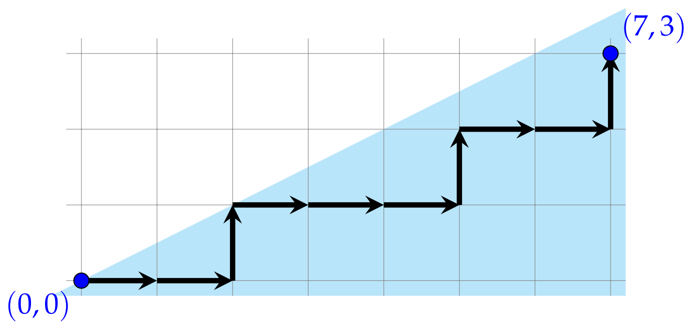
The Catalan LPs are (by definition) the LPs that never stray above the line with equation $x = y$. We know (from Theorem 5.2.5) how to count such LPs between any two points (a, b) and (c, d).

What happens if we replace this line by a different line? One of the simplest options is to replace it by the "slanted diagonal" $x = ky$ for a given $k \in \mathbb{N}$. This leads to the following definition:

Definition 5.4.1. Let $k \in \mathbb{N}$. A LP $\mathbf{v}$ is said to be k-**Catalan** if each $(x, y) \in \mathbf{v}$ satisfies $x \geq ky$.

In other words, a LP $\mathbf{v}$ is k-Catalan if and only if it lies entirely in the halfplane below the straight line given by the equation $x = ky$.

Example 5.4.2. The following picture shows a 2-Catalan LP from $(0, 0)$ to $(7, 3)$:



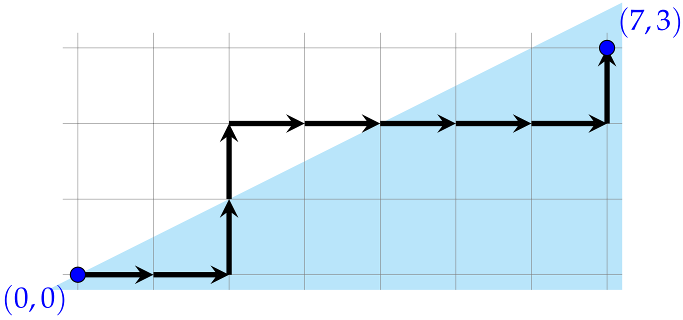
.

(Here, the region shaded cyan is the set of all $(x, y) \in \mathbb{R}^2$ satisfying $x \geq ky$; this is the region in which a k-Catalan LP must stay.)

[16] This is the plural of "tableau" (a loanword from French). That said, you will also find "tableaus" and "tableau" (as well as the even less sensible use of "tableaux" for the singular form) all over the Anglophone literature.

[17] See [Stanle15, Problem 168] for the answer.

The following LP from $(0,0)$ to $(7,3)$ is **not** 2-Catalan:



.

Clearly, a LP is 1-Catalan if and only if it is Catalan. The notion of a k-Catalan LP thus generalizes that of a Catalan LP. The larger k is, the "flatter" the line with equation $x = ky$ is (its slope is $\frac{1}{k}$, after all), and the more restrictive the concept of a k-Catalan LP becomes, at least if the starting point is $(0,0)$.

Encouraged by Theorem 5.2.5, one might try to find a simple formula for the # of k-Catalan LPs from a given point (a,b) to another given point (c,d). Unfortunately, such a formula doesn't seem to exist; the best formulas known involve summation signs (see [Kratte17, Theorem 10.4.7]).

However, all is not lost. The particular case in which $(a,b) = (0,0)$ (that is, counting k-Catalan LPs that start at $(0,0)$) still allows for nice formulas. To state them, we introduce a notation generalizing the $L_{n,m}$ from Definition 5.2.3:

Definition 5.4.3. Let $k \in \mathbb{N}$. If $n, m \in \mathbb{Z}$, then we set

$$L_{n,m,k} := (\# \text{ of } k\text{-Catalan LPs from } (0,0) \text{ to } (n,m)) .$$

Thus, in particular, $L_{n,m} = L_{n,m,1}$ for every $n, m \in \mathbb{Z}$. Also, every LP starting at $(0,0)$ is 0-Catalan, so that we have $L_{n,m,0} = \binom{n+m}{m}$. Now, we can state the formulas for the general case:

Theorem 5.4.4. Let $k \in \mathbb{N}$. Then:

(a) We have $L_{n,m,k} = L_{n-1,m,k} + L_{n,m-1,k}$ for any $n \in \mathbb{Z}$ and $m \in \mathbb{Z}$ satisfying $n \geq km$ and $(n,m) \neq (0,0)$.

(b) If $n \in \mathbb{Z}$ and $m \in \mathbb{Z}$ satisfy $n < km$, then $L_{n,m,k} = 0$.

(c) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq km - 1$, then

$$L_{n,m,k} = \binom{n+m}{m} - k\binom{n+m}{m-1}.$$

(d) If $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy $n \geq km - 1$, then

$$L_{n,m,k} = \frac{n+1-km}{n+1}\binom{n+m}{m}.$$

(e) If $m \in \mathbb{N}$, then

$$L_{km,m,k} = \frac{1}{km+1}\binom{(k+1)\,m}{m}.$$

Proof idea. The straightforward inductive proof of Theorem 5.2.4 in [18s-mt2s, Exercise 2] can be fairly straightforwardly generalized from Catalan LPs to k-Catalan LPs.

Another proof of Theorem 5.4.4 **(d)** can be found in [GouSer03, Theorem 2] (although the statement of [GouSer03, Theorem 2] differs from Theorem 5.4.4 **(d)** in some insubstantial details[18]).

Other authors prove equivalent versions of Theorem 5.4.4 **(d)**:

In [Renaul07], Renault solves a related problem (the so-called "**ballot problem**") in four different ways. This problem asks for the # of LPs $\mathbf{v}$ from $(0,0)$ to (n,m) such that every point $(x,y) \in \mathbf{v}$ other than the starting point $(0,0)$ satisfies the **strict** inequality $x > ky$. We can call such LPs **strictly k-Catalan**. The main result of [Renaul07] is that the # of strictly k-Catalan LPs from $(0,0)$ to (n,m) is $\dfrac{n-km}{n+m}\dbinom{n+m}{m}$. However, it is easy to find a bijection

$$\begin{aligned}&\text{from } \{\text{strictly } k\text{-Catalan LPs from } (0,0) \text{ to } (n+1,m)\}\\ &\text{to } \{k\text{-Catalan LPs from } (0,0) \text{ to } (n,m)\}\end{aligned}$$

(indeed, this bijection removes the first step of a strictly k-Catalan LP, and shifts the rest of the LP by 1 to the left). Thus, the main result of [Renaul07] (applied to $n+1$ instead of n) yields Theorem 5.4.4 **(d)**, which provides four new proofs of Theorem 5.4.4 **(d)**.

In [Kratte17, Theorem 10.4.5], Krattenthaler proves an equivalent version of Theorem 5.4.4 **(d)**: Namely, he shows that

$$L_{n,m,k} = \frac{n+1-km}{n+m+1}\binom{n+m+1}{m}.$$

The equivalence between this and Theorem 5.4.4 **(d)** follows from easy manipulation of BCs. □

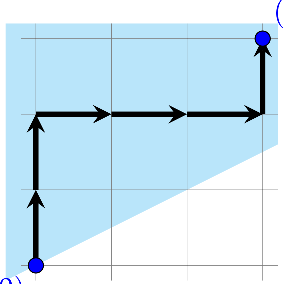
The numbers

$$L_{km,m,k} = \frac{1}{km+1}\binom{(k+1)\,m}{m}$$

from Theorem 5.4.4 **(e)** are known as the **Fuss–Catalan numbers** (see, e.g., [Stanle15, Problem A14]), and appear in many of the same contexts as the Catalan numbers (which are their particular cases for $k=1$: namely, we have $C_m = L_{m,m,1}$). For instance, just as the Catalan number C_n counts the decompositions of an $(n+2)$-gon into triangles, the Fuss–Catalan number $L_{km,m,k}$ counts the decompositions of a $(km+2)$-gon into $(k+2)$-gons. In [HilPed91] (where the $L_{km,m,k}$ are uninventively called "generalized m^{th} Catalan numbers"), this and some other properties are proved.

[18] Namely, the LPs $\mathbf{v}$ counted in [GouSer03, Theorem 2] are required to satisfy $y \geq kx$ (rather than $x \geq ky$) for all $(x,y) \in \mathbf{v}$. However, this difference does not affect their #, since we can go from them to our k-Catalan LPs by replacing each point (x,y) by (y,x). (Geometrically, this means that they are the reflections of our k-Catalan LPs across the diagonal $y = x$.)

Remark 5.4.5. It appears natural to study the counterpart of k-Catalan LPs in which the inequality $x \geq ky$ has been replaced by the opposite inequality $x \leq ky$. Thus, we fix an integer $k \in \mathbb{N}$, and we define a **k-anti-Catalan LP** to be a LP $\mathbf{v}$ such that each $(x, y) \in \mathbf{v}$ satisfies $x \leq ky$. For example, the following LP from $(0, 0)$ to $(3, 3)$ is 2-anti-Catalan:



.

We define $L'_{n,m,k}$ to be the # of k-anti-Catalan LPs from $(0, 0)$ to (n, m). One might now hope for a nice formula for $L'_{n,m,k}$, similar to the formula for $L_{n,m,k}$ given in Theorem 5.4.4 **(d)**. However, such a formula doesn't seem to exist; I am only aware of some formulas that involve summation signs. (Such formulas can be derived from [Kratte17, Theorem 10.4.7] or from [FreSel01, Theorem 2].)

5.5. Rational Catalan LPs

Catalan LPs must not cross a line with slope 1 (the line $y = x$). For k-Catalan LPs, this line is replaced by a line with slope $k \in \mathbb{N}$ (the line $y = kx$).

Can we replace it by a line with rational slope? The simplest example of such a line is the line $ax = by$, where a and b are two given integers. We can WLOG assume that a and b are coprime[19], since the line $ax = by$ doesn't change when we divide the numbers a and b by their greatest common divisor. We also assume that a and b are positive. We now define:

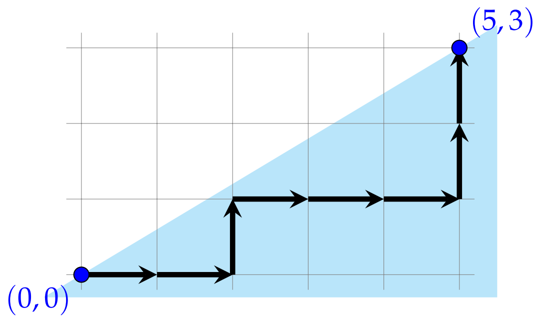
Definition 5.5.1. Let a and b be two coprime positive integers. A LP $\mathbf{v}$ is said to be (a, b)**-Catalan** if each $(x, y) \in \mathbf{v}$ satisfies $ax \geq by$.

Thus, a LP is (a, b)-Catalan if it never strays above the "slanted" diagonal $ax = by$.

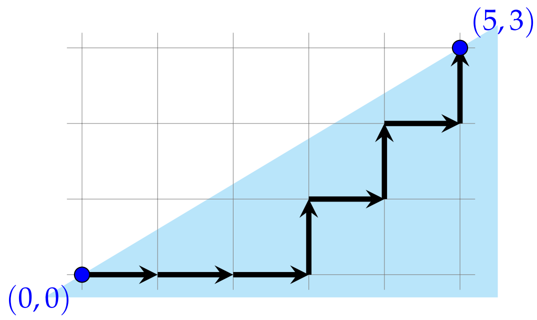
Example 5.5.2. Let $a = 3$ and $b = 5$. Here are two (a, b)-Catalan LPs from $(0, 0)$ to

[19]Recall that two integers a and b are said to be **coprime** if their greatest common divisor is 1. When $b \neq 0$, this is tantamount to saying that the fraction a/b is reduced.

(b, a):

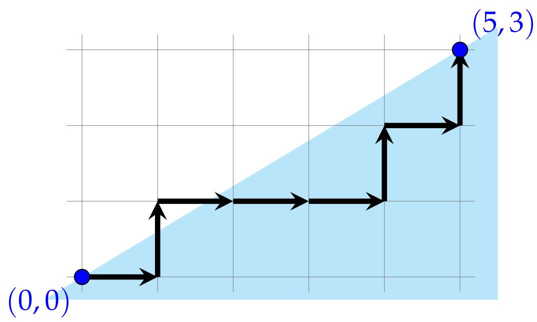


and



.

Here is an LP from $(0, 0)$ to (b, a) that is not (a, b)-Catalan (since it contains the point $(1, 1)$, which fails the $ax \geq by$ condition):



.

Of course, (a, b)-Catalan LPs are a generalization of k-Catalan LPs (since a k-Catalan LP is the same as a $(1, k)$-Catalan LP). Once again, we are paying for the extra generality with some lost simplicity: There is (to my knowledge) no analogue to Theorem 5.4.4 **(d)** for (a, b)-Catalan LPs. There seems to be no simple formula for the # of (a, b)-Catalan LPs from $(0, 0)$ to a given point. However, there is still a beautiful formula for the case when the ending point is (b, a):

Theorem 5.5.3. Let a and b be two coprime positive integers. Then,

$$(\text{\# of } (a, b) \text{-Catalan LPs from } (0, 0) \text{ to } (b, a)) = \frac{1}{a+b}\binom{a+b}{a}.$$

A few words are in order before we prove this theorem. Theorem 5.5.3 is a celebrated result, appearing in [Kratte17, Theorem 10.4.1] and also (in a slightly restated form) in [Loehr17, Theorem 12.1]. The number $\dfrac{1}{a+b}\dbinom{a+b}{a}$ on the right hand side is called a **rational Catalan number**, and the word "rational" here is referring not to its rationality (it is actually an integer, since it counts certain LPs![20]) but to the fact that the "border line" $ax = by$ has rational (as opposed to integer) slope.

Proof of Theorem 5.5.3 (sketched). Let $\mathcal{L}$ denote the set of **all** LPs from $(0,0)$ to (b,a) is $\dbinom{a+b}{a}$. Thus,

$$\begin{aligned} |\mathcal{L}| &= (\text{\# of } \textbf{all} \text{ LPs from } (0,0) \text{ to } (b,a)) \\ &= \binom{a+b}{a} \qquad \text{(by Proposition 5.1.4)}. \end{aligned}$$

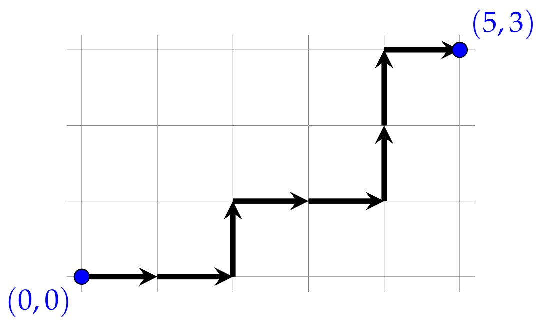
For any LP $\mathbf{v} \in \mathcal{L}$, we define a new LP $S(\mathbf{v})$, which is constructed as follows:

- The LP $S(\mathbf{v})$ starts at $(0,0)$ (just like $\mathbf{v}$ does).
- The 1-st step of $S(\mathbf{v})$ is the 2-nd step of $\mathbf{v}$.
- The 2-nd step of $S(\mathbf{v})$ is the 3-rd step of $\mathbf{v}$.
- The 3-rd step of $S(\mathbf{v})$ is the 4-th step of $\mathbf{v}$.
- And so on, until we get to the last step.
- The last step of $S(\mathbf{v})$ is the first step of $\mathbf{v}$.

Thus, the step sequence of $S(\mathbf{v})$ is precisely the step sequence of $\mathbf{v}$, rotated cyclically one step to the left (so that the first step has become the last). This uniquely defines $S(\mathbf{v})$ (since a LP is uniquely determined by its starting point and its step sequence). Moreover, $S(\mathbf{v})$ is again an LP from $(0,0)$ to (b,a) (since it has the same steps as $\mathbf{v}$, just in a different order), and hence belongs to $\mathcal{L}$. We call $S(\mathbf{v})$ the **cyclic shift** of $\mathbf{v}$.

Example 5.5.4. Assume that $\mathbf{v}$ is the LP from $(0,0)$ to $(5,3)$ whose step sequence is $RRURRUUR$. This LP looks as follows:

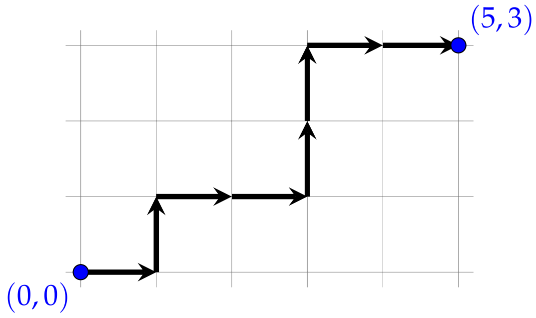
$\mathbf{v} =$



.

[20] Fun exercise: Prove that $\dfrac{1}{a+b}\dbinom{a+b}{a}$ is an integer **without** using Theorem 5.5.3! (Don't forget to use the assumption that a and b are coprime, since this statement doesn't hold otherwise.)

Then, $S(\mathbf{v})$ is the LP from $(0,0)$ to $(5,3)$ whose step sequence is $RURRUURR$ (note that this is just $RRURRUUR$, rotated cyclically one step to the left). This LP looks as follows:

$S(\mathbf{v}) =$ 

.

Geometrically, $S(\mathbf{v})$ can be obtained from $\mathbf{v}$ by cut-and-pasting the first step of $\mathbf{v}$ to the very end of $\mathbf{v}$ and then translating[21] the resulting LP to make it start at $(0,0)$ again.

Thus, we have defined a map

$$S : \mathcal{L} \to \mathcal{L},$$
$$\mathbf{v} \mapsto S(\mathbf{v}).$$

Note that each LP $\mathbf{v}$ in $\mathcal{L}$ is an LP from $(0,0)$ to (b,a), and thus has $b+a=a+b$ steps. Hence, if we rotate its step sequence cyclically $a+b$ times (each time rotating it one step to the left), then we recover its original step sequence. Therefore, $S^{a+b}(\mathbf{v}) = \mathbf{v}$ (since the step sequence of $\mathbf{v}$ gets rotated $a+b$ times when we apply S^{a+b}). This holds for every $\mathbf{v} \in \mathcal{L}$; thus, we conclude that $S^{a+b} = \operatorname{id}$.

Hence, the map S is bijective (with inverse S^{a+b-1}), and thus is a permutation of the finite set $\mathcal{L}$. Consider the orbits of this permutation. We claim the following:

Claim 1: Each orbit of S has size $a+b$.

Claim 2: Each orbit of S contains exactly one (a,b)-Catalan LP.

Once Claims 1 and 2 are proven, it will follow that exactly one in $a+b$ LPs in $\mathcal{L}$ is (a,b)-Catalan (because each orbit of S has exactly one (a,b)-Catalan LP among its $a+b$ many elements), and therefore we will conclude that

$$(\text{\# of } (a,b)\text{-Catalan LPs in } \mathcal{L}) = \frac{1}{a+b}|\mathcal{L}| = \frac{1}{a+b}\binom{a+b}{a}$$

(since $|\mathcal{L}| = \binom{a+b}{a}$). This will immediately yield Theorem 5.5.3 (since $\mathcal{L}$ is the set of all LPs from $(0,0)$ to (b,a)). Thus, it remains to prove Claims 1 and 2.

In order to prove both claims, let us fix an orbit C of S. We must prove that $|C| = a+b$ and that C contains exactly one (a,b)-Catalan LP.

[21] To **translate** a LP by a given pair $(u,v) \in \mathbb{Z}^2$ means to replace each point (p,q) of this LP by the point $(p+u, q+v)$. Geometrically, this means that the LP undergoes a parallel translation by the vector (u,v). Clearly, any LP can be made to start at $(0,0)$ by an appropriate translation.

Choose any $\mathbf{v} \in C$. Then, C is the orbit of S containing $\mathbf{v}$. In other words, $C = [\mathbf{v}]_\sim$, where $\sim$ is the equivalence relation $\overset{S}{\sim}$. Hence,

$$\begin{aligned} C &= \{S^m(\mathbf{v}) \mid m \in \mathbb{N}\} \qquad \text{(by Proposition 4.3.4 **(a)** in Lecture 27)} \\ &= \left\{S^0(\mathbf{v}),\ S^1(\mathbf{v}),\ \ldots,\ S^{a+b-1}(\mathbf{v})\right\} \end{aligned} \tag{9}$$

(since $S^{a+b}(\mathbf{v}) = \mathbf{v}$). We set out to show the following claim (from which Claims 1 and 2 will easily follow):

> *Claim 3:* There is exactly one $i \in \{0, 1, \ldots, a+b-1\}$ such that the LP $S^i(\mathbf{v})$ is (a,b)-Catalan.

The easiest way to prove Claim 3 is to regard the LPs $S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})$ not as separate LPs, but rather as pieces of one longer LP $\mathbf{w}$, which goes from $(0,0)$ to $(2b, 2a)$ (and thus has twice as many steps as $\mathbf{v}$). We will construct this longer LP by "copy-pasting" the LP $\mathbf{v}$ and translating its copy by the vector (b,a) (so that its starting point becomes (b,a)).

Here is this construction in detail. The LP $\mathbf{v}$ is a LP from $(0,0)$ to (b,a); thus, it has b right-steps and a up-steps. Consequently, $\mathbf{v}$ has $b + a = a + b$ steps in total. Thus, we can write $\mathbf{v}$ in the form $\mathbf{v} = (v_0, v_1, \ldots, v_{a+b})$. Let us do so. Thus, $v_0 = (0,0)$ and $v_{a+b} = (b,a)$.

Let $\mathbf{v}'$ be the LP $\mathbf{v}$ translated by the vector (b,a). That is, we set

$$\mathbf{v}' := (v_0 + (b,a),\ v_1 + (b,a),\ \ldots,\ v_{a+b} + (b,a)). \tag{10}$$

The starting point of this new LP $\mathbf{v}'$ is $\underbrace{v_0}_{=(0,0)} + (b,a) = (0,0) + (b,a) = (b,a) = v_{a+b}$. Let us denote the remaining points of $\mathbf{v}'$ by $v_{a+b+1}, v_{a+b+2}, \ldots, v_{2a+2b}$, respectively (since v_i has so far been defined only for $i \in \{0, 1, \ldots, a+b\}$). Thus,

$$\mathbf{v}' = (v_{a+b}, v_{a+b+1}, \ldots, v_{2a+2b}). \tag{11}$$

Comparing this with (10), we obtain

$$(v_{a+b}, v_{a+b+1}, \ldots, v_{2a+2b}) = (v_0 + (b,a),\ v_1 + (b,a),\ \ldots,\ v_{a+b} + (b,a)).$$

In other words,

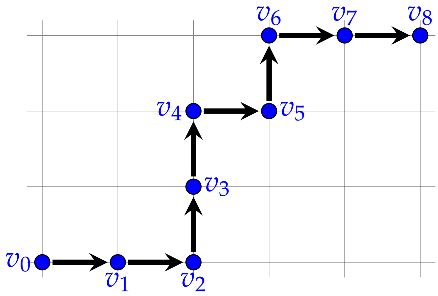
$$v_{a+b+i} = v_i + (b,a) \qquad \text{for each } i \in \{0, 1, \ldots, a+b\}. \tag{12}$$

Now, the LP $\mathbf{v} = (v_0, v_1, \ldots, v_{a+b})$ ends at the same point at which the LP $\mathbf{v}' = (v_{a+b}, v_{a+b+1}, \ldots, v_{2a+2b})$ starts (namely, at v_{a+b}). Hence, we can combine these two LPs into the LP

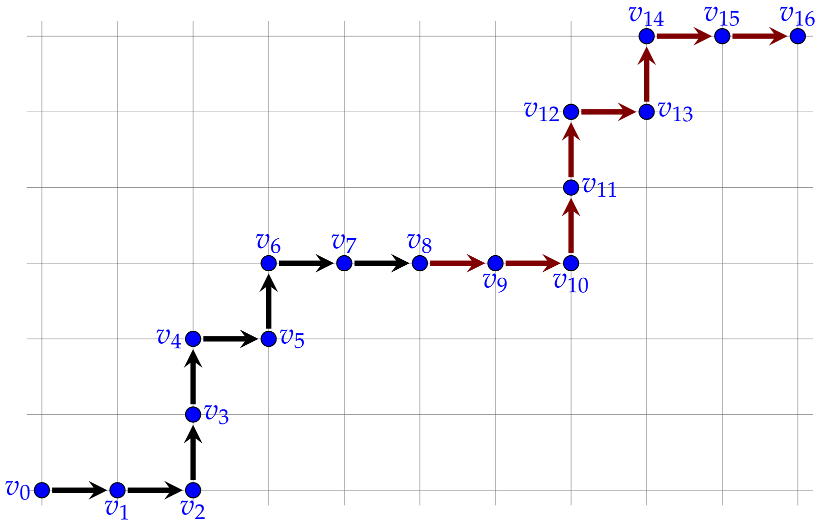
$$\mathbf{w} := (v_0, v_1, \ldots, v_{2a+2b}),$$

which has $2a + 2b$ steps.

Example 5.5.5. For example, let $a = 3$ and $b = 5$, and let $\mathbf{v}$ be the following LP:



(a LP from $(0,0)$ to $(5,3)$). Then, $\mathbf{w}$ is the following LP:



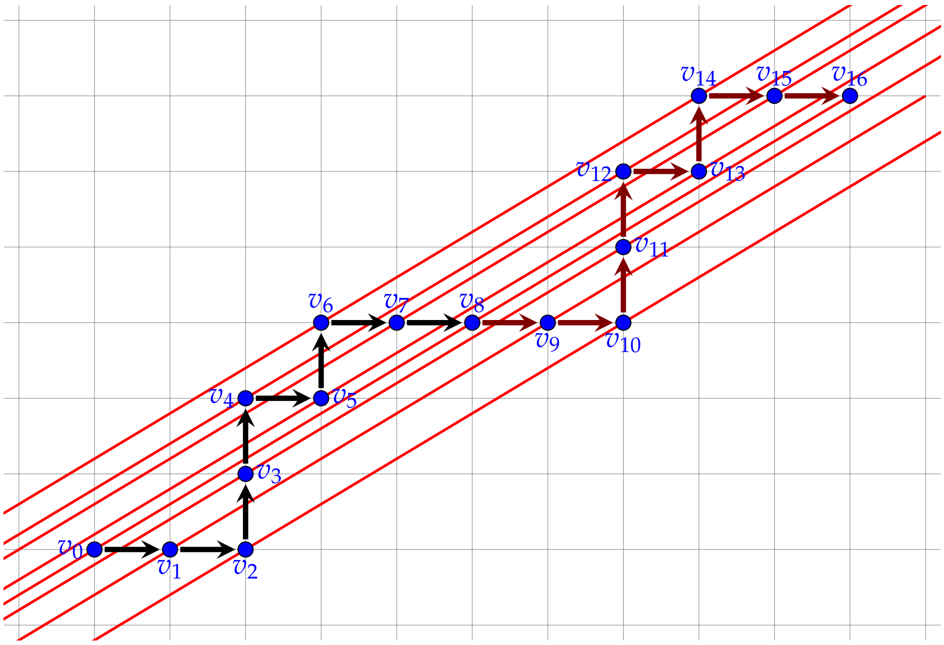
(where we painted all steps of $\mathbf{v}$ black and all steps of $\mathbf{v}'$ dark-red).

Back to the general case. For each $i \in \{0, 1, \ldots, 2a + 2b\}$, we let ℓ_i be the line[22] with slope $\dfrac{a}{b}$ (that is, parallel to the line with equation $ax = by$) passing through the point v_i. Thus, $\ell_0, \ell_1, \ldots, \ell_{2a+2b}$ are $2a + 2b + 1$ parallel lines. We will soon see that some of these lines coincide, while others don't. First, an example:

Example 5.5.6. If a, b and $\mathbf{v}$ are as in Example 5.5.5, then let us draw these $2a + 2b + 1$

[22]The word "line" always "straight line" in this proof.

lines in red:



.

Note that there are only $8 = a + b$ distinct red lines here, and these are (listed from top to bottom)

$$\ell_6 = \ell_{14}, \qquad \ell_4 = \ell_{12}, \qquad \ell_7 = \ell_{15}, \qquad \ell_5 = \ell_{13},$$
$$\ell_0 = \ell_8 = \ell_{16}, \qquad \ell_3 = \ell_{11}, \qquad \ell_1 = \ell_9, \qquad \ell_2 = \ell_{10}.$$

Some of the things you see in this example hold in full generality.

First of all, we have $\ell_i = \ell_{a+b+i}$ for each $i \in \{0, 1, \ldots, a+b\}$ [23]. In other words, the sequence $(\ell_0, \ell_1, \ldots, \ell_{2a+2b})$ is periodic with period $a + b$. Therefore, the $2a + 2b + 1$ lines $\ell_0, \ell_1, \ldots, \ell_{2a+2b}$ are just the $a + b$ lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$, each repeated twice or (in the case of ℓ_0) thrice.

Next, we shall show that the $a + b$ lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$ are distinct.

[*Proof:* Assume the contrary. Thus, two of these $a + b$ lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$ coincide. In other words, there exist two elements $i < j$ of $\{0, 1, \ldots, a+b-1\}$ such that $\ell_i = \ell_j$. Consider these i and j.

Write the points v_i and v_j as $v_i = (x_i, y_i)$ and $v_j = (x_j, y_j)$. Note that the path $\mathbf{w}$ takes $j - i$ steps to get from v_i to v_j. Among these $j - i$ steps, there must be exactly $x_j - x_i$ right-steps (since $v_i = (x_i, y_i)$ and $v_j = (x_j, y_j)$) and exactly $y_j - y_i$ up-steps (for the

[23] *Proof.* Let $i \in \{0, 1, \ldots, a+b\}$. Then, $v_{a+b+i} = v_i + (b, a)$ (by (12)). Thus, the line joining the points v_i and v_{a+b+i} has slope $\dfrac{a}{b}$. Therefore, the line with slope $\dfrac{a}{b}$ passing through v_i coincides with the line with slope $\dfrac{a}{b}$ passing through v_{a+b+i}. But the former line has been called ℓ_i, while the latter has been called ℓ_{a+b+i}. Thus, ℓ_i coincides with ℓ_{a+b+i}. In other words, $\ell_i = \ell_{a+b+i}$, qed.

same reason). Hence, we must have

$$j - i = \left(x_j - x_i\right) + \left(y_j - y_i\right). \tag{13}$$

Thus, $\left(x_j - x_i\right) + \left(y_j - y_i\right) = j - i > 0$ (since $i < j$), so that the two numbers $x_j - x_i$ and $y_j - y_i$ cannot both be 0. Their ratio $\dfrac{y_j - y_i}{x_j - x_i}$ is thus well-defined, if we understand a fraction with zero denominator (but nonzero numerator) to be the symbol ∞.

However, the line ℓ_i passes through v_i, whereas the line ℓ_j passes through v_j. Since we have assumed that $\ell_i = \ell_j$, we thus conclude that the line $\ell_i = \ell_j$ passes through both v_i and v_j. The slope of this line must therefore be $\dfrac{y_j - y_i}{x_j - x_i}$ (by the standard formula for the slope of the line through two given points, since $v_i = (x_i, y_i)$ and $v_j = \left(x_j, y_j\right)$). Thus, $\dfrac{y_j - y_i}{x_j - x_i} = \dfrac{a}{b}$ (since the slope of the line $\ell_i = \ell_j$ is $\dfrac{a}{b}$ by the definition of this line). In other words, $b\left(y_j - y_i\right) = a\left(x_j - x_i\right)$. Hence, the product $b\left(y_j - y_i\right)$ is divisible by a. In other words, $a \mid b\left(y_j - y_i\right)$.

However, a classical number-theoretical fact (see, e.g., [19s, Theorem 2.10.6], where I use the symbol "$\perp$" for "coprime to") shows that if some integer c satisfies $a \mid bc$, then $a \mid c$ (since a and b are coprime). Applying this to $c = y_j - y_i$, we obtain $a \mid y_j - y_i$ (since $a \mid b\left(y_j - y_i\right)$). In other words, $y_j - y_i = ma$ for some integer m. Consider this m. Hence, the equality $b\left(y_j - y_i\right) = a\left(x_j - x_i\right)$ (which we proved above) can be rewritten as $bma = a\left(x_j - x_i\right)$. Cancelling a from this equality, we obtain $bm = x_j - x_i$ (since a is positive and thus nonzero). Hence, $x_j - x_i = bm$. Now, (13) becomes

$$j - i = \underbrace{\left(x_j - x_i\right)}_{=bm} + \underbrace{\left(y_j - y_i\right)}_{=ma} = bm + ma = m(a+b).$$

Since both numbers $j - i$ and $a + b$ in this equality are positive, the factor m must also be positive. Hence, $m \geq 1$, so that $m(a+b) \geq a + b$. Thus, $j - i = m(a+b) \geq a + b$.

However, from $i, j \in \{0, 1, \ldots, a+b-1\}$, we obtain $i \geq 0$ and $j \leq a + b - 1 < a + b$, so that $j - \underbrace{i}_{\geq 0} \leq j < a + b$. This contradicts $j - i \geq a + b$. This contradiction shows that our assumption was false. Hence, we have proved that the $a + b$ lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$ are distinct.]

We note that the line ℓ_0 is defined to be the line through $v_0 = (0, 0)$ with slope $\dfrac{a}{b}$. In other words, ℓ_0 is the line with equation $ax = by$.

Now, we shall show the following:

> *Claim 4:* Let $i \in \{0, 1, \ldots, a+b-1\}$. Then, the LP $S^i(\mathbf{v})$ is (a,b)-Catalan if and only if the line ℓ_i is the highest[24] among the parallel lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$.

[*Proof of Claim 4:* We observe that the LP $S^i(\mathbf{v})$ has starting point $(0, 0) = v_0$.

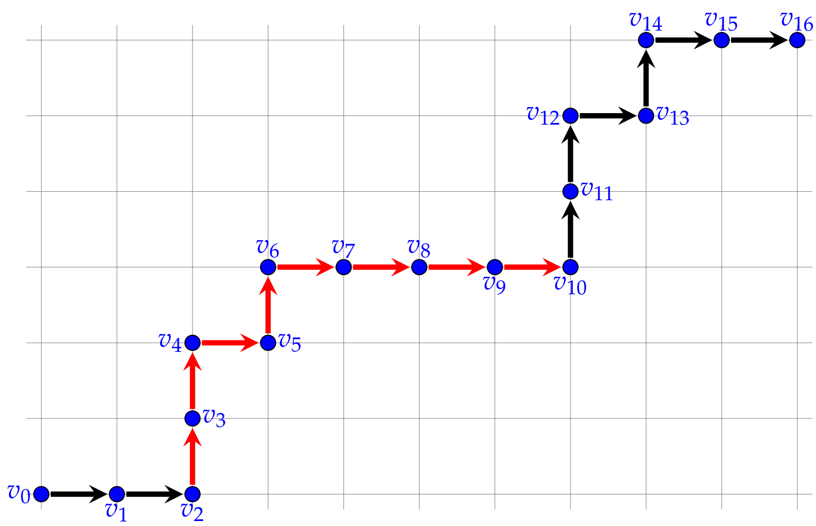
[24] The lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$ are parallel, and their direction is not vertical (since their slope is $\dfrac{a}{b} \neq \infty$). Thus, they can be ranked by their height. (For example, you can define the height of a line as the y-coordinate of the point where it intersects the y-axis.)

Let $\mathbf{v}^{+i}$ be the LP $(v_i, v_{i+1}, v_{i+2}, \ldots, v_{i+a+b})$. This LP $\mathbf{v}^{+i}$ is a piece of the LP $\mathbf{w} = (v_0, v_1, \ldots, v_{2a+2b})$, and thus its step sequence consists of the last $a+b-i$ steps of $\mathbf{v}$ followed by the first i steps of $\mathbf{v}'$ (since the path $\mathbf{w}$ is obtained by combining $\mathbf{v}$ with $\mathbf{v}'$). Since the steps of $\mathbf{v}'$ are precisely the steps of $\mathbf{v}$ (by the definition of $\mathbf{v}'$), we can restate this as follows: The step sequence of $\mathbf{v}^{+i}$ consists of the last $a+b-i$ steps of $\mathbf{v}$ followed by the first i steps of $\mathbf{v}$. In other words, the step sequence of $\mathbf{v}^{+i}$ is the step sequence of $\mathbf{v}$, rotated i steps to the left (so that the first i steps are moved to the end). But this is precisely how we characterized the step sequence of $S^i(\mathbf{v})$ above.[25]

Hence, the LPs $\mathbf{v}^{+i}$ and $S^i(\mathbf{v})$ have the same sequence. This does not mean that these two LPs are literally identical, since they start at different points (unless $i = 0$). However, it does mean that the LP $\mathbf{v}^{+i}$ can be obtained from the LP $S^i(\mathbf{v})$ by a translation – namely, by the translation that sends the starting point v_0 of the latter LP to the starting point v_i of the former LP. Let τ denote this translation. Thus, $\tau(v_0) = v_i$ and $\tau\left(S^i(\mathbf{v})\right) = \mathbf{v}^{+i}$.

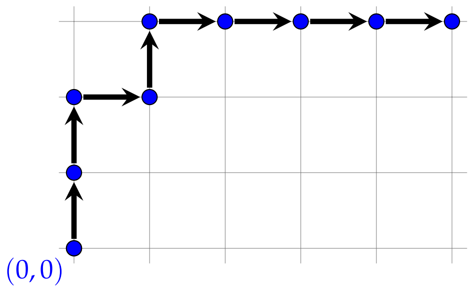
A translation sends any line to a line parallel to it. Hence, the image $\tau(\ell_0)$ of the line ℓ_0 under the translation τ is a line parallel to ℓ_0 and passing through the point $\tau(v_0)$

[25]Here is an illustration (for $i = 2$): If $\mathbf{v}$ is the LP from Example 5.5.5, then the LP $\mathbf{v}^{+2}$ consists of the red steps in the following picture:



,

whereas $S^2(\mathbf{v})$ is the following LP:



(since ℓ_0 passes through v_0). In other words, $\tau(\ell_0)$ is a line parallel to ℓ_0 and passing through v_i (since $\tau(v_0) = v_i$). But this means that it must be the line ℓ_i (since the line ℓ_i is parallel to ℓ_0 and passes through v_i). Thus, we have shown that $\tau(\ell_0) = \ell_i$.

Now, recall that a LP is (a,b)-Catalan if and only if it never strays above the line with equation $ax = by$. But this line is precisely the line ℓ_0. Hence, a LP is (a,b)-Catalan if and only if it never strays above the line ℓ_0.

Therefore, we have the following chain of logical equivalences:

$$\begin{aligned}
& \left(\text{the LP } S^i(\mathbf{v}) \text{ is } (a,b)\text{-Catalan}\right) \\
\Longleftrightarrow\ & \left(\text{the LP } S^i(\mathbf{v}) \text{ never strays above the line } \ell_0\right) \\
\Longleftrightarrow\ & \left(\text{the LP } \tau\left(S^i(\mathbf{v})\right) \text{ never strays above the line } \tau(\ell_0)\right) \\
& \qquad \left(\begin{array}{c} \text{since the translation } \tau \text{ preserves the} \\ \text{“never strays above” relation} \end{array}\right) \\
\Longleftrightarrow\ & \left(\text{the LP } \mathbf{v}^{+i} \text{ never strays above the line } \ell_i\right) \qquad \left(\begin{array}{c} \text{since } \tau\left(S^i(\mathbf{v})\right) = \mathbf{v}^{+i} \\ \text{and } \tau(\ell_0) = \ell_i \end{array}\right) \\
\Longleftrightarrow\ & \left(\text{none of the points } v_i, v_{i+1}, v_{i+2}, \ldots, v_{i+a+b} \text{ lies above the line } \ell_i\right) \\
& \qquad \left(\text{since } \mathbf{v}^{+i} = (v_i, v_{i+1}, v_{i+2}, \ldots, v_{i+a+b})\right) \\
\Longleftrightarrow\ & \left(\text{none of the lines } \ell_i, \ell_{i+1}, \ell_{i+2}, \ldots, \ell_{i+a+b} \text{ is higher than the line } \ell_i\right) \\
& \qquad \left(\begin{array}{c} \text{since a point } v_j \text{ lies above the line } \ell_i \\ \text{if and only if the line } \ell_j \text{ is higher than } \ell_i \\ \text{(because } \ell_j \text{ is the line parallel to } \ell_i \text{ through } v_j\text{)} \end{array}\right) \\
\Longleftrightarrow\ & \left(\ell_i \text{ is the highest of the } a+b+1 \text{ lines } \ell_i, \ell_{i+1}, \ell_{i+2}, \ldots, \ell_{i+a+b}\right) \\
\Longleftrightarrow\ & \left(\ell_i \text{ is the highest of the } 2a+2b+1 \text{ lines } \ell_0, \ell_1, \ldots, \ell_{2a+2b}\right) \\
& \qquad \left(\begin{array}{c} \text{since } \left\{\ell_i, \ell_{i+1}, \ell_{i+2}, \ldots, \ell_{i+a+b}\right\} = \left\{\ell_0, \ell_1, \ldots, \ell_{2a+2b}\right\} \text{ as sets} \\ \text{(because the sequence } (\ell_0, \ell_1, \ldots, \ell_{2a+2b}) \text{ is periodic} \\ \text{with period } a+b, \text{ and thus any } a+b \text{ consecutive} \\ \text{entries of this sequence contain all its entries)} \end{array}\right) \\
\Longleftrightarrow\ & \left(\ell_i \text{ is the highest of the } a+b \text{ lines } \ell_0, \ell_1, \ldots, \ell_{a+b-1}\right) \\
& \qquad \text{(for a similar reason as the previous equivalence)} \\
\Longleftrightarrow\ & \left(\ell_i \text{ is the highest among the parallel lines } \ell_0, \ell_1, \ldots, \ell_{a+b-1}\right).
\end{aligned}$$

This proves Claim 4.]

Claims 1, 2 and 3 are now at arm's reach:

[*Proof of Claim 3:* We must prove that there is exactly one $i \in \{0, 1, \ldots, a+b-1\}$ such that the LP $S^i(\mathbf{v})$ is (a,b)-Catalan. In view of Claim 4, this is equivalent to proving that there is exactly one $i \in \{0, 1, \ldots, a+b-1\}$ such that the line ℓ_i is the highest among the parallel lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$. In other words, we must prove that among the $a+b$ lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$, exactly one is the highest (i.e., there is no tie). But this follows from the fact (proved above) that the $a+b$ lines $\ell_0, \ell_1, \ldots, \ell_{a+b-1}$ are distinct (and thus have different heights). Thus, Claim 3 is proven.]

Now, Claim 3 says that there is exactly one (a,b)-Catalan LP among the $a+b$ LPs $S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})$. Therefore, the set $\{S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})\}$ contains exactly one (a,b)-Catalan LP. In view of (9), this means that the set C contains exactly one (a,b)-Catalan LP.

Moreover, it is now easy to prove that the $a+b$ LPs $S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})$ are distinct.

[*Proof.* Assume the contrary. Thus, there exist two elements $p<q$ of $\{0,1,\ldots,a+b-1\}$ such that $S^p(\mathbf{v}) = S^q(\mathbf{v})$. Consider these p and q.

The map S is bijective, so its inverse S^{-1} and its negative powers S^k for $k<0$ exist. Applying S^{-p} to both sides of the equality $S^p(\mathbf{v}) = S^q(\mathbf{v})$, we obtain $S^{-p}(S^p(\mathbf{v})) = S^{-p}(S^q(\mathbf{v})) = S^{-p+q}(\mathbf{v}) = S^{q-p}(\mathbf{v})$. Hence, $S^{q-p}(\mathbf{v}) = S^{-p}(S^p(\mathbf{v})) = \mathbf{v}$.

From $q \le a+b-1$ and $p \ge 0$, we obtain $q - \underbrace{p}_{\ge 0} \le q \le a+b-1$, so that $q-p \in [a+b-1]$ (since $p<q$ leads to $q-p>0$).

However, Claim 3 (with i renamed as k) shows that there is exactly one $k \in \{0,1,\ldots,a+b-1\}$ such that the LP $S^k(\mathbf{v})$ is (a,b)-Catalan. Consider this k.

But $\mathbf{v}$ is just an arbitrary element of C. Now, $S^k(\mathbf{v})$ is an element of C as well (since $C = \{S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})\}$ clearly contains $S^k(\mathbf{v})$). Thus, we can apply Claim 3 to $S^k(\mathbf{v})$ instead of $\mathbf{v}$. We thus conclude that there is exactly one $i \in \{0,1,\ldots,a+b-1\}$ such that $S^i(S^k(\mathbf{v}))$ is (a,b)-Catalan. Since 0 is such an i (because $S^0(S^k(\mathbf{v})) = S^k(\mathbf{v})$ is (a,b)-Catalan), we thus conclude that 0 is the **only** such i. In other words, for any $i \in [a+b-1]$, the LP $S^i(S^k(\mathbf{v}))$ is **not** (a,b)-Catalan. Applying this to $i = q-p$, we conclude that $S^{q-p}(S^k(\mathbf{v}))$ is **not** (a,b)-Catalan (since $q-p \in [a+b-1]$).

However,

$$S^{q-p}\left(S^k(\mathbf{v})\right) = S^{(q-p)+k}(\mathbf{v}) = S^{k+(q-p)}(\mathbf{v}) = S^k\left(\underbrace{S^{q-p}(\mathbf{v})}_{=\mathbf{v}}\right) = S^k(\mathbf{v}),$$

which is (a,b)-Catalan (as we know). This contradicts the fact that $S^{q-p}(S^k(\mathbf{v}))$ is **not** (a,b)-Catalan. This contradiction shows that our assumption was false. Hence, we have shown that the $a+b$ LPs $S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})$ are distinct.]

This entails that $|\{S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})\}| = a+b$. In other words, $|C| = a+b$ (since $C = \{S^0(\mathbf{v}), S^1(\mathbf{v}), \ldots, S^{a+b-1}(\mathbf{v})\}$). In other words, C has size $a+b$. Recall also that C contains exactly one (a,b)-Catalan LP (as we have proved above).

Forget that we fixed C. We thus have shown that every orbit C of S has size $a+b$ and contains exactly one (a,b)-Catalan LP. This proves Claims 1 and 2. As explained above, this completes the proof of Theorem 5.5.3. □

Recall that k-Catalan paths are the same as $(1,k)$-Catalan paths. However, Theorem 5.5.3 is specific to LPs from $(0,0)$ to (b,a), which are not very interesting in the particular case when $a=1$. Thus, it may appear that Theorem 5.5.3 is unrelated to Theorem 5.4.4. However, this appearance is deceptive. Thanks to a subtle trick, we can actually derive Theorem 5.4.4 **(e)** from Theorem 5.5.3:

Proof of Theorem 5.4.4 ***(e)*** *using Theorem 5.5.3 (sketched).* Let $m \in \mathbb{N}$. WLOG assume that $m>0$ (since otherwise, Theorem 5.4.4 **(e)** is obvious).

The positive integers m and $km+1$ are coprime (since basic properties of the greatest common divisor yield $\gcd(m,\ km+1) = \gcd(m,\ 1) = 1$). Hence, we can apply Theorem 5.5.3 to $a = m$ and $b = km+1$. Thus we obtain

$$\begin{aligned}
&(\text{\# of } (m,\ km+1)\text{-Catalan LPs from } (0,0) \text{ to } (km+1,\ m)) \\
&= \frac{1}{m+km+1}\binom{m+km+1}{m} \\
&= \frac{1}{km+1}\binom{m+km}{m} \qquad \left(\begin{array}{c}\text{by some easy manipulations using} \\ \text{the factorial formula for BCs}\end{array}\right) \\
&= \frac{1}{km+1}\binom{(k+1)\,m}{m}.
\end{aligned}$$

It remains to relate the $(m,\ km+1)$-Catalan LPs from $(0,0)$ to $(km+1,\ m)$ to the k-Catalan LPs from $(0,0)$ to (km,m) (which are counted by $L_{km,m,k}$).

It turns out that the former are just one step away from the latter (literally): If $\mathbf{v} = (v_0, v_1, \ldots, v_n)$ is a k-Catalan LP from $(0,0)$ to (km,m), then

$$\mathbf{v}' := \left((0,0)\,,\ \underbrace{v_0+(1,0)\,,\ v_1+(1,0)\,,\ \ldots,\ v_n+(1,0)}_{\text{This is just the LP } \mathbf{v}\text{, translated by the vector } (1,0)}\right)$$

is a $(m,\ km+1)$-Catalan LP from $(0,0)$ to $(km+1,\ m)$ (check this![26]). Essentially, $\mathbf{v}'$ is just the LP $\mathbf{v}$, shifted to the right by 1 unit and reconnected to the old starting point $(0,0)$ by an extra right-step inserted at the front. Thus, we obtain a map

$$\begin{aligned}
&\text{from } \{k\text{-Catalan LPs from } (0,0) \text{ to } (km,m)\} \\
&\text{to } \{(m,\ km+1)\text{-Catalan LPs from } (0,0) \text{ to } (km+1,\ m)\}
\end{aligned}$$

that sends each $\mathbf{v}$ to $\mathbf{v}'$. This map is furthermore bijective[27]. Hence, by the bijection principle, we have

$$\begin{aligned}
&(\text{\# of } k\text{-Catalan LPs from } (0,0) \text{ to } (km,m)) \\
&= (\text{\# of } (m,\ km+1)\text{-Catalan LPs from } (0,0) \text{ to } (km+1,\ m)) \\
&= \frac{1}{km+1}\binom{(k+1)\,m}{m}.
\end{aligned}$$

Since the left hand side of this equality is $L_{km,m,k}$ (by definition of $L_{km,m,k}$), this yields precisely the claim of Theorem 5.4.4 **(e)**. □

[26] This boils down to checking that if a point $(x,y) \in \mathbf{v}$ satisfies $x \geq ky$, then the corresponding translated point $(x+1,y) \in \mathbf{v}'$ satisfies $m(x+1) \geq (km+1)\,y$. In order to check this, observe that $y \leq m$ (why?).

[27] Injectivity is obvious, but surjectivity is a bit tricky: It is clear that every $(m,\ km+1)$-Catalan LP from $(0,0)$ to $(km+1,\ m)$ must start with a right-step and thus has the form $\mathbf{v}'$ for **some** LP $\mathbf{v}$ from $(0,0)$ to (km,m). What is less clear is that this $\mathbf{v}$ must be k-Catalan. To show this, you need to prove that each of its points $(x,y) \in \mathbf{v}$ satisfies $x \geq ky$; you are given that $m(x+1) \geq (km+1)\,y$, since $\mathbf{v}'$ is $(m,\ km+1)$-Catalan. However, $y \in \{0,1,\ldots,m\}$ (why?). If $y > 0$, then dividing the inequality $m(x+1) \geq (km+1)\,y$ by m yields $x+1 \geq ky + \frac{y}{m} > ky$ and thus $x \geq ky$ (since x and ky are integers). If $y = 0$, then $x \geq ky$ holds anyway. Thus, in either case, we get $x \geq ky$, as desired.

5.6. Dyck paths

So far, we have been discussing LPs that consist of up-steps and right-steps. Many variations on this concept can be obtained by requiring other kinds of steps. Here is one of the most popular ones:

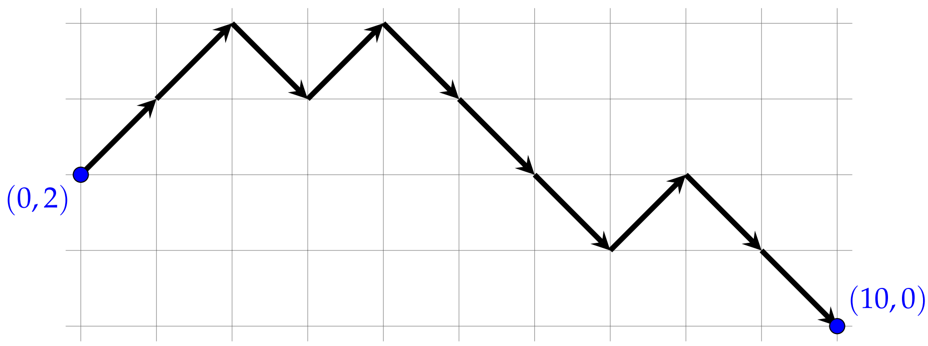
Definition 5.6.1. Let $(a,b) \in \mathbb{Z}^2$ and $(c,d) \in \mathbb{Z}^2$ be two points. Then, a **diagonal path** from (a,b) to (c,d) is

- informally understood to be a path from (a,b) to (c,d) in the plane that uses only the following two kinds of steps:
  - "**positive steps**" (denoted "P") that go from a point (p,q) to $(p+1, q+1)$;
  - "**negative steps**" (denoted "N") that go from a point (p,q) to $(p+1, q-1)$.
- rigorously defined to be a tuple $(v_0, v_1, \ldots, v_n)$ of points $v_0, v_1, \ldots, v_n \in \mathbb{Z}^2$ such that

$$v_0 = (a,b) \qquad \text{and} \qquad v_n = (c,d) \qquad \text{and}$$
$$v_i - v_{i-1} \in \{(1,1), \ (1,-1)\} \qquad \text{for each } i \in [n].$$

  If $(v_0, v_1, \ldots, v_n)$ is a diagonal path from (a,b) to (c,d), then the differences $v_i - v_{i-1}$ (for $i \in [n]$) are called the **steps** of this path. The pair $(1,1)$ is called a **positive step** and is denoted by P; the pair $(1,-1)$ is called a **negative step** and is denoted by N.

Example 5.6.2. Here is an example of a diagonal path from $(0,2)$ to $(10,0)$:



.

Its steps are $P, P, N, P, N, N, N, P, N, N$.

Definition 5.6.3. Let $\mathbf{v} = (v_0, v_1, \ldots, v_n)$ be a diagonal path from a point (a,b) to a point (c,d). Let $p \in \mathbb{Z}^2$ be a point. We say that $p \in \mathbf{v}$ (in words: p **lies on** $\mathbf{v}$) if $p \in \{v_0, v_1, \ldots, v_n\}$.

A fundamental observation about diagonal paths is the following:

Proposition 5.6.4. Assume that all points $(x, y) \in \mathbb{Z}^2$ are colored in two colors: Namely, each point $(x, y) \in \mathbb{Z}^2$ is colored black if $x + y$ is even and white if $x + y$ is odd. (This is called the **chessboard coloring**, for obvious reasons.)

Let $\mathbf{v}$ be a diagonal path. Then, all points on $\mathbf{v}$ have the same color.

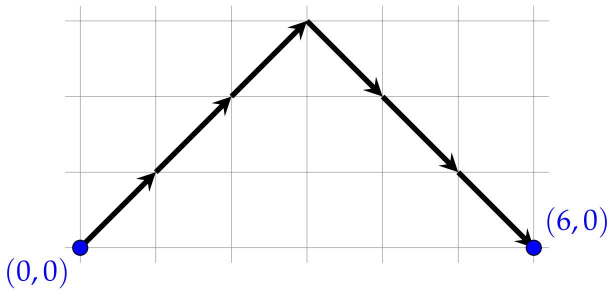
Proof. Left to the reader. □

We also leave it to the reader to find the # of diagonal paths from a point (a, b) to a point (c, d). (Note that this # will be 0 if the numbers $a + b$ and $c + d$ have different parities, because of Proposition 5.6.4.) We will instead focus on a special kind of diagonal paths (compare to Catalan paths):

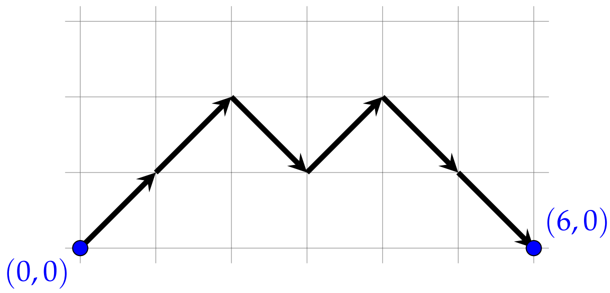
Definition 5.6.5. A diagonal path $\mathbf{v}$ is said to be **Dyck** if each $(x, y) \in \mathbf{v}$ satisfies $y \geq 0$.

In other words, a diagonal path is Dyck if and only if it never strays below the x-axis.[28]

Example 5.6.6. Here are all Dyck paths from $(0, 0)$ to $(6, 0)$:



,

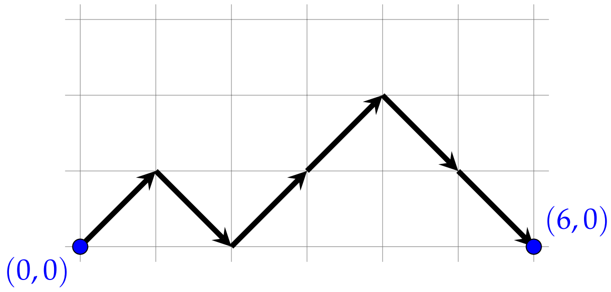


,

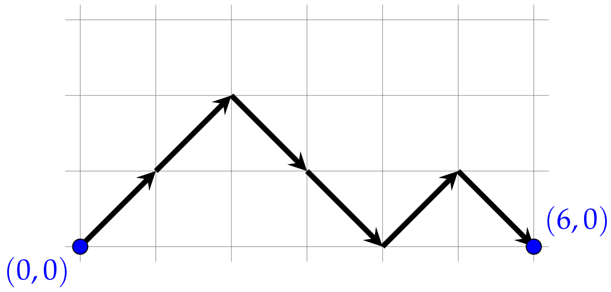
[28]Note that the name "Dyck path" does not have a standardized meaning across the literature. Some authors instead use it for what we call Catalan paths, or (like Loehr in [Loehr17]) for the reflections of the Catalan paths across the $x = y$ diagonal. This is understandable, since all of these objects are closely related (as we will soon see).

What we call "Dyck paths", on the other hand, is called "mountain ranges" in some sources (such as [Davis16, §1.2]).

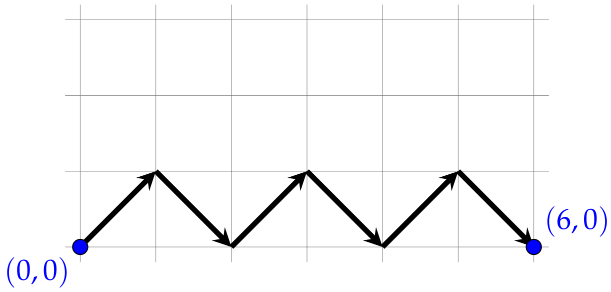
The name "Dyck paths" honors the German mathematician Walther Franz Anton von Dyck, who appears to have introduced these paths as an equivalent concept to the so-called "Dyck words" (which are precisely the "legal sequences of parentheses" we discussed above).



,



,



.

In particular, there are 5 of them.

The 5 is not a random number here; it is the Catalan number C_3. This generalizes:

Theorem 5.6.7. Let $n \in \mathbb{N}$. Then,

$$(\# \text{ of Dyck paths from } (0,0) \text{ to } (2n,0)) = C_n.$$

Proof idea. There is a map

$$\text{from } \{\text{Dyck paths from } (0,0) \text{ to } (2n,0)\}$$
$$\text{to } \{\text{Catalan paths from } (0,0) \text{ to } (n,n)\},$$

which replaces each point (x,y) on the Dyck path by $\left(\dfrac{x+y}{2}, \dfrac{x-y}{2}\right)$. Check that this map is well-defined (Proposition 5.6.4 ensures that $\dfrac{x+y}{2}$ and $\dfrac{x-y}{2}$ are integers!) and bijective (construct the inverse explicitly!). Thus, the bijection principle yields

$$\begin{aligned}
&(\text{\# of Dyck paths from } (0,0) \text{ to } (2n,0)) \\
&= (\text{\# of Catalan paths from } (0,0) \text{ to } (n,n)) \\
&= L_{n,n} \qquad (\text{by the definition of } L_{n,n}) \\
&= C_n \qquad (\text{by the definition of } C_n)\,.
\end{aligned}$$

Thus, Theorem 5.6.7 follows. □

More generally, if (a,b) and (c,d) are two black points in $\mathbb{Z}^2$ (where colors are defined as in Proposition 5.6.4), then there is a canonical bijection

$$\begin{aligned}
&\text{from } \{\text{Dyck paths from } (a,b) \text{ to } (c,d)\} \\
&\text{to } \left\{\text{Catalan paths from } \left(\frac{a+b}{2}, \frac{a-b}{2}\right) \text{ to } \left(\frac{c+d}{2}, \frac{c-d}{2}\right)\right\}.
\end{aligned}$$

A similar bijection exists for Dyck paths between two white points. Thus, studying Dyck paths is equivalent to studying Catalan paths.

5.7. More surprising integralities

For any $n \in \mathbb{N}$, the n-th Catalan number C_n is given by

$$\begin{aligned}
C_n &= \frac{1}{n+1}\binom{2n}{n} \\
&= \frac{1}{n+1} \cdot \frac{(2n)!}{n!n!} \qquad (\text{by the factorial formula for BCs}) \\
&= \frac{(2n)!}{(n+1)\cdot n! \cdot n!} = \frac{(2n)!}{(n+1)! \cdot n!} \qquad (\text{since } (n+1)\cdot n! = (n+1)!)\,.
\end{aligned}$$

Thus, we conclude that the ratio $\dfrac{(2n)!}{(n+1)! \cdot n!}$ is an integer (since C_n is an integer by its definition). This is an instance of a class of remarkable results, which claim that a certain ratio of factorials turns out to be an integer. Of course, an even simpler result of this type is the integrality of the binomial coefficient $\dbinom{n}{k} = \dfrac{n!}{k! \cdot (n-k)!}$ for all $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$. But there are also subtler results in this family. Here is another striking result of the same nature:

Proposition 5.7.1. For any $m \in \mathbb{N}$ and $n \in \mathbb{N}$, define a rational number $T(m,n)$ by

$$T(m,n) := \frac{(2m)!\,(2n)!}{m!n!\,(m+n)!}.$$

Then:
(a) We have $4T(m,n) = T(m+1,n) + T(m,n+1)$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.
(b) We have $T(m,n) \in \mathbb{N}$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(c) If $m \in \mathbb{N}$ and $n \in \mathbb{N}$ are such that $(m, n) \neq (0, 0)$, then the integer $T(m, n)$ is even.

(d) If $m \in \mathbb{N}$ and $n \in \mathbb{N}$ are such that $m + n$ is odd and $m + n > 1$, then $4 \mid T(m, n)$.

(e) We have $T(m, 0) = \dbinom{2m}{m}$ for every $m \in \mathbb{N}$.

(f) We have $T(m, n) = \dfrac{\dbinom{2m}{m}\dbinom{2n}{n}}{\dbinom{m+n}{m}}$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(g) We have $T(m, n) = T(n, m)$ for every $m \in \mathbb{N}$ and $n \in \mathbb{N}$.

(h) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m, n\}$. Then,

$$\sum_{k=-p}^{p} (-1)^k \binom{m+n}{m+k}\binom{m+n}{n+k} = \binom{m+n}{m}.$$

(i) Let $m \in \mathbb{N}$ and $n \in \mathbb{N}$. Let $p = \min\{m, n\}$. Then,

$$T(m, n) = \sum_{k=-p}^{p} (-1)^k \binom{2m}{m+k}\binom{2n}{n-k}.$$

This proposition is proved in [Grinbe15, Exercise 3.25]. The numbers $T(m, n)$ introduced in it are the so-called **super-Catalan numbers**[29]; much has been written about them (e.g., [Gessel92] and [AleGhe14]). So far, no good combinatorial interpretation has been found to "justify" their integrality (i.e., it is not clear what $T(m, n)$ counts); the known proofs of $T(m, n)$ are either number-theoretical (such as the proof in [21f-lec5, solution to Exercise 5.3.5 **(a)**], which proves an even more general result) or algebraic (such as the proof in [Grinbe15, solution to Exercise 3.25]).

There are several other integrality results like this. As already mentioned, [21f-lec5, Exercise 5.3.5 **(a)**] generalizes Proposition 5.7.1 **(b)**. Yet another generalization of Proposition 5.7.1 **(b)** is the following fact (which I have learnt from Ira Gessel):

Proposition 5.7.2. Let $a, b, c \in \mathbb{N}$. Then, $\dfrac{(a+2b)!\,(a+2c)!}{a!b!c!\,(a+b+c)!} \in \mathbb{N}$.

Proof. This is not hard to do using the methods of [21f-lec5] (comparing p-valuations of numerator and denominator), although this is far from a nice proof. Is there any combinatorial or algebraic approach? $\square$

For other examples of ratios of factorials being unexpectedly integral, see [WarZud19].

Incidentally, here is a somewhat unexpected number-theoretic property of Catalan numbers:

[29] Unlike the Catalan numbers, they have in fact been first discovered by Eugène Catalan (1874).

Proposition 5.7.3. Let $n \in \mathbb{N}$. Then, the Catalan number C_n is odd if and only if $n+1$ is a power of 2.

Proof. Exercise! (See [KosSal06] for a short proof.) □

References

[18s-mt2s] Darij Grinberg, *UMN Spring 2018 Math 4707 midterm #2 with solutions*, `http://www.cip.ifi.lmu.de/~grinberg/t/18s/mt2s.pdf`

[19s] Darij Grinberg, *Introduction to Modern Algebra (UMN Spring 2019 Math 4281 notes)*, 29 June 2019.
`http://www.cip.ifi.lmu.de/~grinberg/t/19s/notes.pdf`

[21f-lec5] Darij Grinberg, *Math 235: Mathematical Problem Solving, Fall 2021, Worksheet 5: p-valuations*.
`http://www.cip.ifi.lmu.de/~grinberg/t/21f/lec5.pdf`

[AleGhe14] Emily Allen, Irina Gheorghiciuc, *A weighted interpretation for the super Catalan numbers*, Journal of Integer Sequences, Vol. 17 (2014), Article 14.10.7.
A preprint is arXiv:1403.5246v2.

[Bona22] Miklos Bóna, *Combinatorics of Permutations*, 3rd edition, Taylor&Francis 2022.
`https://doi.org/10.1201/9780429274107`

[Davis16] Tom Davis, *Catalan Numbers*, 19 February 2016.
`http://www.geometer.org/mathcircles/catalan.pdf`

[FreSel01] Darrin D. Frey, James A. Sellers, *Generalizing Bailey's Generalization of the Catalan Numbers*, Fibonacci Quarterly, 2001, pp. 142–148.

[Gessel92] Ira M. Gessel, *Super Ballot Numbers*, Journal of Symbolic Computation, Volume 14, Issues 2–3, August-September 1992, pp. 179–194.
`https://doi.org/10.1016/0747-7171(92)90034-2`

[GouSer03] I. P. Goulden, Luis G. Serrano, *Maintaining the spirit of the reflection principle when the boundary has arbitrary integer slope*, Journal of Combinatorial Theory, Series A **104** (2003), pp. 317–326.

[Grinbe15] Darij Grinberg, *Notes on the combinatorial fundamentals of algebra*, 15 September 2022.
`http://www.cip.ifi.lmu.de/~grinberg/primes2015/sols.pdf`
The numbering of theorems and formulas in this link might shift

when the project gets updated; for a "frozen" version whose numbering is guaranteed to match that in the citations above, see `https://github.com/darijgr/detnotes/releases/tag/2022-09-15c` .

[HilPed91] Peter Hilton, Jean Pedersen, *Catalan Numbers, Their Generalization, and Their Uses*, Mathematical Intelligencer **13** (1991), no. 2, pp. 64–75.

[Kitaev11] Sergey Kitaev, *Patterns in Permutations and Words*, Springer 2011.

[KosSal06] Thomas Koshy, Mohammad Salmassi, *Parity and Primality of Catalan Numbers*, The College Mathematics Journal **37** (2006), no. 1, pp. 52–53.

[Kratte17] Christian Krattenthaler, *Lattice Path Enumeration*, arXiv:1503.05930v3, published in: Handbook of Enumerative Combinatorics, M. Bóna (ed.), Discrete Math. and Its Appl., CRC Press, Boca Raton-London-New York, 2015, pp. 589–678.
`https://arxiv.org/abs/1503.05930v3`

[Loehr17] Nicholas A. Loehr, *Combinatorics*, 2nd edition, CRC Press 2017.

[Pak15] Igor Pak, *History of Catalan numbers*, an appendix to Richard Stanley, Catalan Numbers, Cambridge Univ. Press, 2015, pp. 177–189.

[Reiner05] Victor Reiner, *An old, but cute, proof of the Catalan formula*, 2005.
`https://www-users.cse.umn.edu/~reiner/Papers/MessingCatalan.pdf`

[Renaul07] Marc Renault, *Four Proofs of the Ballot Theorem*, Mathematics Magazine **80** (2007), no. 5, pp. 345–352.

[Roman15] Steven Roman, *An Introduction to Catalan Numbers*, Birkhäuser 2015.

[Singma78] David Singmaster, *An Elementary Evaluation of the Catalan Numbers*, The American Mathematical Monthly **85**, 1978, issue 5, pp. 366–368.

[Stanle15] Richard P. Stanley, *Catalan Numbers*, 1st edition 2015.
See `http://math.mit.edu/~rstan/catalan/` for errata.

[WarZud19] S. Ole Warnaar, Wadim Zudilin, *q-rious and q-riouser*, arXiv:1909.07045v1.