

Bijections

Yufei Zhao

yufeiz@mit.edu

In this lecture, we will look at using bijections to solve combinatorics problems. Given two sets \mathcal{A} and \mathcal{B} , a *bijection* (also called *bijective correspondence*) is a map $f : \mathcal{A} \rightarrow \mathcal{B}$ that is both injective and surjective, meaning that no two elements of \mathcal{A} get mapped onto the same element in \mathcal{B} , and every element of \mathcal{B} is the image of some element of \mathcal{A} . This gives us a way of pairing up every element from \mathcal{A} with some element of \mathcal{B} . In particular, when the sets are finite, the existence of a bijection implies that $|\mathcal{A}| = |\mathcal{B}|$.

This explanation of a bijection may seem a little abstract. Let us take a look at some examples of how bijections can be used.

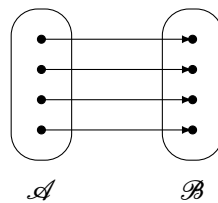
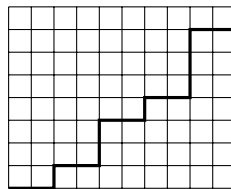


Figure 1: Illustrating a bijection from \mathcal{A} to \mathcal{B} .

Problem 1. Determine the number of walks from $(0, 0)$ to (m, n) allowing only unit steps up or to the right.



The idea of using bijection is that we want to transform the problem into something that we know better how to count. In this case, we can encode the path as a sequence of letters U and R , corresponding to whether the step taken was a unit up step (U) or a unit right step (R). The path displayed above, for instance, is encoded as

$RRURRUURRURRUUURRU$.

Note that the resulting sequence of letters always has m copies of R and n copies of U , since it takes in total of m right unit steps and n up unit steps to move from $(0, 0)$ to

(m, n) . Also, you can check that for any sequence consisting of m copies of R and n copies of U , we can construct a corresponding path encoded by this sequence. Thus we have constructed a bijection between the following two sets:

the set of walks from $(0, 0)$ to (m, n) using only unit up or right steps
 \updownarrow
the set of sequences consisting of m copies of R and n copies of U .

We know how to count the latter set. It has exactly $\binom{m+n}{m}$ elements, which can be thought of as choosing m spots to place the R 's in a sequence of length $m+n$. Therefore, the number of lattice walks from $(0, 0)$ to (m, n) must be $\binom{m+n}{m}$ as well.

The above example illustrates the use of the bijection technique. Counting lattice paths is a fairly important problem by itself that we will come back to it later in the lecture.

Usually, a proof involving a bijection between two sets \mathcal{A} and \mathcal{B} should explain the following:

1. How to obtain an element of \mathcal{B} from any element of \mathcal{A} .
2. How to recover the element of \mathcal{A} from any element of \mathcal{B} .
3. Why the above two constructions are inverses of each other.

Can you spot these elements in the solution presented above? Sometimes, like in the solution above, once the map $\mathcal{A} \rightarrow \mathcal{B}$ (i.e., the first step) has been constructed, it is easy to construct the inverse map. However, there are times when the inverse construction is not at all obvious.

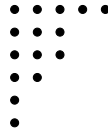
Next, let us look at a problem involving *partitions*. A partition of a positive integer n is a way of writing n as a sum of positive integers, where the order of the summands is irrelevant, though we usually write the summands in non-increasing order. For instance, the number 4 has exactly five partitions:

$$\begin{aligned} 4 &= 1 + 1 + 1 + 1 \\ 4 &= 2 + 1 + 1 \\ 4 &= 2 + 2 \\ 4 &= 3 + 1 \\ 4 &= 4 \end{aligned}$$

We will also write a partition as a non-decreasing tuple of parts, so the partitions of 4 are denoted by $(1, 1, 1, 1)$, $(2, 1, 1)$, $(2, 2)$, $(3, 1)$, (4) .

Problem 2. Let n and k be positive integers. Show that the number of partitions of n with exactly k parts equals the number of partitions of n whose largest part is exactly k .

To solve this problem, we consider a graphical representation of a partition. For each partition $n = a_1 + a_2 + \cdots + a_r$, where $a_1 \geq a_2 \geq \cdots \geq a_r > 0$, we consider a diagram with a_i dots on the i -th row, all left aligned. This is known as the *Ferrar diagram* of the partition. For instance, the partition $15 = 5+3+3+2+1+1$ corresponds to the following diagram:



The key insight here is that there exists a natural bijection between partitions of n and the set of Ferrar diagrams with n dots. The number of parts of the partition corresponds to the number of rows in the Ferrar diagram, and the size of the largest part corresponds to the number of columns of the Ferrar diagram.

For each partition, let us consider its *conjugate*, whose Ferrar diagram is formed by reflecting the original diagram across the main diagonal. The following diagram shows a pair of conjugate partitions.



Observe that the conjugation operation switches the number of rows and columns in a Ferrar diagram. In particular, we get a bijection between the number of Ferrar diagrams with k rows and the number of Ferrar diagrams with k columns, thereby giving us a bijection between the set of partitions of n with exactly k parts and the set of partitions of n whose largest part is k .

By the way, the conjugation operation is known as an *involution*, since applying it twice gives us back the original element. Every involution is a bijection of a set with itself (why?).

Here is a very classic result about partitions originally discovered by Euler.

Problem 3. Prove that the number of partitions of n into distinct parts is equal to the number of partitions of n into odd parts.

We will construct a bijection between the set of partitions of n into distinct parts with the set of partitions of n into odd parts.

Starting from partition of n into distinct parts, let us write each part of n as $a \cdot 2^b$, where a is odd, and then split the part into 2^b parts all equal to a . This gives a partition of n into odd parts. For instance, starting from the partition $(12, 7, 6, 4, 1)$ of 30, we have

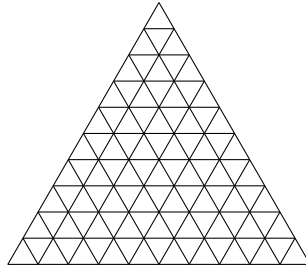
$$\begin{aligned} 30 &= 12 + 7 + 6 + 4 + 1 \\ &= 3 \cdot 2^2 + 7 \cdot 2^0 + 3 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^0 \\ &= (3 + 3 + 3 + 3) + 7 + (3 + 3) + (1 + 1 + 1 + 1) + 1 \\ &= 7 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 1 + 1 + 1 + 1 + 1. \end{aligned}$$

So we obtain the partition $(7, 3, 3, 3, 3, 3, 3, 1, 1, 1, 1, 1)$ into distinct parts. For the reverse direction, we start with a partition of n into distinct parts. Suppose that there are k parts equal to a , where a is odd, let $k = 2^{k_1} + 2^{k_2} + \dots + 2^{k_r}$ for distinct positive integers k_1, k_2, \dots, k_r (this is equivalent to the binary representation of k). Then create the parts $a \cdot 2^{k_i}$ for each i . Note that the results parts are all distinct, since every integer can be uniquely written as $a \cdot 2^b$ where a is odd. For instance, starting from the partition $(7, 3, 3, 3, 3, 3, 3, 1, 1, 1, 1, 1)$ of 30, we have

$$\begin{aligned} 30 &= 7 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 1 + 1 + 1 + 1 + 1 \\ &= 7 \cdot 1 + 3 \cdot 6 + 1 \cdot 5 \\ &= 7 \cdot 2^0 + 3 \cdot (2^2 + 2^1) + 1 \cdot (2^2 + 2^0) \\ &= 7 \cdot 2^0 + 3 \cdot 2^2 + 3 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^0 \\ &= 7 + 12 + 6 + 4 + 1 \\ &= 12 + 7 + 6 + 4 + 1. \end{aligned}$$

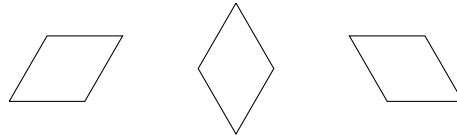
So we get back the partition $(12, 7, 6, 4, 1)$. You should convince yourself that the two procedures just described are inverses of each other. Thus we have shown that the number of partitions of n into distinct parts is equal to the number of partitions of n into odd parts.

Problem 4. A triangular grid is obtained by tiling an equilateral triangle of side length n by n^2 equilateral triangles of side length 1. Determine the number of parallelograms bounded by line segments of the grid.



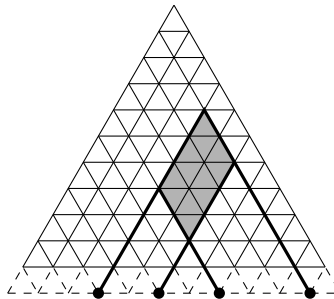
One approach is to carefully count all the parallelograms directly. But one has to be careful to not make mistakes, and we end up with a summation that has to be dealt with. We present a different, much more elegant solution using bijections.

Observe that the parallelograms come in three different orientations:



By symmetry, there is an equal number of parallelograms of each orientation, so let us just count the parallelograms with the middle type of orientation (i.e., no horizontal sides).

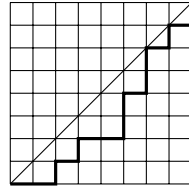
Extend the triangular grid by one extra row at the bottom. The key (and clever) observation is that starting from any such parallelogram in the original grid, we can extend its sides to meet the lines to meet the bottom edge of the new row in the large triangular grid, and there would be four distinct intersection points, as shown below.



Conversely, starting from any four distinct grid points in new bottom edge, we can extend 60° lines from the first two points and 120° lines from last two points to obtain a parallelogram in the original grid. This gives us a bijection between the set of parallelograms in the original grid with no horizontal sides with set of four distinct points in the new bottom edge, and hence there must be $\binom{n+2}{4}$ of them. Accounting for all three orientations, we find that the total number of parallelograms in the original grid is $3\binom{n+2}{4}$.

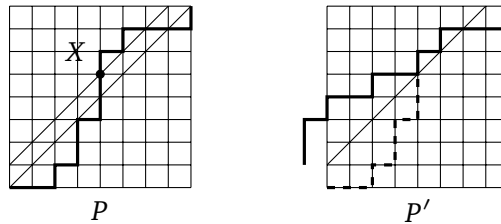
Next, let us explore the *Catalan numbers*, which is a sequence that comes up in many counting problems, which provide a rich source for bijections.

Problem 5. Let n be a positive integer. Determine the number of lattice paths from $(0, 0)$ to (n, n) using only unit up and right steps, such that the path stays in the region $x \geq y$.



We saw previously that the total number of lattice paths from $(0, 0)$ to (n, n) without the $x \geq y$ restriction is equal to $\binom{2n}{n}$. Let us count the number of paths that go into the $x < y$ region. Call these paths *bad paths*.

Suppose that P is a bad path. Since P goes into the region $x < y$, it must hit the line $y = x + 1$ at some point. Let X be the first point on the path P that lies on the line $y = x + 1$. Now, reflect the portion of path P up to X about the line $y = x + 1$, keeping the latter portion of P the same. This gives us a new path P' .



We claim that this gives us a bijection between the set of bad paths to the set of lattice paths from $(-1, 1)$ to (n, n) using only up and right unit steps.

Here is the inverse construction. For any lattice path Q from $(-1, 1)$ to (n, n) , let X be the first point on the path lying on the line $y = x + 1$, and let Q' be constructed from Q by reflecting the first portion of Q up to X through the line $y = x + 1$ and keeping the rest the same. Then the inverse of the bijection given above sends Q to Q' .

To complete the proof of this claim, we need to check a number of details, which we outline below. The reader should think about why claim is true.

- The inverse construction is well defined. That is, we can always find such a point X , and also, the resulting Q' is always a bad path.
- The two constructions are inverses of each other.

The number of bad paths is equal to the number of lattice paths from $(-1, 1)$ to (n, n) using only unit up and right steps, and there are $\binom{2n}{n+1}$ such paths (using again

Problem 1). Therefore, the total number of “good” paths, i.e., those that do not go into the region $x < y$, is equals to

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

This is our first example of something that is counted by the Catalan numbers. There are many more to come.

The n -th Catalan number is

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

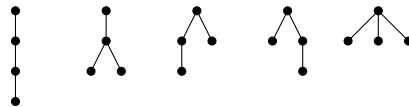
Problem 6. Show that the n -th Catalan number counts the number of expressions containing n pairs of parentheses which are correctly matched. E.g., for $n = 3$,

((())) ((()()) ()()()) ()()() ()()() ()()()

We could solve this problem by counting it using techniques similar to the one used to count paths above. A much quicker solution is to find a bijection between these parentheses expressions and the lattice paths counted in the previous problem. Indeed, note that by interpreting each (as a unit right step and each) as a unit up step, we obtain the desired bijection. The condition that the parentheses expression is correctly matched corresponds exactly to the condition that the lattice path do not go into the $x < y$ region (why?). This bijection shows that the number of expressions of n pairs of parentheses which are correctly matched is also equal to n -th Catalan number, as desired.

The previous bijection was rather simple. Let us look at a more involved Catalan number bijection.

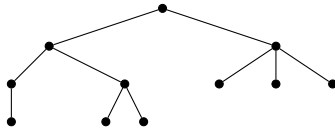
A *plane tree* is an object with the following structure. We start with a root vertex (drawn at the top), and then with each node we attach a number of new vertices (possibly none), where the order of the attached vertices matters. For instance, there are exactly 5 plane trees with 4 vertices:



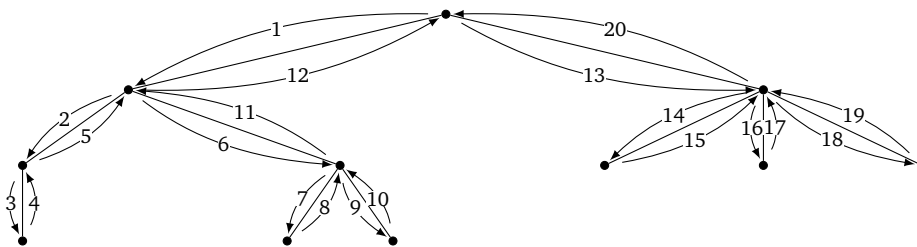
Problem 7. Show that the n -th Catalan number counts the number of plane trees with $n + 1$ vertices.

We will produce a bijection between plane trees and the parentheses expressions considered in the previous problem. We first describe an algorithm to turn a plane tree into a parentheses expression.

Given a plane tree, starting from the top vertex, let us perform a depth-first search walk, meaning that we go as further down as possible until we hit a dead-end, and then backtrack to a branch point, where we then explore a new branch. We will always explore the branches of a vertex in order from left to right. For instance, starting with the plane



we obtain the following walk, where the steps are labeled in order.



Now we record the sequence of steps we took, writing down a (each time we stepped downward along an edge, and a) every time we stepped upward along an edge. For example, the above walk corresponds to

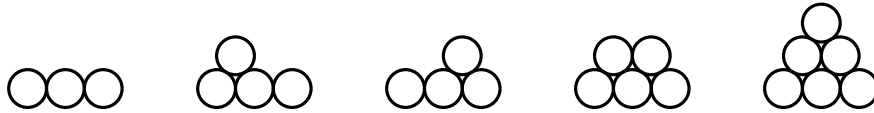
$$((()())(())(())(())).$$

A plane tree with $n + 1$ vertices always produces a correctly matched expression of n pairs of parentheses correctly matched (why is it correctly matched?). Conversely, given an expression of n pairs of correctly matched parentheses, it is possible to reverse this construction to produce a plane tree that corresponds to it. You should first convince yourself that this is the case why writing down a few parentheses expressions and then figure out what the corresponding tree is. Then, you should write down a description of this bijection.

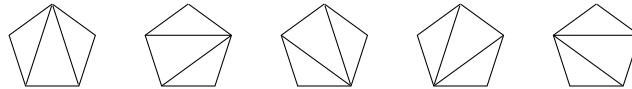
Practice problems:

1. Let n be a positive integer. In how many ways can one write a sum of at least two positive integers that add up to n ? Consider the same set of integers written in a different order as being different. (For example, there are 3 ways to express 3 as $3 = 1 + 1 + 1 = 2 + 1 = 1 + 2$.)
2. Let m, n be positive integers. Determine the number of m -tuples of positive integers (x_1, x_2, \dots, x_m) satisfying $x_1 + x_2 + \dots + x_m = n$.

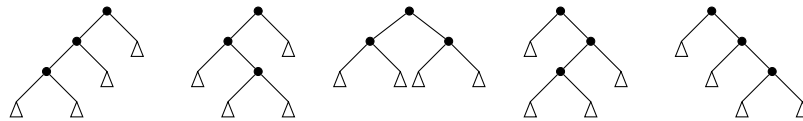
- Let m, n be positive integers. Determine the number of m -tuples of nonnegative integers (x_1, x_2, \dots, x_m) satisfying $x_1 + x_2 + \dots + x_m = n$. (Hint: can you find a bijection that reduces this problem to the previous problem?)
- Let n be a positive integer. Prove that the number of partitions of n equals the number of partitions of $2n$ with n parts.
- Show that the number of ways of stacking coins in the plane so that the bottom row consists of n consecutive coins is C_n . E.g., for $n = 3$,



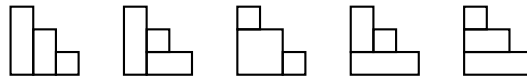
- Show that the number of triangulations of a convex $(n + 2)$ -gon into n triangles by $n - 1$ diagonals that do not intersect their interiors is the n -th Catalan number, C_n . E.g., for $n = 3$,



- Show that the number of complete binary trees with n internal vertices is the n -th Catalan number C_n . E.g., for $n = 3$,



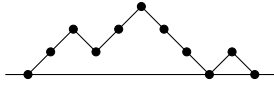
- Show that the number of ways to tile a stair-step shape of height n with n rectangles is the n -th Catalan number, C_n . E.g., for $n = 3$,



- Show that the number of ways that $n + 1$ factors can be completely parenthesized is the n -th Catalan number, C_n . E.g., for $n = 3$,

$$(((ab)c)d) \quad ((a(bc))d) \quad ((ab)(cd)) \quad (a((bc)d)) \quad (a(b(cd)))$$

- (Putnam 2003) A Dyck n -path is a lattice path of n upsteps $(1, 1)$ and n downsteps $(1, -1)$ that starts at the origin and never dips below the x -axis. A return is a maximal sequence of contiguous downsteps that terminates on the x -axis. For example, the Dyck 5-path illustrated has two returns, of length 3 and 1 respectively.



Show that there is a one-to-one correspondence between the Dyck n -paths with no return of even length and the Dyck $(n - 1)$ -paths.

11. (USAMO 1996) An n -term sequence (x_1, x_2, \dots, x_n) in which each term is either 0 or 1 is called a *binary sequence of length n* . Let a_n be the number of binary sequences of length n containing no three consecutive terms equal to 0, 1, 0 in that order. Let b_n be the number of binary sequences of length n that contain no four consecutive terms equal to 0, 0, 1, 1 or 1, 1, 0, 0 in that order. Prove that $b_{n+1} = 2a_n$ for all positive integers n .
12. Form a 2000×2002 screen with unit screens. Initially, there are more than 1999×2001 unit screens which are on. In any 2×2 screen, as soon as there are 3 unit screens which are off, the 4th screen turns off automatically. Prove that the whole screen can never be totally off.
13. (IMO 2002) Let n be a positive integer. Each point (x, y) in the plane, where x and y are non-negative integers with $x + y < n$, is colored red or blue, subject to the following condition: if a point (x, y) is red, then so are all points (x', y') with $x' \leq x$ and $y' \leq y$. Let A be the number of ways to choose n blue points with distinct x -coordinates, and let B be the number of ways to choose n blue points with distinct y -coordinates. Prove that $A = B$.