UNIVERSITY OF GHANA, LEGON
DEPARTMENT OF MATHEMATICS

Catalan Numbers

By

Sebastian Etornam Ladzekpo
(10886672)

December, 2024

A DISSERTATION TO THE DEPARTMENT OF MATHEMATICS, UNIVERSITY OF GHANA, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR BACHELOR OF ARTS DEGREE.

DECLARATION

I declare that this dissertation is my original work. Except where due acknowledgement has been provided, this dissertation contains no material previously published by any other individual. To the best of my knowledge, there is no content in this dissertation that has been accepted as part of any other academic degree requirements at the University of Ghana or any other university.

This dissertation was carried out in the Department of Mathematics, University of Ghana in partial fulfilment of a Bachelor of Science Degree under the supervision of Dr. Kenneth Dadedzi.

Student: Sebastian Etornam Ladzekpo

Signature:

Date:

Supervisor: Dr. Kenneth Dadedzi

Signature:

Date:

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to my supervisor - Dr. Kenneth Dadedzi , whose unwavering support and dedication have been instrumental in my understanding of Catalan numbers. His hard work, patience, and relentless effort to guide me through the complexities of the subject have been invaluable. His deep knowledge and willingness to break down intricate concepts into comprehensible elements made even the most challenging aspects of this project achievable. I truly appreciate the time and energy he invested in ensuring that I grasp the full depth and breadth of the topic, and for always encouraging me to explore further.

DEDICATION

I dedicate this work to Dr. Ruby Appiah-Campbell and my lovely mother Madam Peace Anyanah.

Abstract

When counting the number of ways a hiker can move from point A to point B given that he can only move North and East or the number of ways a set of operators can be arranged back into an equation, the answer often lies in combinatorial mathematics. Catalan numbers, a sequence of natural numbers that counts various combinatorial structures, provides solutions to problems like this. The focus of this research is in twofold: to explore the recursive nature of Catalan numbers and to analyze their representation through generating functions. By investigating these, we gain insight into how Catalan numbers emerge in diverse combinatorial structures.

Key structures examined in this study include valid parenthesization, the triangulation of convex polygons, and Dyck paths, each of which demonstrates a natural correspondence with Catalan numbers. Through a detailed analysis of these structures, the recursive properties of Catalan numbers are illustrated, alongside their combinatorial interpretations. Generating functions are employed to provide efficient means of deriving these numbers, offering a deeper understanding of their underlying patterns and relationships.

The results presented in this project underscore the usefulness of Catalan numbers in solving a wide range of mathematical problems, from counting parenthesis to evaluating Dyck paths. By highlighting their theoretical significance, this research contributes to a more comprehensive understanding of Catalan numbers and their role in combinatorial mathematics.

Contents

Chapter 1

Introduction

In this chapter, we introduce the Catalan numbers. We talk about the history of how the Catalan sequence came about, highlighting the works of some mathematicians. We move on to the motivation, problem statement and the nature of the remaining chapters.

1.1 Catalan Numbers History and Motivation

Catalan numbers form an interesting and important sequence in mathematics, with wide-ranging applications across combinatorics, geometry, algebra, and computer science. Catalan numbers is a sequence of positive integers $\mathbb{Z}$. An example of such application is the number of ways to place n indistinguishable balls in n numbered boxes $B_1, ..., B_n$ such that at most a total of k balls are placed in boxes $B_1, ..., B_k$, for $k = 1, ..., n$. According to the Online Encyclopedia of Integers Sequence (O.E.I.S), there are 256 results where the Catalan numbers appear as a solution to combinatorial problems.

A Belgian mathematician Eugene C. Catalan discovered Catalan numbers in 1838, while studying well-formed sequences of parentheses. Although they are named after Catalan, they were not first discovered by him. Around 1751, Euler found them while studying the triangulations of convex polygons. However, according to a 1988 article by Chinese mathematician J. J. Luo, Chinese mathematician Antu Ming discovered them about 1730 through his geometric models. Ming's work was published in Chinese, so it was not known in the West [3]. The real reason for the Catalan number's name is that it was Eugene Charles Catalan, who was the first to actually describe these numbers as a sequence and give a well-defined formula for the $n - th$ Catalan number.

Catalan numbers appear in a range of counting problems hence the need to study these structures from both a combinatorial and geometric view. We will analyze how these structures relate to Catalan numbers through recursive relations. By studying these relationships and with the use of generating functions we can derive a closed form for the Catalan numbers, providing a unified understanding of their behavior across different structures.

This project explores some mathematical structures that give rise to Catalan numbers. We will study these structures using some mathematical tools and establish the relationship among them. Specifically, we will focus on parenthesization, the triangulation of convex polygons, and Dyck paths.

1.2 Structure Of The Project

The main aim of this work has been introduced in Chapter 1. We will introduce the Catalan numbers further in Chapter 2. In this same chapter, we will also introduce some definitions, notations, and theorems relevant to this work. Chapter 3 delves into how a string of parenthesis count the Catalan numbers. In this chapter, a more detailed work is observed. The chapter covers the definition of a valid string of parenthesis, how the recurrence relation is gotten from the parenthesization and with the use of generating functions we come out with a closed form of deriving any term of the Catalan sequence. In Chapter 4 we study two other structures, that is triangulation of convex polygons and Dyck paths. We establish a combinatorial and geometric relation between these structures and parenthesization which goes to prove that the same process can be used to arrive at the closed form of the Catalan sequence.

Chapter 2

Basic Notions

2.1 Introduction

In this chapter, we will talk in detail about the Catalan sequence, looking at the sequence and a typical representation of the recurrence. We also delve deeper by studying generating functions in relation to using it to solve the Catalan recurrence relation to get a closed form for the Catalan sequence. In the last part of this chapter we take a look at the definitions of some mathematical terms and theorems that would later be used in Chapter 3 and Chapter 4 of this project.

2.2 Calatan Numbers

Catalan numbers emerge in a wide variety of combinatorial structures, each representing a specific counting problem. Solving these counting problems gives rise to the Catalan sequence. A brief explanation of some of these counting problems like, valid parenthesizations of expressions, the triangulation of convex polygons, and Dyck paths are given here to help us establish some concepts in this chapter, these definitions are found in [5]. A more detailed study of these structures is made in Chapter 3 and Chapter 4.

- **Parenthesization:** One of the earliest applications of Catalan numbers involves counting the number of ways to fully parenthesize a sequence of operations. Given a set of binary operations, the number of distinct ways to group them with parentheses corresponds to a Catalan number.

- **Triangulation of Convex Polygons:** This problem involves counting the number of ways to triangulate a convex polygon. For a polygon with $n+2$ vertices, the number of distinct triangulations is given by the nth Catalan number.

- **Dyck Paths:** A Dyck path is a staircase walk from the origin to a specific point on a Cartesian plane, constrained to remain above the diagonal line. The number of such paths for a given number of steps also corresponds to a Catalan number. Dyck paths provide a useful geometric and visual representation of these numbers, illustrating their recursive nature.

Definition 2.2.1 (Catalan Sequence,[4]). The Catalan sequence is an infinite sequence of positive integers given by $1, 1, 2, 5, 14, 42, \ldots$. Let C_n to represent the $n - th$ Catalan number. We have that $C_0 = 1, C_1 = 1, C_2 = 2, C_3 = 5, \ldots$.

Table 2.1: Catalan Numbers

n	Recursive Nature	C_n
0	-	1
1	C_0C_0	1
2	$C_0C_1 + C_1C_0$	2
3	$C_0C_2 + C_1C_1 + C_2C_0$	5
4	$C_0C_3 + C_1C_2 + C_2C_1 + C_3C_0$	14
5	$C_0C_4 + C_1C_3 + C_2C_2 + C_3C_1 + C_4C_0$	42
$\vdots$	$\vdots$	$\vdots$

As n grows large, we observed that the Catalan numbers grow quickly, following an asymptotic formula: $C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$ [4]. Table 2.1 draws our attention to the recursive relation of the Catalan sequence, hence our next topic for discussion.

Definition 2.2.2 (Recursive Nature of Catalan Numbers,[6]). Recursion in combinatorics refers to a method of defining sequences, sets, or other combinatorial objects in terms of themselves, that is each term is defined in terms of previous terms. A recursive definition typically consists of two parts:

- **Base case:** A direct definition for the initial element(s) of the sequence or set.
- **Recursive step:** A rule that defines each subsequent element in terms of the previous ones.

A sequence is called the solution of a recurrence relation if the terms of this sequence satisfies the relation.

From Table 2.1, we clearly observe that the Catalan numbers is also a solution to the recursion.

$$C_0 = 1,$$

$$C_n = \sum_{i=0}^{n-1} C_i C_{n-1-i}, \quad \text{for } n \geq 1.$$

Catalan numbers count many combinatorial structures. We will explore three of such structures. Later, we will see that the closed form for the recursion is given by:

$$C_n = \frac{1}{n+1}\binom{2n}{n}, \quad \text{for } n \geq 0.$$

We will prove this formula by using combinatorial arguments, later in Chapter 3 by considering the set of strings of n pairs of parentheses. The number of valid strings of parenthesis corresponds to the Catalan number C_n. A set of parenthesis is valid if it never has more right parentheses than left parentheses at any point.

Recursion is a powerful tool in combinatorics, providing a method to define and analyze sequences and sets. By using combinatorial arguments and arriving at the recurrence relation, we employ generating functions to simplify further helping us get the closed form of the Catalan sequence. In the next section we take a look at generating functions.

2.3 Generating Functions

Generating functions are an essential tool in combinatorics, particularly in the study of sequences. They allow us to encode sequences as coefficients in a formal power series, providing a bridge between discrete and continuous mathematics (In an informal sense, generating functions help transform problems about sequences into problems about functions).

Definition 2.3.1 (Ordinary Generating Function,[2]). Given a sequence $\{a_n\}$ with $n \geq 0$, the ordinary generating function $G(x)$ for the sequence is typically defined as:

$$G(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Hence, $G(x)$ is a polynomial if a_n is finite and a power series when a_n is infinite.

Not all generating functions are ordinary, but those are the ones considered in this study. Here, x is a formal variable (Regard x as a place holder rather than a number. So we generally ignore the issue of convergence), and the coefficients a_n correspond to the terms of the sequence. Generating functions enable the analysis of sequences through algebraic manipulation and can be used to solve recurrence relations, find closed-form expressions, and understand deeper combinatorial properties. The way of finding the closed-form of the Catalan sequence would be treated in more details in Chapter 3. We take a look at some basic applications of generating functions by considering these examples.

Example 2.3.2. Define a sequence, $\{a_n\}_{n\geq 0} = \{1, 1, 1, 1, 1, \ldots\}$ by definition,

$$G(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Hence,

$$\begin{aligned} G(x) &= \sum_{n=0}^{\infty} x^n \\ G(x) &= 1 + x + x^2 + x^3 + \cdots \\ G(x) - 1 &= x + x^2 + x^3 + x^4 + \cdots \\ G(x) - 1 &= x(1 + x + x^2 + x^3 + \cdots) \\ G(x) - 1 &= xG(x) \\ G(x) &= \frac{1}{1-x}. \end{aligned}$$

Example 2.3.3. Define a sequence, $\{a_n\}_{n\geq 0} = \{1,0,1,0,1,0,\cdots\}$ by definition,

$$G(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Hence,

$$G(x) = 1 + x^2 + x^4 + x^6 + \cdots$$
$$G(x) - 1 = x^2 + x^4 + x^6 + \cdots$$
$$G(x) - 1 = x^2(1 + x^2 + x^4 + x^6 + \cdots)$$
$$G(x) = 1 + x^2 G(x)$$
$$G(x) = \frac{1}{1-x^2}.$$

Theorem 2.3.4 (Properties of Ordinary Generating Functions,[5]). *Having two generating functions defined as,*

$$A(x) = \sum_{n=0}^{\infty} a_n x^n \text{ and } B(x) = \sum_{n=0}^{\infty} b_n x^n$$

where $\{a_n\} = \{a_0, a_1, a_2, \ldots\}$ *also* $\{b_n\} = \{b_0, b_1, b_2, \ldots\}$ *with* $n \geq 0$ *are the respective sequences of the ordinary generating functions* $A(x)$ *and* $B(x)$*, then*

1. *the sequence* $c_n = a_n + b_n$ *has generating function* $C(x) = A(x) + B(x)$*,*
2. *the sequence* $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ *has generating function* $C(x) = A(x) \cdot B(x)$*,*
3. *the sequence* $c_n = \alpha a_n$ *has generating function* $C(x) = \alpha A(x)$ *where* α *is any constant,*
4. *the sequence* $c_n = a_{n+m}$ *has generating function* $C(x) = \frac{A(x) - \sum_{n=0}^{m-1} a_n x^n}{x^m}$*,*
5. *the sequence* $c_n = n a_n$ *has generating function* $C(x) = xA'(x)$*,*
6. *the sequence* $c_n = \sum_{k=0}^{n} a_k$ *has generating function* $C(x) = \frac{A(x)}{1-x}$*.*

$Proof:$ *Each of the statements can be obtained by simple arithmetic:*

1.
$$C(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n = \sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} b_n x^n.$$

2.
$$C(x) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k} x^n = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} a_k b_{n-k} x^n = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} a_k x^k b_{n-k} x^{n-k}$$
$$= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_k x^k b_l x^l = \left(\sum_{k=0}^{\infty} a_k x^k\right)\left(\sum_{l=0}^{\infty} b_l x^l\right) = A(x) \cdot B(x).$$

3.

$$C(x) = \sum_{n=0}^{\infty} \alpha a_n x^n = \alpha \sum_{n=0}^{\infty} a_n x^n = \alpha A(x).$$

4.

$$C(x) = \sum_{n=0}^{\infty} a_{n+m} x^n = \sum_{n=m}^{\infty} a_n x^{n-m} = x^{-m} \sum_{n=m}^{\infty} a_n x^n = \frac{A(x) - \sum_{n=0}^{m-1} a_n x^n}{x^m}.$$

5.

$$C(x) = \sum_{n=0}^{\infty} n a_n x^n = x \sum_{n=0}^{\infty} n a_n x^{n-1} = x \sum_{n=0}^{\infty} a_n \frac{d}{dx} x^n = x \frac{d}{dx} \sum_{n=0}^{\infty} a_n x^n = x A'(x).$$

6.

$$C(x) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k x^n = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} a_k x^n = \sum_{k=0}^{\infty} a_k \sum_{n=0}^{\infty} x^{n+k}$$
$$\sum_{k=0}^{\infty} a_k x^k \sum_{n=0}^{\infty} x^n = \sum_{k=0}^{\infty} a_k x^k \cdot \frac{1}{1-x} = \frac{A(x)}{1-x}.$$

Example 2.3.5. For the sequence $\{a_n\} = \{1, 1, 1, 1, \cdots\}$ and $\{b_n\} = \{1, 0, 1, 0, \cdots\}$ with ordinary generating function

$$A(x) = 1 + x + x^2 + x^3 + \cdots = \sum_{n=0}^{\infty} a_n x^n = \frac{1}{1-x}$$

and

$$B(x) = 1 + x^2 + x^4 + x^6 + \cdots = \sum_{n=0}^{\infty} b_n x^n = \frac{1}{1-x^2}$$

respectfully would sum up to

$$C(x) = A(x) + B(x)$$
$$C(x) = \sum_{n=0}^{\infty} a_n x^n + \sum_{n=0}^{\infty} b_n x^n$$
$$C(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n$$
$$C(x) = \frac{1}{1-x} + \frac{1}{1-x^2}.$$

Lets consider some important theorems in combinatorics that will be relevant to this work.

Theorem 2.3.6 (Binomial Coefficient, [2]). *The number of possible ways to choose a subset of k elements from a set of n elements (the order is not relevant) is given by*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Remark 2.3.7. For $n < k$ or $k < 0$, we set the binomial coefficient to 0. However, it is important to note that we can define $\binom{\alpha}{k}$ for any real number α and integer k as

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}.$$

The generating function for the Catalan numbers, $C(x)$, is given by the following equation:

$$C(x) = \sum_{n=0}^{\infty} C_n x^n.$$

This generating function not only entail the entire sequence but also provides a direct method for computing individual Catalan numbers. By solving the generating function, we can arrive at the well-known closed form for the nth Catalan number:

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

Theorem 2.3.8 (Binomial Theorem, [2])**.** *For any integer $n \geq 0$, we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Remark 2.3.9. We have the binomial series

$$(1+x)^{\alpha} = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k.$$

This is the Taylor series of $(1+x)^{\alpha}$ at $x = 0$.

2.4 Geometric Similarities in Structures Counting Catalan Numbers

Beyond their combinatorial interpretations, the structures that give rise to Catalan numbers also share some geometric similarities. The problem of counting valid parenthesizations, triangulating convex polygons, and enumerating Dyck paths can all be visualized in terms of plane structures. Each structure involves a form of partitioning or subdivision, whether it is splitting parentheses, dividing a polygon into triangles, or dividing a plane into distinct paths. This makes it possible to connect all the structures, also known as the bijection between the structures [3]. These geometric similarities highlight the underlying unity of the different counting problems associated with Catalan numbers.

Moreover, these structures exhibit a form of recursive self-similarity, where smaller versions of the structure are nested within larger versions. This recursive property, reflected both in the recurrence relation and in the visual patterns of these structures, further emphasizes the central role of Catalan numbers in organizing combinatorial and geometric objects.

Chapter 3

Catalan Number and Parenthesization

3.1 Introduction

One of the insightful applications of Catalan numbers is found in the study of parenthesization. Parenthesization refers to the different ways of grouping expressions using parentheses, and it serves as a powerful combinatorial tool. Imagine you have a sequence of elements that need to be multiplied in different orders. The question of how many distinct ways you can fully parenthesize the product without changing the order of the elements is a classic problem, whose solution is given by the Catalan numbers.

This chapter delves into the combinatorial structure of parenthesization, exploring its properties and connection to Catalan numbers. We begin by defining the problem of parenthesization in a mathematical framework, followed by a detailed enumeration of the valid parenthesizations for a given number of elements. Through examples and step-by-step constructions, we will study the recursive nature of the Catalan numbers and demonstrate how they solve the parenthesization problem.

In conclusion, our study will end with the derivation of the explicit formula for the Catalan sequence.

3.2 Parenthesization

One of the structures counted by Catalan numbers is "Parenthesis". In counting this structure, an understanding of what a valid string of parenthesis must first be established.

Definition 3.2.1 (Valid Parenthesis, [4]). What is meant by a valid parenthesis is that to every open parenthesis, there has to be a corresponding closed parenthesis. That is, '$()'$, '$(())'$, '$(()())'$. Also a group of valid parenthesis can either be enclosed in each other $(())$ or stacked together $()()$.

Theorem 3.2.2 ([4]). *The number of well-formed sequences of parentheses of length $2n$ is the Catalan number C_n.*

The rest of this Chapter, will be focused on establishing the proof of Theorem 3.2.2.

We will let S_n represent the number of ways of arranging an $n - string$ of valid parenthesis. The aim is to show that $S_n = C_n, n \geq 0$.

Now, for $n = 0$, we set $S_0 = 1$, since there is only one way to arrange $0 - string$ of valid parenthesis. That is do nothing.

We have that

Table 3.1: Parenthesis

n	Valid Parenthesis	S_n
0		1
1	$()$	1
2	$()(), (())$	2
$\vdots$	$\vdots$	$\vdots$

Lets now explore the geometric structure of valid parenthesis to deduce a recursion formula.

To get a valid parenthesis with $n - pairs$ of strings, we do the following

a. enclose a valid parenthesis with $k - pairs$ of strings, $k < n$ and

b. attach a valid parenthesis with $(n - k - 1) - pairs$ of strings.

Let P_n represent a valid parenthesis with $n - pairs$ of strings. Then we have

$$P_n = (P_k)P_{n-k-1}, 0 \leq k < n.$$

Clearly, a valid parenthesis with $n - strings$ can be obtained from valid parenthesis with $n = 0, 1, 2, 3, \dots, n - 1$ strings. This leads us to the recursive formula;

$$S_n = \sum_{k=0}^{n-1} S_k S_{n-1-k}, \text{ where } S_0 = 1. \tag{3.2.1}$$

Example 3.2.3. Let us use the valid parenthesis in Table 3.1 and the recursive formula to find all valid parenthesis with three pairs of strings. We have that ;

1. Enclose all parenthesis with $k = 0$ string and attach all parenthesis with $k = 2$ strings to get ()()() and ()(())
2. Enclose all parenthesis with $k = 1$ string and attach all parenthesis with $k = 1$ strings to get (())()
3. Enclose all valid parenthesis with $k = 2$ strings and attach all parenthesis with $k = 0$ strings to get (()()) and ((())).

This gives us 5 valid parenthesis with 3 pairs of strings.

Clearly, we see that $S_n = C_n$. Therefore, the recurrence formula for the $n-th$ Catalan number is given by

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-k-1}, C_0 = 1. \tag{3.2.2}$$

Example 3.2.4. Find C_7

$$\begin{aligned}C_7 &= C_0 \cdot C_6 + C_1 \cdot C_5 + C_2 \cdot C_4 + C_3 \cdot C_3 + C_4 \cdot C_2 + C_5 \cdot C_1 + C_6 \cdot C_0 \\ &= 132 + 42 + 28 + 25 + 28 + 42 + 132 = 429\end{aligned}$$

3.3 Explicit formula for the n-th Catalan number.

As n gets large, it becomes difficult to compute $n-th$ Catalan number using the recurrence formula (3.2.2). There is therefore the need to find an explicit formula for the $n-th$ Catalan number. To do so, we make use of the generating functions.

Let $C(x)$ be the ordinary generating function for the Catalan numbers. We have that

$$\begin{aligned}C(x) &= C_0 + C_1 x + C_2 x^2 + C_3 x^3 + C_4 x^4 + \cdots \\ &= \sum_{n=0}^{\infty} C_n x^n.\end{aligned}$$

By Theorem 2.3.4, we have that

$$\begin{aligned}&C(x)^2 = C_0C_0 + [C_0C_1 + C_1C_0]\,x + [C_2C_0 + C_1C_1 + C_0C_2]\,x^2 + \cdots C_1 + C_2x + C_3x^2 + \cdots \\ &Hence: \\ &xC(x)^2 = xC_1 + C_2x^2 + C_3x^3 + \cdots \\ &xC(x)^2 = C(x) - C_0 \\ &xC(x)^2 = C(x) - 1 \\ &\quad C(x) = \frac{1 \pm \sqrt{1-4x}}{2x}.\end{aligned}$$

We know $C(0) = C_0 = 1$. So we consider

$$\lim_{x\to 0} \frac{1-(1-4x)^{\frac{1}{2}}}{2x} = \frac{\frac{-1}{2}(1-4x)^{\frac{-1}{2}}(-4)}{2} = 1$$

and

$$\lim_{x\to 0} \frac{1+(1-4x)^{\frac{1}{2}}}{2x} = \frac{\frac{1}{2}(1-4x)^{\frac{-1}{2}}(-4)}{2} = -1.$$

Since the Catalan numbers consist of non-negative numbers only, we consider

$$C(x) = \frac{1-\sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty} C_n x^n.$$

Lets apply the binomial series to find C_n.

We have

$$1 - 2xC(x) = (1-4x)^{\frac{1}{2}}$$

$$1 - 2x\sum_{n=0}^{\infty} C_n x^n = \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (-4)^k x^k$$

$$1 - 2\sum_{n=0}^{\infty} C_n x^{n+1} = \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (-4)^k x^k$$

Considering the coefficients of x^{n+1} to get

$$\begin{aligned}
-2C_n &= \binom{\frac{1}{2}}{n+1}(-4)^{n+1} \\
&= (-4)^{n+1}\left[\frac{(\frac{1}{2})\cdot(\frac{1}{2}-1)\cdots(\frac{1}{2}-n)}{(n+1)!}\right](-1)^{n+1}(2^2)^{n+1}\left[\frac{(\frac{1}{2})\cdot(\frac{-1}{2})\cdot(\frac{-3}{2})\cdot(\frac{-5}{2})\cdots(\frac{-(2n-1)}{2})}{(n+1)!}\right] \\
&= (-1)^{n+1}(2)^{2n+2}\left[\frac{(1)\cdot(-1)\cdot(-3)\cdot(-5)\cdots(-(2n-1))}{2^{n+1}(n+1)!}\right] \\
-2C_n &= (-1)^{n+1}(2)^{2n+2}\cdot(-1)^n\left[\frac{(1)\cdot(1)\cdot(3)\cdot(5)\cdots(2n-1)}{2^{n+1}(n+1)!}\right] \\
-2C_n &= (-1)^{2n+1}(2)^{n+1}\left[\frac{(1)\cdot(3)\cdot(5)\cdots(2n-1)\times(2\cdot4\cdot6\cdot8\cdots2n)}{(n+1)!(2\cdot4\cdot6\cdot8\cdots2n)}\right] \\
C_n &= \frac{-1\cdot2^{n+1}}{-2}\left[\frac{(2n)!}{(n+1)!2^n(n)!}\right] \\
C_n &= \frac{2n!}{(n+1)\cdot n!\cdot n!} \\
&= \left(\frac{1}{n+1}\right)\left(\frac{2n!}{n!\cdot n!}\right) \\
&= \left(\frac{1}{n+1}\right)\binom{2n}{n}.
\end{aligned}$$

Example 3.3.1. Find the values of $C_1, C_2, C_3, andC_{10}$, using the formula for C_n

$$C_1 = \frac{1}{(1+1)}\binom{2}{1} = \frac{1}{(2)}(2) = 1,$$

$$C_2 = \frac{1}{(2+1)}\binom{4}{2} = \frac{1}{(3)}(6) = 2,$$

$$C_3 = \frac{1}{(3+1)}\binom{6}{3} = \frac{1}{(4)}(20) = 5,$$

$$C_{10} = \frac{1}{(10+1)}\binom{20}{10} = \frac{1}{(11)}(184756) = 16796.$$

The primary focus was understanding how the structure of parenthesization directly counts Catalan numbers. By defining valid sequences of nested parentheses, we highlighted that the number of ways to correctly parenthesize an expression with $(n+1)$ factors is precisely the $(n)-th$ Catalan number, (C_n).

Lets now explore other structures that are counted by the Catalan numbers. The main tool is to find a bijection between these structures and the number of ways to validly parenthesize.

Chapter 4

Bijection Between Parenthesization and Other Mathematical Structures

4.1 Introduction

In this chapter, we will talk about the combinatorial and geometric similarities between parenthesization and both Dyck paths and triangulation of convex polygons. We establish that these structures are related geometrically and from a combinatorial stand point, hence arriving at the same recursive relation. This implies that we can use the generating function to solve them as demonstrated in Chapter 3.

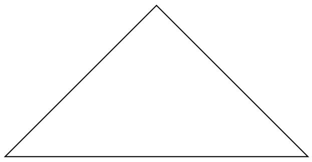
4.2 Triangulation of $(N+2)$-gon

For an $(n+2)$-gon, the task involves dividing the polygon into triangles using its diagonals, such that no diagonals intersect within the polygon. The number of ways to triangulate an $(n+2)$-gon is given by the n-th Catalan number. This section aims to develop an understanding on the concept of triangulating a polygon and the conditions that make a triangulation valid, explore different methods to count the number of triangulations, starting from smaller polygons and building up to the general case, demonstrating how the triangulation problem relates to Catalan numbers and derive the recursion formula for the number of triangulations. Also this section focuses on the bijection between the triangulations of $(n+2)$- gon and Parenthesization.

Definition 4.2.1 (Valid Triangulation,[1])**.** Triangulating an $(n+2)$-gon is the number of ways of drawing non-intersecting diagonals so that the interior of the $(n+2)$-gon is partitioned into n triangles.

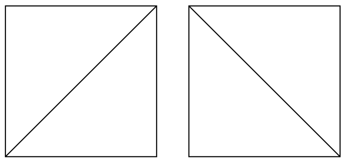
Let T_n represent the number of possible triangulations of an $(n+2)-gon$.

for $n=1$, we have the possible triangulations of a (3)-gon(triangle) is



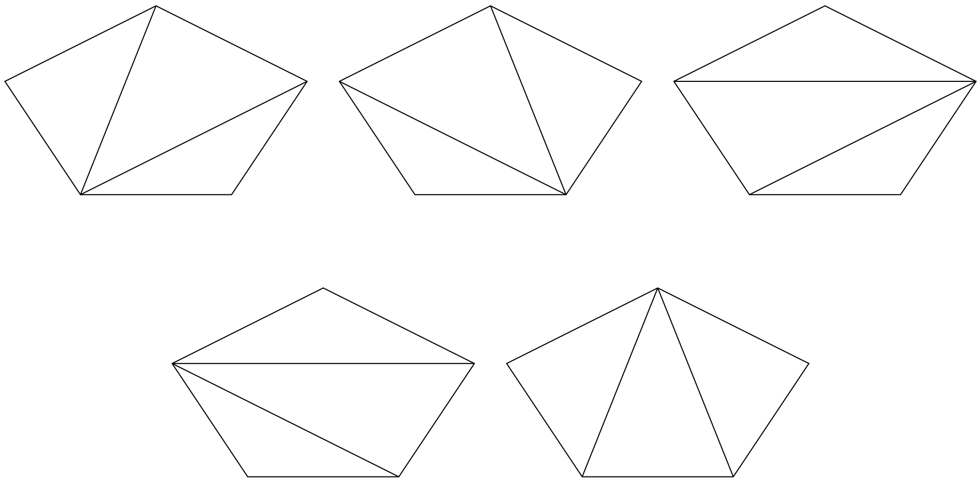
Hence $T_1 = 1$

For $n = 2$, we have



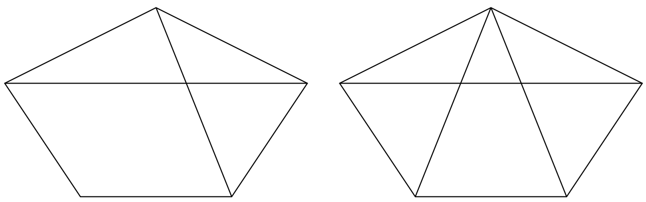
Hence $T_2 = 2$.

For $n = 3$, we have



Hence $T_3 = 5$.

Not-valid triangulations includes



To triangulate an $(n+2)$- gon, we first find a side as a point of reference. Label the vertices as depicted in Figure 4.1 .Choose any other vertex (label K) which is not on the base side and join it to the vertices on the base side to form a triangle. The triangle formed partitions the $(n+2)$-gon into two main parts, namely part A and part B as depicted in Figure 4.1. Part A will contain any possible triangulation of $(k+2)$-gon and part B will contain any possible triangulation of $(n-k+1)$-gon

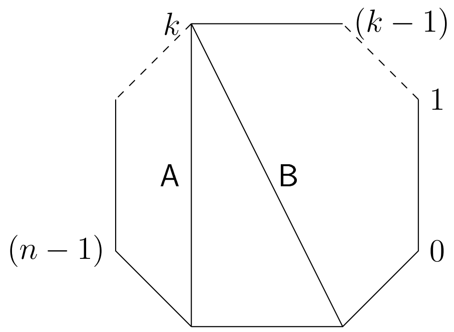
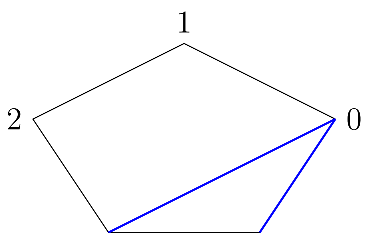


Figure 4.1: Triangulation of (n+2)-gon.

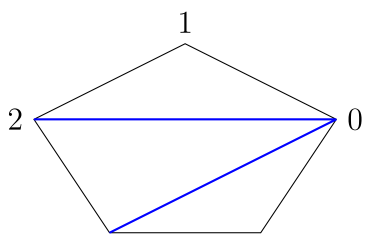
Summing all possible triangulations leads to the recurrence relation

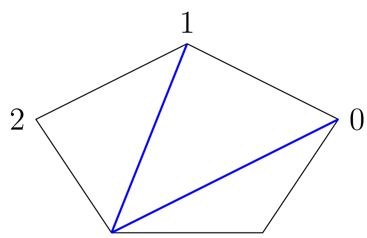
$$T_n = \sum_{k=0}^{n-1} T_k T_{n-k-1}, \text{ where } T_0 = 0$$

Example 4.2.2. $n = 3$ and $k = 0$

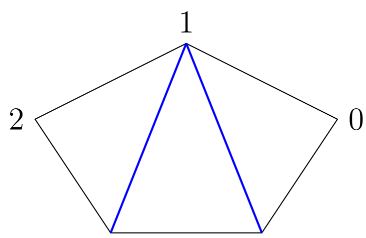


We have

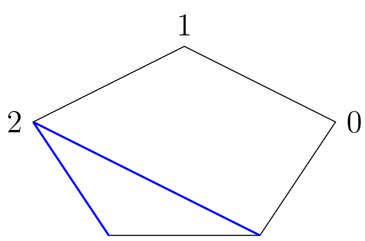




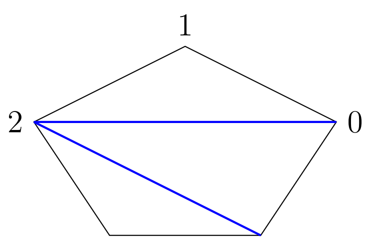
When $k = 1$, we have

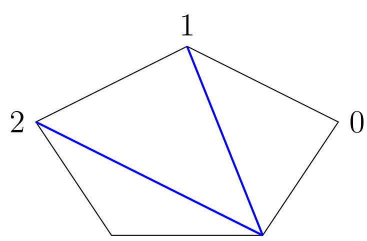


When $k = 2$



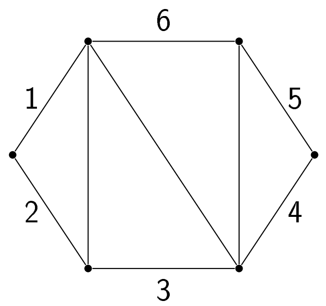
We have

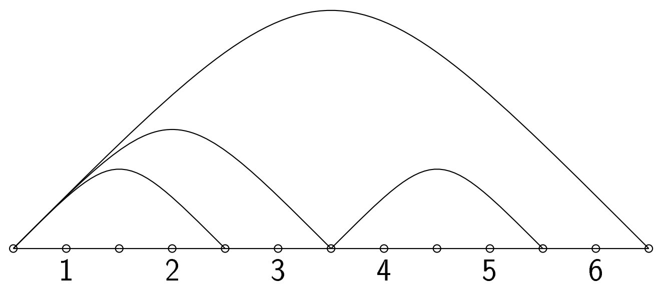




Clearly $T_n = S_n = C_n$

4.3 Bijection Between Parenthesization and Triangulation of $(n+2)$-gons





To establish the bijection, we consider $[(4+2)-gon$ where $n = 4$ and number of way of counting is $C_4]$ and the remaining sides are opened up to lie flat. The internal line segments of the shape that divides the internals into non-intersecting triangles becomes arcs.

The beginning of an arc is an open parenthesis and the end of the arc is a closed parenthesis. An arc within an arc is a valid string of parenthesis contained within another valid string of parenthesis. Note that, from the above diagrams the flat out line segments with arcs yields a parenthesization of $((())())$, which is a valid string of parenthesis for $n = 4$.

It is seen that both parenthesization and triangulations of $(n+2)$-gons have the same recursion relation. Here a structural relationship between the two structures is established. It shows they both count the sequence of Catalan numbers (C_n) and therefore the same procedure for deriving a closed formula for C_n as demonstrated for parenthesization in Chapter 3 applies to the triangulation of $(n+2)$-gons.

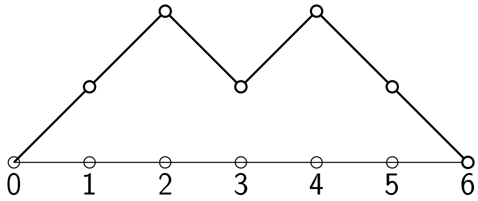
4.4 Dyck Path

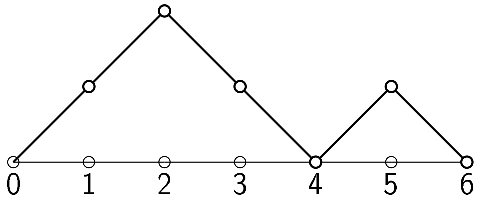
One representation of Catalan numbers is through Dyck paths. In this section, we take a look at Dyck paths, talking about definitions, properties, and significance within the broader context of Catalan numbers. We will study their combinatorial structure, understanding how these paths illustrate the counting principles underlying Catalan numbers. Through step-by-step examples, we will demonstrate the construction of Dyck paths, highlighting their recursive nature and the bijection with parenthesizations.

Definition 4.4.1 (Valid Dyck Paths, [3])**.** A Dyck path is a staircase walk on a grid that starts at the origin $(0, 0)$, ends at $(2n, 0)$, and consists of steps either upwards or downwards and never dipping below the horizontal axis $(y = 0)$.

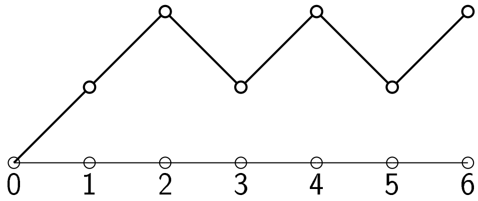
A valid Dyck path satisfies the additional constraint of remaining on or above the horizontal axis at all points along the path. This means for any point (x, y) along the path, $y \geq 0$.

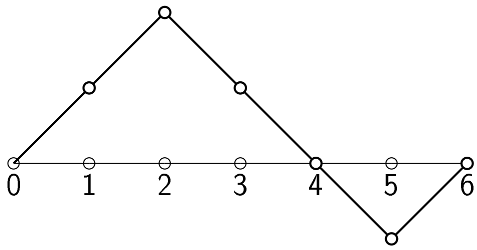
$Valid \quad Dyck \quad Paths$





$Not-Valid \quad Dyck \quad Paths$

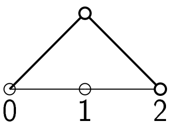




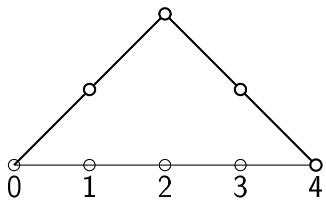
4.5 Deriving The Recursion

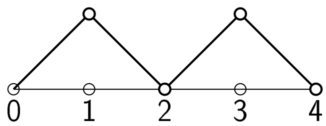
We will let D_n denote the number of Dyck paths with $2n$ steps. We have that

when $n = 1$ then $D_n = 1$, thus

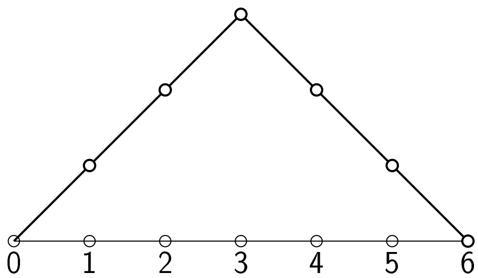


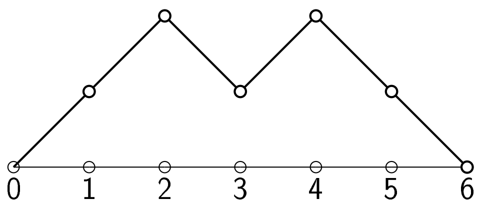
when $n = 2$ then $D_n = 2$, thus

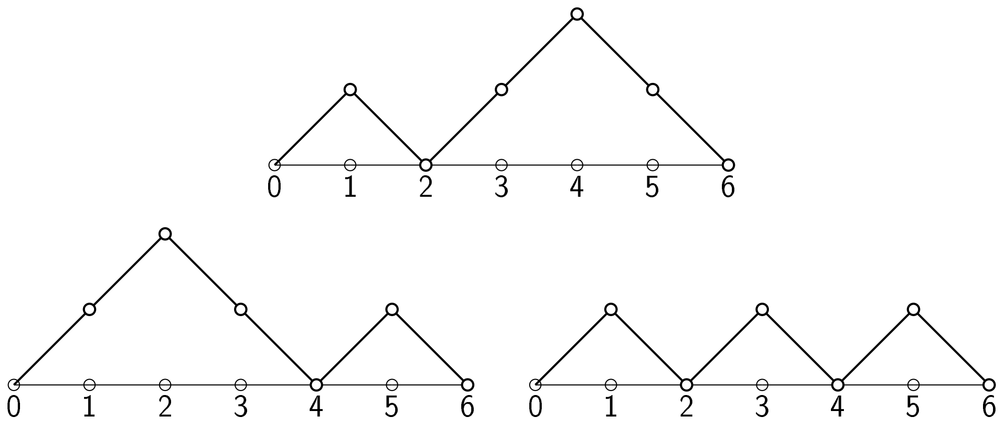




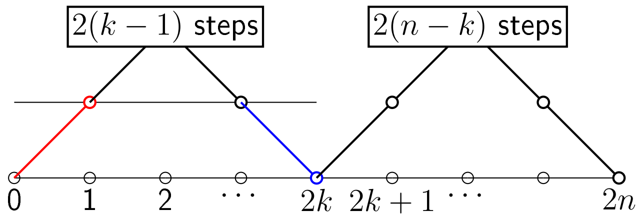
when $n = 3$ then $D_n = 5$, thus







Let us consider a valid Dyck path with $2n$ steps. We focus on the first time the Dyck path touches the $x - axis$. let $(2k, 0)$ be that point. Note that $0 \leq k \leq n$. We have that the left side of the point consist of a Dyck path with $2(k-1)$ steps, starting from $(1,1)$ to $((2k-1),1)$. The other side consist of a Dyck path with $2(n-k)$ steps, this is depicted in Figure 4.5.

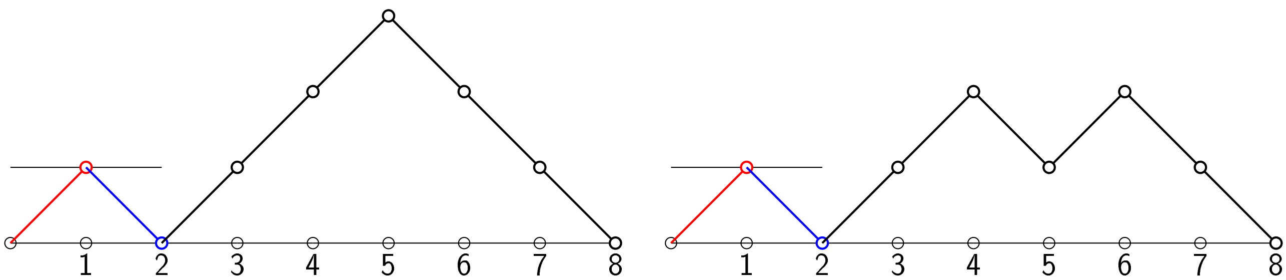


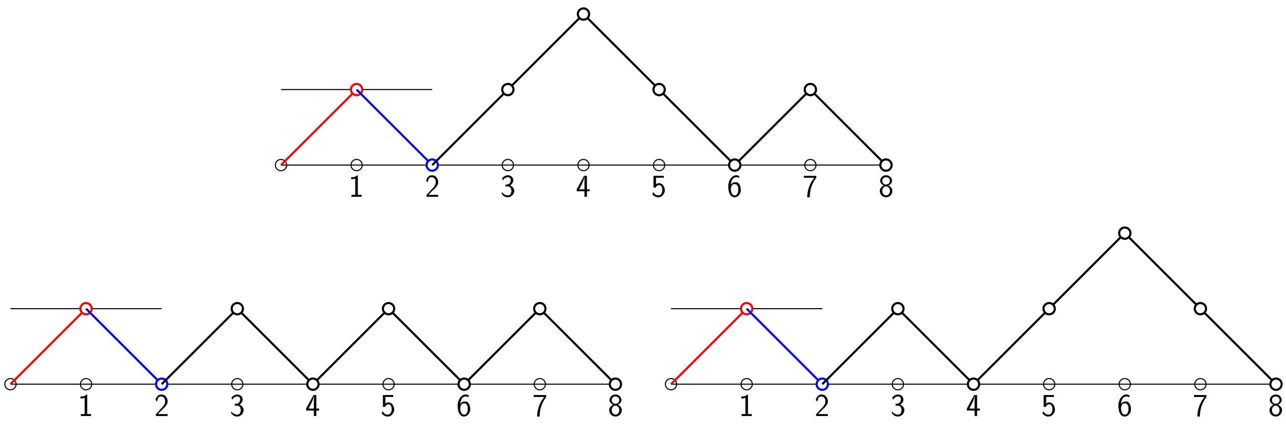
Summing over all such possibilities yield the recurrence formula;

$$D_n = \sum_{k=0}^{n-1} D_k D_{n-k-1}, \text{ where } D_0 = 1.$$

Example 4.5.1. Consider the D_4 case. the Dyck path touches the $y = 0$ line at the point $(2n, 0)$ for $n = 1, 2, 3, 4$ as mentioned in the definition, but for $n = 1, 2, 3, 4$ only applies for the Dyck paths of D_4.

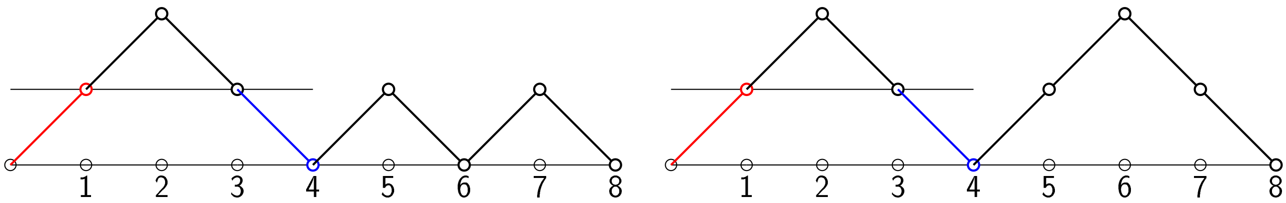
CASE.I. At Point $(2, 0)$





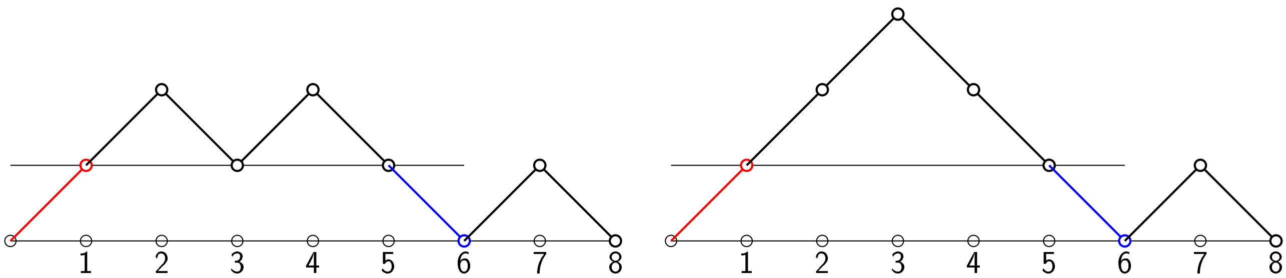
When the path touches $y = 0$ for the first time at $(2, 0)$ and the first leg from $(0, 0)$ [ie. the red path from $(0, 0)$] and the last leg before reaching $(2, 0)$ [ie. the blue path going to $(2, 0)$] are omitted. On counting the path above the omitted legs before $(2, 0)$ yields D_0, since there are no paths above these omitted legs. And counting the remaining path after $(2, 0)$ yields C_3. The result of the two counts is $D_0 D_3$. Where $D_0 D_3 = 5$, which means when the Dyck path touches $y = 0$ at $(2, 0)$ it yields five different paths as demonstrated above.

CASE.II. At Point $(4, 0)$



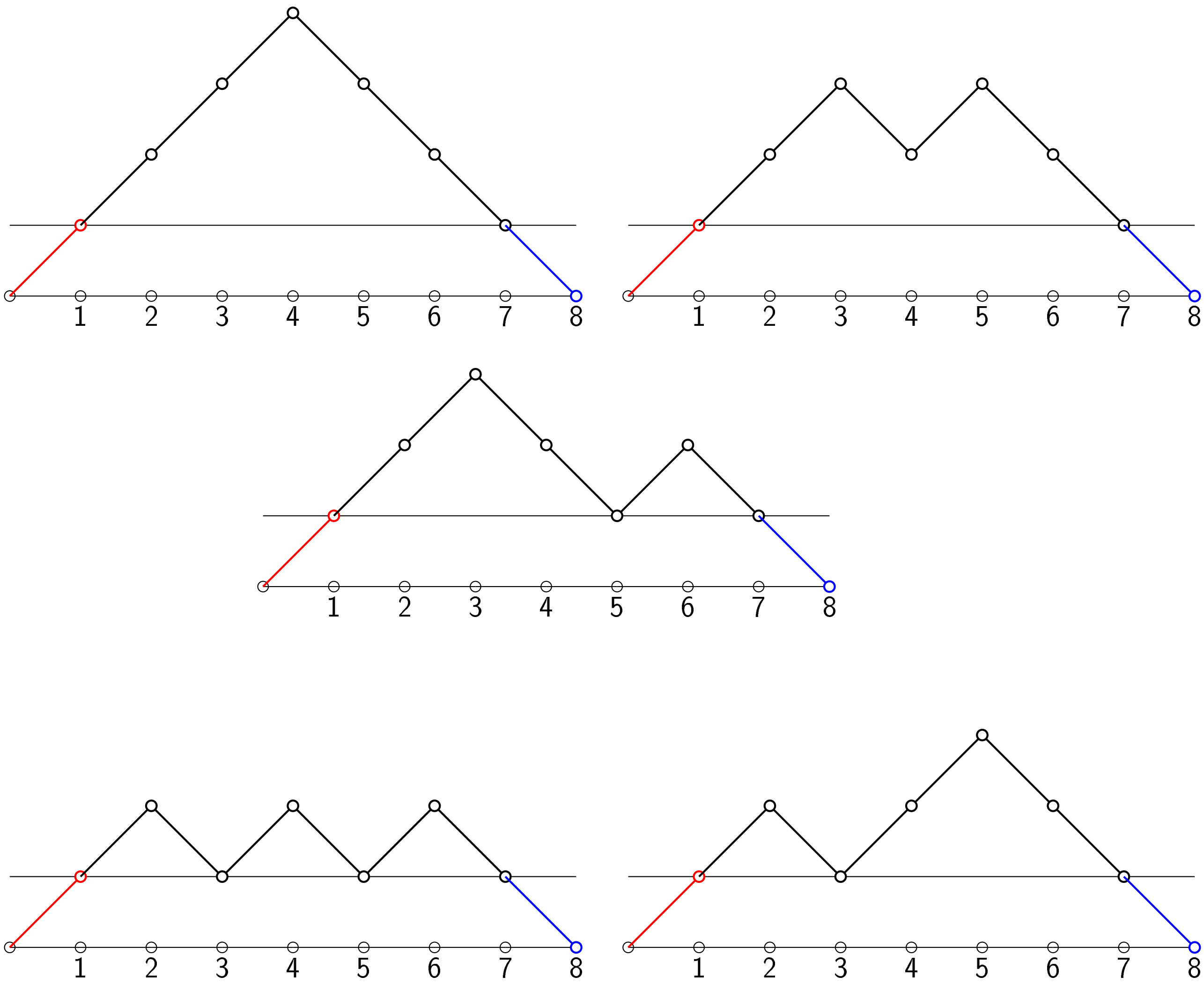
The Dyck path touches $y = 0$ for the first time at $(4, 0)$, above the red and blue leg which are omitted there is a Dyck path of one up-step and one down-step which is D_1 and the remaining path after $(4, 0)$ is a Dyck path of two up-steps and two down-steps which is D_2, the total yields $D_1 D_2 = 2$.

CASE.III. At Point $(6, 0)$



The Dyck path touches $y = 0$ for the first time at $(6, 0)$, above the red and blue leg which are omitted there is a Dyck path of two up-steps and two down-steps which is D_2 and the remaining path after $(6,0)$ is a Dyck path of one up-step and one down-step which is D_1, the total yields $D_2D_1 = 2$.

CASE.IV. At Point $(8,0)$



The Dyck path touches $y = 0$ for the first time at $(8,0)$, above the red and blue leg which are omitted there is a Dyck path of three up-steps and three down-steps which is D_3 and the remaining path after $(4,0)$ is a Dyck path with no up-steps and down-steps which is D_0, the total yields $D_0D_3 = 5$.

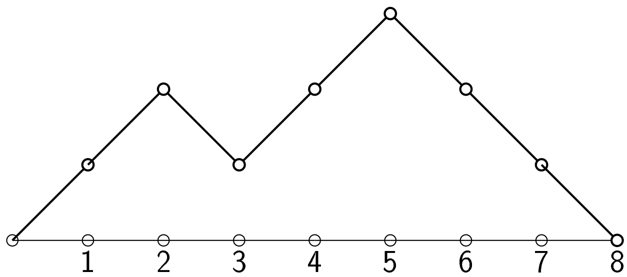
From the above it is sound to conclude that, for Dyck path of four upstairs and four downstairs the number of ways of counting how the four upstairs and downstairs can be arranged as valid Dyck paths is D_4,

$$D_4 = D_0 \cdot D_3 + D_1 \cdot D_2 + D_2 \cdot D_1 + D_3 \cdot D_0 = 5 + 2 + 2 + 5 = 14$$

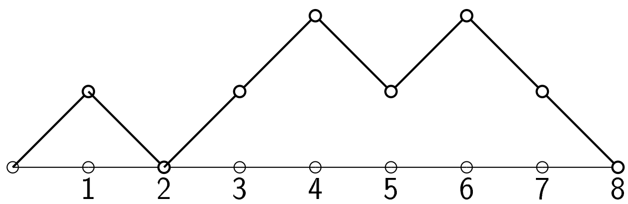
Clearly $C_n = D_n = S_n = T_n$.

4.6 Bijection Between Parenthesization and Dyck Paths

To translate valid Dyck paths into corresponding parenthesizations, we map each up step to an open parenthesis and each down step to a closed parenthesis. This process ensures that the Dyck path's properties of balanced steps correspond to correctly matched parentheses, thus forming valid parenthesizations.



(()(()))



()(()())

In a more technical sense, each step U [up-step] in a Dyck path can be mapped to an open parenthesis (, and each step D [down-step] to a closing parenthesis). This way, a valid Dyck path corresponds directly to a valid sequence of parentheses because the Dyck path's constraint [not falling below the horizontal $y = 0$] ensures that the parentheses are properly nested.

Example 4.6.1. Consider $n = 4$: A Dyck path $UUDUUDDD$ corresponds to the valid parenthesization $(()(()))$. Another Dyck path $UDUUDUDD$ corresponds to the valid parenthesization $()(()())$.As demontrated above.

Both structures share the same length of $2n$, where n is the number of pairs of steps in the Dyck path or the number of pairs of parentheses. In a Dyck path, the characteristic constraint is that the path must stay on or above the horizontal. In valid parenthesization, the characteristic constraint is that every prefix must have at least as many opening parentheses as closing parentheses, ensuring proper nesting.

Chapter 5

Conclusion

In this project, we explored Catalan numbers by studying their recursive nature through three main structures: Dyck paths, valid parenthesizations, and triangulation of convex polygons. We examined how each of these structures could be described by the same recursive relation and solved this relation to find a closed formula for the Catalan numbers: $C_n = \frac{1}{n+1}\binom{2n}{n}$. This shows how Catalan numbers count many different types of combinatorial objects, all connected through a common pattern.

We also found interesting connections between these structures by studying bijections, or one-to-one correspondences, between them. For example, we showed that Dyck paths, ways of placing parentheses, and ways of dividing polygons into triangles can all be linked to each other. This highlights how Catalan numbers provide a unified way to count different types of combinatorial problems.

Beyond combinatorics, Catalan numbers have applications in fields like computer vision, where they are used in image segmentation and contour recognition, and in algebra, where they appear in the study of free Lie algebras and group theory. Catalan numbers continue to play a significant role in both theory and practical problems, connecting different areas of mathematics and computer science.

References

[1] Ralph Grimaldi. *Fibonacci and Catalan Numbers: an introduction*. John Wiley & Sons, 2012.

[2] John M Harris. *Combinatorics and graph theory*. Springer, 2008.

[3] Thomas Koshy. *Catalan numbers with applications*. Oxford University Press, 2008.

[4] John G Michaels and Kenneth H Rosen. *Applications of discrete mathematics*, volume 267. McGraw-Hill New York, 1991.

[5] Robert Sedgewick and Philippe Flajolet. *An Introduction to the Analysis of Algorithms*. Addison-Wesley, 2013.

[6] Richard P Stanley. *Catalan numbers*. Cambridge University Press, 2015.

Chapter 6

Appendices

First 50 Catalan numbers

Table 6.1: Catalan Numbers

n	C_n
1	1
2	2
3	5
4	14
5	42
6	132
7	429
8	1,430
9	4,862
10	16,796
11	58,786
12	208,012
13	742,900
14	2,674,440
15	9,694,845
16	35,357,670
17	129,644,790
18	477,638,700
19	1,767,263,190
20	6,564,120,420

21	24,466,267,020
22	91,482,560,400
23	343,059,613,650
24	1,286,252,828,740
25	4,896,019,301,370
26	18,416,375,290,888
27	69,128,516,849,108
28	259,957,273,244,603
29	978,341,001,600,745
30	3,670,352,887,287,448
31	13,922,800,157,649,009
32	52,989,400,714,478,516
33	201,718,688,047,297,840
34	769,822,933,791,781,509
35	2,938,530,011,546,256,216
36	11,224,532,155,017,653,614
37	42,813,221,375,446,443,840
38	163,542,917,501,910,299,914
39	625,251,601,985,984,238,421
40	2,390,713,731,107,255,715,809
41	9,146,518,788,666,181,618,087
42	35,357,870,433,930,762,422,512
43	136,849,005,344,903,637,551,680
44	530,433,749,766,563,828,163,580
45	2,057,221,843,277,053,017,620,788
46	7,990,766,641,660,907,697,683,200
47	31,218,774,986,795,734,046,899,044
48	122,053,146,674,038,064,928,531,755
49	477,638,700,544,850,632,389,888,001
50	1,874,322,367,705,234,616,265,085,375