4.1 A Warmup Problem

Consider the following recurrence: \( F_0 = 0 \) and \( F_{n+1} = 2F_n + 1 \) for every \( n \geq 0 \). You might recall that this is the recurrence for the Tower of Hanoi problem. We’ll now introduce the method of generating function to solve this recurrence. First, define

\[
f(x) = F_0 + F_1 x + F_2 x^2 + \cdots = \sum_{n=0}^{\infty} F_n x^n.
\]

(Do not worry about the convergence of the infinite sum.) The function \( f(x) \) is called the generating function of the sequence \( \{F_n\} \).

**Step 1: Solving \( f(x) \).** First, we will find a closed-form expression of the generating function \( f(x) \) using the recurrence:

\[
f(x) = F_0 + F_1 x + F_2 x^2 + F_3 x^3 \cdots \\
= F_0 + (2F_0 + 1)x + (2F_1 + 1)x^2 + (2F_2 + 1)x^3 \cdots \\
= F_0 + x(2F_0 x + 2F_1 x^2 + 2F_2 x^3 + \cdots) + (x + x^2 + x^3 + \cdots) \\
= 2xf(x) + x(1 + x + x^2 + \cdots)
\]

Moreover, for \(-1 < x < 1\),

\[
1 + x + x^2 + \cdots = \lim_{n \to \infty} (1 + x + \cdots + x^n) = \lim_{n \to \infty} \frac{1 - x^{n+1}}{1 - x} = \frac{1}{1 - x}.
\]

(Note: The identity \( 1 + x + x^2 + \cdots = \frac{1}{1-x} \) is something that appears frequently in manipulating generating functions, so you may want to memorize it.) Now, we have an equation of \( f(x) \):

\[
f(x) = 2xf(x) + \frac{x}{1-x}.
\]

Solving this equation we obtain

\[
f(x) = \frac{x}{(1-2x)(1-x)}.
\]

**Step 2: Breaking \( f(x) \) into Partial Fractions.** We are interested in expanding \( f(x) \) into an infinite sum, so that we can have a closed-form solution for \( F_n \). In order to achieve this, we must first break \( f(x) \) into some partial fractions so that it’s easier to manipulate. In this context, we want to write \( f(x) \) in the form

\[
f(x) = \frac{A}{1-2x} + \frac{B}{1-x}
\]

for some constant \( A \) and \( B \). There are many ways to obtain \( A \) and \( B \); here’s a sure (but tedious) approach:

\[
\frac{A}{1-2x} + \frac{B}{1-x} = \frac{A(1-x) + B(1-2x)}{(1-2x)(1-x)} = \frac{A + B - (A + 2B)x}{(1-2x)(1-x)}.
\]
Thus, if we want that to be \( \frac{1}{(1-2x)(1-x)} \), it means that \( A + B - (A + 2B)x = x \), and thus \( A + B = 0 \) and \( -(A + 2B) = 1 \). Solving this system of equations we have \( A = 1 \) and \( B = -1 \). Thus

\[
f(x) = \frac{1}{1-2x} - \frac{1}{1-x}.
\]

**Expanding \( f(x) \) into an infinite sum.** Now, on the one hand,

\[
\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n.
\]

On the other hand,

\[
\frac{1}{1-2x} = \sum_{n=0}^{\infty} (2x)^n.
\]

Subtracting side by side, we have

\[
f(x) = \frac{1}{1-2x} - \frac{1}{1-x} = \sum_{n=0}^{\infty} (2^n - 1)x^n.
\]

Thus

\[
f(x) = \sum_{n=0}^{\infty} F_n x^n = \sum_{n=0}^{\infty} (2^n - 1)x^n.
\]

Hence \( F_n = 2^n - 1 \).

### 4.2 Review of Partial Fractions

**First example.** In the section above, once we solve the function \( f(x) \), we need to turn it into partial fractions. We now review the knowledge of partial fractions. Let’s consider the following function

\[
f(x) = \frac{x}{x^2 - 5x + 6}.
\]

The first step is to check whether the degree of the numerator is *strictly* less than that of the denominator. If this condition is not satisfied, the function is not yet in a proper form to handle. Here we are lucky: the degree of the numerator is 1, whereas that of the denominator is 2. To turn \( f(x) \) into partial fractions, we need to factor out the denominator: \( x^2 - 5x + 6 = (x - 2)(x - 3) \). This is the simplest situation we can wish for: the exponents of both factors \( (x - 2) \) and \( (x - 3) \) are 1. Thus the partial-fraction form of \( f(x) \) will be

\[
f(x) = \frac{A}{x-2} + \frac{B}{x-3} = \frac{(A + B)x - (3A + 2B)}{(x-2)(x-3)}.
\]

Hence \( (A + B)x - (3A + 2B) = x \), and consequently \( A + B = 1 \) and \( 3A + 2B = 0 \). Solving this system of linear equations, we have \( A = -2 \) and \( B = 3 \). Hence

\[
f(x) = \frac{-2}{x-2} + \frac{3}{x-3}.
\]

**Second example.** Next, let’s consider

\[
f(x) = \frac{x}{(x-2)^2(x-3)}.
\]
Now the degree of the numerator is 1, which is strictly smaller than the degree of the denominator. The function \( f(x) \) is already in the proper form to find partial fractions. Moreover, the denominator is already factored. In this case, the factor \((x - 3)\) has exponent 1, but the factor \((x - 2)\) has exponent 2. Thus the partial fractions would have the form

\[
\frac{A}{x - 2} + \frac{B}{(x - 2)^2} + \frac{C}{x - 3}
\]

where \(A, B,\) and \(C\) are constant. To find those constants, again

\[
\frac{A}{x - 2} + \frac{B}{(x - 2)^2} + \frac{C}{x - 3} = \frac{A(x - 2) - (x - 3) + B(x - 3) + C(x - 2)^2}{(x - 2)^2(x - 3)}
\]

\[
= \frac{(A + C)x^2 - (5A - B + 4C)x + (6A - 3B + 4C)}{(x - 2)^2(x - 3)}
\]

If we want that to be \(\frac{x}{(x - 2)^2(x - 3)}\), we must have

\[
(A + C)x^2 - (5A - B + 4C)x + (6A - 3B + 4C) = x
\]

Hence

\[
A + C = 0, \\
-(5A - B + 4C) = 1, \\
6A - 3B + 4C = 0.
\]

Solving this system of linear equations, we have \(A = -3, B = -2,\) and \(C = 3\). Hence

\[
f(x) = \frac{-3}{x - 2} + \frac{-2}{(x - 2)^2} + \frac{3}{x - 3}.
\]

**Third example.** Now consider

\[
f(x) = \frac{x^3 + x^2}{(x - 2)(x - 3)}.
\]

First, note that both the degree of the numerator is 3, whereas that of the denominator is just 2. This is not yet a clean form to manipulate: we want the degree of the numerator to be strictly less than the denominator. You want to turn your \(f(x)\) into the form

\[
f(x) = A(x) + \frac{B(x)}{(x - 2)(x - 3)},
\]

where \(A(x)\) and \(B(x)\) are polynomials, and the degree of \(B(x)\) is strictly less than that of \((x - 2)(x - 3)\). How would we do that? Generally, one has to use Long Division of Polynomials. You would divide \(x^3 + x^2\) by \((x - 2)(x - 3)\), and let \(A(x)\) be the quotient, and \(B(x)\) be the remainder. Here I briefly illustrate the steps for this particular example; you can Google the term Long Division of Polynomials for more details.

The division of polynomials resembles your high-school division. Here we divide \(x^3 + x^2\) by \((x - 2)(x - 3) = x^2 - 5x + 6\). Now we need to focus on just the highest order terms on both the dividend and the divisor. In this case the former is \(x^3\) and the latter is \(x^2\). Dividing \(x^3\) by \(x^2\), we obtain \(x\). Now, what’s left is

\[
x^3 + x^2 - x(x^2 - 5x + 6) = 6x^2 - 6x.
\]

Again, repeat the prior step, meaning that we have to divide \(6x^2\) by \(x^2\), which is 6. What is left is

\[
6x^2 - 6x - 6(x^2 - 5x + 6) = 24x - 36.
\]
Now, the degree of $24x - 36$ is just 1, which is smaller than the degree of $x^2 - 5x + 6$, and we have to stop. Thus the quotient $A(x)$ is $x + 6$, and the remainder $B(x) = 24x - 36$. In other words,

$$f(x) = x + 6 + \frac{24x - 36}{(x - 2)(x - 3)} .$$

Now, we need to turn the fraction $\frac{24x - 36}{(x - 2)(x - 3)}$ into partial fractions. This is the simple case, as both factors $(x - 2)$ and $(x - 3)$ have exponent 1. Then the partial fractions would be of the form

$$\frac{C}{x - 2} + \frac{D}{x - 3} ,$$

where $C$ and $D$ are constant. Again

$$\frac{C}{x - 2} + \frac{D}{x - 3} = \frac{C(x - 3) + D(x - 2)}{(x - 2)(x - 3)} = \frac{(C + D)x - (3C + 2D)}{(x - 2)(x - 3)} .$$

If we want that to be $\frac{24x - 36}{(x - 2)(x - 3)}$, we must have

$$(C + D)x - (3C + 2D) = 24x - 36 ,$$

and thus $C + D = 24$ and $3C + 2D = 36$. Solving this system of linear equations, we have $D = 36$ and $C = -12$. Hence

$$f(x) = x + 6 - \frac{12}{x - 2} + \frac{36}{x - 3} .$$

### 4.3 Infinite Expansion

**First example.** Suppose that we want to expand, say $\frac{1}{x - 3}$ into an infinite sum. How would we do that? Of course we want to use the identity

$$\frac{1}{1 - t} = 1 + t + t^2 + \cdots$$

but what would be $t$ in this case? Note that

$$\frac{1}{x - 3} = \frac{-1}{3 - x} = \frac{-1/3}{1 - x/3} .$$

Now, let $t = x/3$. Then we have

$$\frac{1}{1 - x/3} = 1 + (x/3) + (x/3)^2 + \cdots .$$

Multiplying both sides by $-1/3$, we have

$$\frac{1}{x - 3} = \sum_{n=0}^{\infty} \frac{-1}{3} \cdot 3^n x^n = \sum_{n=0}^{\infty} -3^{n-1} x^n .$$

**Second example.** Now, suppose that we want to expand $\frac{1}{(1-x)^2}$ into an infinite sum. There are two ways to do that. Here's the first way. First, recall that

$$\frac{1}{1 - x} = 1 + x + x^2 + x^3 + \cdots .$$
If we take derivative of both sides, we’ll have
\[
\frac{1}{(1-x)^2} = 1 + 2x + 3x^3 + \cdots.
\]

Alternatively, here’s a general formula:
\[
(1 + t)^\alpha = 1 + \alpha t + \frac{\alpha(\alpha-1)}{2!} t^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!} t^3 + \cdots
\]  
(4.1)

If \(\alpha\) is a positive integer then you get back the well-known Binomial Theorem. However, the expansion above is a lot more generalized: it works for any real number \(\alpha\), even a negative one. So for \(t = -x\) and \(\alpha = -2\), we have
\[
(1 - x)^{-2} = 1 + (-2)(-x) + \frac{(-2) \cdot (-3)}{2!} (-x)^2 + \frac{(-2) \cdot (-3) \cdot (-4)}{3!} (-x)^3 + \cdots
\]

THIRD EXAMPLE. Now suppose that you have \(f(x) = \sqrt{1 - 4x}\). In order to expand it, we need to use the identity in Equation 4.1. Use \(t = -4x\) and \(\alpha = 1/2\), we have
\[
\sqrt{1 - 4x} = 1 - \frac{1}{2} \cdot 4x + \frac{1/2 \cdot (-1/2)}{2!} (4x)^2 - \frac{1/2 \cdot (-1/2) \cdot (-3/2)}{3!} (4x)^3 + \cdots
\]
\[
= 1 - 2x - \sum_{n=2}^{\infty} \frac{2^n \cdot 1 \cdot 3 \cdot 5 \cdots (2n - 3)}{n!} x^n.
\]

The coefficients in the infinite sum are somewhat complicated, but we can simplify them further:
\[
\frac{2^n}{n!} \cdot 1 \cdot 3 \cdot 5 \cdots (2n - 3) = \frac{2^n}{n!} \cdot \frac{(2n - 2)!}{2 \cdot 4 \cdot 6 \cdots (2n - 2)}
\]
\[
= \frac{2^n}{n!} \cdot \frac{(2n - 2)!}{2^{n-1} \cdot (n-1)!}
\]
\[
= \frac{2^n}{n!} \cdot \left(\frac{2n - 2}{n-1}\right).
\]

In conclusion,
\[
\sqrt{1 - 4x} = 1 - 2x - \sum_{n=2}^{\infty} \frac{2^n}{n!} \cdot \left(\frac{2n - 2}{n-1}\right) x^n = 1 - 2x - \sum_{k=1}^{\infty} \frac{2}{k+1} \cdot \left(\frac{2k}{k}\right) x^{k+1},
\]
where the last equation is obtained by setting \(k = n - 1\).

### 4.4 Multiplying Generating Functions

Suppose that we have two generating functions
\[
f(x) = F_0 + F_1 x + F_2 x^2 + \cdots, \quad \text{and} \quad g(x) = G_0 + G_1 x + G_2 x^2 + \cdots.
\]

In manipulating generation functions, we often need to add or multiply them. Addition is straightforward
\[
f(x) + g(x) = (F_0 + G_0) + (F_1 + G_1)x + (F_2 + G_2)x^2 + \cdots,
\]
but multiplication is not so simple

\[ f(x)g(x) = \sum_{n=0}^{\infty} \left( \sum_{k=0}^{n} F_k G_{n-k} \right) x^n. \]

(4.2)

In particular,

\[ (f(x))^2 = \sum_{n=0}^{\infty} \left( \sum_{k=0}^{n} F_k F_{n-k} \right) x^n. \]

(4.3)

4.5 Another Example: Fibonacci Sequence

Let’s use the generating-function method to find a closed-form solution of the famous Fibonacci sequence: \( F_0 = F_1 = 1 \) and \( F_{n+2} = F_{n+1} + F_n \) for every \( n \geq 0 \). First, we’ll solve the generating function of \( \{F_n\} \).

\[
\begin{align*}
f(x) &= F_0 + F_1 x + F_2 x^2 + \cdots \\
&= F_0 + F_1 x + (F_0 + F_1) x^2 + (F_1 + F_2) x^3 + \cdots \\
&= F_0 + F_1 x + x^2(F_0 + F_1 x + \cdots) + x(F_1 x + F_2 x^2 + \cdots) \\
&= F_0 + F_1 x + x^2 f(x) + x(f(x) - F_0) \\
&= 1 + x^2 f(x) + x f(x).
\end{align*}
\]

Hence

\[
f(x) = \frac{1}{1 - x - x^2}.
\]

Now, the degree of the numerator is 0, which is strictly less than that of the denominator, and thus \( f(x) \) is already in proper form to be turned into partial fractions. In order to do that, we need to factor the denominator \( g(x) = 1 - x - x^2 \) into the form \( (1 - ax)(1 - bx) \). By comparing the corresponding coefficients of \( 1 - x - x^2 \) and \( (1 - ax)(1 - bx) \), we have

\[
\begin{align*}
a + b &= 1, \\
ab &= 1.
\end{align*}
\]

By solving this system of equations, we have \( a = \frac{1 + \sqrt{5}}{2} \) and \( b = \frac{1 - \sqrt{5}}{2} \). Hence the partial fraction of \( f(x) \) should have the form

\[
\frac{A}{1 - ax} + \frac{B}{1 - bx} = \frac{A + B - (Ab + Ba)x}{(1 - ax)(1 - bx)}.
\]

Hence \( A + B = 1 \) and \( Ab + Ba = 0 \). Solving this system of linear equations we have \( B = \frac{b}{b-a} \) and \( A = \frac{-a}{b-a} \). Then

\[
\begin{align*}
f(x) &= \frac{-a}{b-a} \cdot \frac{1}{1 - ax} + \frac{b}{b-a} \cdot \frac{1}{1 - bx} \\
&= \frac{-a}{b-a} \left( 1 + ax + a^2 x^2 + \cdots \right) + \frac{b}{b-a} \left( 1 + bx + b^2 x^2 + \cdots \right) \\
&= \sum_{n=0}^{\infty} \frac{b^{n+1} - a^{n+1}}{b-a} x^n.
\end{align*}
\]

Hence \( F_n = (b^{n+1} - a^{n+1})/(b - a) \).
4.6 Another Example: Binary Trees

Recall that a binary tree is a (rooted) tree in which each node has at most 2 children. We now use the generating-function method to compute the number \( C_n \) of binary trees of exactly \( n \) nodes. The number \( C_n \) is also known as the \( n \)-th Catalan number. As illustrated below, we have \( C_1 = 1 \), \( C_2 = 2 \), and \( C_3 = 5 \).

![Figure 4.1: Enumeration of binary trees of 1, 2, and 3 nodes.](image)

The exact value. Suppose that on the left subtree, we have \( k \) nodes, meaning that there are \( C_k \) choices of the left subtree. Moreover, there are \( n - k - 1 \) nodes on the right subtree, and thus there are \( C_{n-k-1} \) choices for the right subtree. Hence for each \( k \), there are corresponding \( C_k \cdot C_{n-k-1} \) binary trees of \( n \) nodes, and thus

\[
C_n = \sum_{k=0}^{n-1} C_k \cdot C_{n-k-1} .
\]

For the base case, we have \( C_0 = 1 \). Let’s solve the generating function of \( \{C_n\} \).

\[
f(x) = \sum_{n=0}^{\infty} C_n x^n = C_0 + \sum_{n=1}^{\infty} \left( \sum_{k=0}^{n-1} C_k C_{n-k-1} \right) x^n
= C_0 + x \left( \sum_{k=0}^{\infty} C_k x^k \right)^2 = 1 + x f(x)^2 .
\]

The quadratic equation \( t = 1 + xt^2 \) has two roots \( \frac{1 + \sqrt{1-4x}}{2x} \) and \( \frac{1 - \sqrt{1-4x}}{2x} \), but we know that

\[
\lim_{x \to 0} f(x) = C_0 = 1 ,
\]

whereas

\[
\lim_{x \to 0} \frac{1 + \sqrt{1-4x}}{2x} = \infty .
\]

Since \( f(x) > C_0 = 1 \) for every \( x > 0 \), the only choice for \( f(x) \) is

\[
f(x) = \frac{1 - \sqrt{1-4x}}{2x} .
\]
To expand $f(x)$ into an infinite sum, recall that
\[
\sqrt{1 - 4x} = 1 - 2x - \sum_{k=1}^{\infty} \frac{2}{k+1} \binom{2k}{k} x^{k+1}.
\]
Hence
\[
f(x) = 1 + \sum_{k=1}^{\infty} \frac{1}{k+1} \binom{2k}{k} x^k,
\]
and thus $C_k = \frac{1}{\pi^{3/2}} \binom{2k}{k}$.

**Approximating Catalan numbers.** While it’s nice to know the exact value of $C_n$, in many situations, it’s better to have just an approximation in a more convenient form. The term $1/(n + 1)$ in $C_n$ can be simplified as
\[
\frac{1}{n+1} \approx \frac{1}{n}.
\]
Thus what is left is to approximate the term
\[
\binom{2n}{n} = \frac{(2n)!}{n! \cdot n!}.
\]
(4.4)
When dealing with factorials, it’s useful to know Stirling’s approximation:
\[
N! \approx \sqrt{2\pi N} \left(\frac{N}{e}\right)^N.
\]
In particular, if we apply Stirling’s approximation for the numerator of the right-hand side of Equation 4.4,
\[
(2n)! \approx 2\sqrt{\pi n} \cdot \left(\frac{2n}{e}\right)^{2n}\]
Likewise, we can approximate the denominator of the right-hand side of Equation 4.4 as
\[
n! \cdot n! \approx 2\pi n \cdot \left(\frac{n}{e}\right)^{2n}\]
Hence
\[
\binom{2n}{n} \approx \frac{4^n}{\sqrt{n} \cdot n!}
\]
and thus
\[
C_n \approx \frac{4^n}{\sqrt{\pi} \cdot n!}.
\]
**Encoding binary trees.** Binary trees occur frequently in computer science, and thus it’s desirable to find a succinct representation for them. Since there are $C_n$ binary trees of $n$ nodes, one needs at least $\log_2(C_n) \approx 2n - 1.5 \log_2(n)$ bits to encode such a tree. We now give a nearly optimal encoding, using up to $2n + 1$ bits. For convenience, we view each node as having exactly two (possibly null) children. Thus there are totally at most $2n + 1$ nodes, and exactly $n$ of them are non-null. Label each non-null node 1, and label other nodes 0. To encode this tree, we will print out the labels in the preorder traversal. This encoding has at most $2n + 1$ bits, since there are at most $2n$ nodes, and we use only one bit per node. See Figure 4.2 for an illustration. There’s a simple recursive, linear-time algorithm to decode this encoding; this is left as an exercise.
Figure 4.2: **Left:** The original tree. **Right:** An extended tree with null nodes, and the labeling. This tree will be encoded as 111001001011000.