Applications of the Goulden-Jackson cluster method to counting Dyck paths by occurrences of subwords

A Dissertation

Presented to
The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Mathematics
Ira Gessel, Advisor

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
Chao-Jen Wang

August, 2011
This dissertation, directed and approved by Chao-Jen Wang’s committee, has been accepted and approved by the Faculty of Brandeis University in partial fulfillment of the requirements for the degree of:

DOCTOR OF PHILOSOPHY

Malcolm Watson, Dean of Arts and Sciences

Dissertation Committee:

Ira Gessel, Dept. of Mathematics, Chair.

Susan Parker, Dept. of Mathematics

Ji Li, Dept. of Mathematics, Babson College
Acknowledgments

The work in this thesis would not have begun without the guidance and advice of my supervisor, Professor Ira Gessel, to whom I give my sincerest thanks. His many insights and good ideas, as well as his generosity and patience during my moments of confusion, made this thesis possible.

I am grateful to the members of my dissertation defense committee, Professor Susan Parker and Professor Ji Li.

I would also like to thank my friends and classmates for their support, as well as the Chu family for their encouragement and help throughout my days at Brandeis.

Finally, I wish to thank my family for always being there for me.
Abstract

Applications of the Goulden-Jackson cluster method to counting Dyck paths by occurrences of subwords

A dissertation presented to the Faculty of the Graduate School of Arts and Sciences of Brandeis University, Waltham, Massachusetts

by Chao-Jen Wang

Goulden and Jackson introduced the cluster method for counting words avoiding a prescribed set of subwords in [14, 15]. Noonan and Zeilberger [17] generalized it and wrote many Maple programs to implement the method and its extensions. We count Dyck paths according to the number of occurrences of certain patterns, using a variation of the Goulden-Jackson cluster method. We will give several examples of counting Dyck paths by occurrences of subwords and show how to use the cluster method to compute generating functions for those examples. Then we show more applications to count paths with bounded height by occurrences of subwords and more applications to count $r$–Dyck paths.
CHAPTER 1

Introduction

We first give some definitions and describe the general Goulden-Jackson cluster method with examples in Chapter 1. Then we give several examples of counting Dyck paths by occurrences of subwords and show how to use the cluster method to compute generating functions for those examples in Chapter 2. We apply the cluster method to count paths with bounded height by occurrences of subwords in Chapter 3. We show more applications to count $r$–Dyck paths in Chapter 4.

A Dyck path is a path in the first quadrant which begins at the origin. It ends at $(2n, 0)$ and consists of steps $(1, 1)$, called rises, and $(1, -1)$, called falls. We will refer to $n$ as the semilength of the path. It is well-known that the number of all Dyck paths of semilength $n$ is the $n$th Catalan number

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

The Catalan number generating function is

$$C(x) = \sum_n c_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x}.$$  

We can encode each rise by a letter $U$ for an up step and each fall by a letter $D$ for a down step, obtaining the encoding of a Dyck path by a Dyck word.

A Motzkin path is a path in the first quadrant which begins at the origin. It ends at $(n, 0)$ and consists of steps $(1, 1)$, $(1, 0)$, and $(1, -1)$. Here $n$ is the length of the path. The number of all Motzkin $n$-paths (paths with length $n$) is the $n$th Motzkin
number \( m_n \). The Motzkin number generating function is

\[
M(x) = \sum_{n} m_n x^n = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2}.
\]

We can encode \((1,1)\) by a letter \( U \) for an up step, \((1,-1)\) by a letter \( D \) for a down step, and \((1,0)\) by a letter \( F \) for a flat step, obtaining the encoding of a Motzkin path by a Motzkin word.

Goulden and Jackson introduced the cluster method for counting words avoiding a prescribed set of subwords in \([14, 15]\). Noonan and Zeilberger \([17]\) generalized it and wrote many Maple programs to implement the method and its extensions. See also Stanley \([19, \text{Ch. 4, Ex. 14}]\), Kupin and Yuster \([16]\), and the references given there. We first illustrate the approach and describe the general Goulden-Jackson cluster method with an example:

Let \( w = w_1w_2\cdots w_n \) be a word in an alphabet \( A = \{a_1, a_2, \ldots, a_k\} \) and let \( A^* \) be the set of words made up by letters in \( A \). Define the length of \( w \) as \( l(w) = n \). A marked subword of \( w \) is a pair \((i,v)\) such that

\[
v = w_iw_{i+1}\cdots w_{i+l(v)-1}
\]

where \( l(v) \geq 2 \). Here \( i \) indicates where the marked subword starts in \( w \) and \( v \) is the subword. A marked word is a word \( w \) together with a (possibly empty) set of marked subwords of \( w \).

For example, the word \( abbaba \) together with the set of marked subwords

\[
\{(1, abb), (3, ba), (5, ba)\}
\]

is a marked word which we represent as

\[
\text{(a b) (b a) (b a)}
\]
We can concatenate marked words in the obvious way. For example, concatenating
\[ a \overline{b} a \quad \text{and} \quad \overline{b} a b a \quad \text{gives} \quad a \overline{b} a \overline{b} a b a \]

A marked word is a \textit{cluster} if it is not a concatenation of two nonempty marked words.

A marked word is the same as a word in the set of single letters and clusters. We can define \( f(t) \) as the generating function for a set of marked words.

Given a set \( S \) of words of length at least 2, we may consider the generating function

\[
f(t) = \sum_w wt^{n(w)}
\]

where the sum runs over all words \( w \in A^* \) and \( n(w) \) is the number of occurrences of marked words in \( S \) in \( w \). We think of the letters as noncommuting variables and \( t \) as commuting with these variables. It is easier to compute

\[
f(1 + t) = \sum_w w(1 + t)^{n(w)}
\]

\[
= \sum_w w \sum_k \binom{n(w)}{k} t^k
\]

\[
= \sum_w \sum_{T \subseteq B} t^{|T|}
\]

where \( B \) is the set of occurrences of words in \( S \) in \( w \).

So \( f(1 + t) \) is the sum of the weights of the marked words whose marked subwords are in \( S \), where the weight of a marked word \( w \) is the underlying word times \( t^{m(w)} \), where \( m(w) \) is the number of marked subwords in \( w \).

Therefore, we have

\[
f(1 + t) = (1 - a_1 - a_2 - \cdots - a_k - L(t))^{-1}
\]
where $A = \{a_1, a_2, \ldots, a_k\}$ is an alphabet and $L(t)$ is the generating function for clusters.

More generally, we can use different weights for different words. By the same reasoning, we have the following result:

**Theorem 1.** Let $A = \{a_1, a_2, \ldots, a_i\}$ be an alphabet. Let $S = \{v_1, v_2, \ldots, v_k\}$ be a set of words of length at least 2. Let $f(t_1, t_2, \ldots, t_k)$ be the generating function for counting words in $A^*$ by occurrences of $v_1, v_2, \ldots, v_k$, where we assign the weight $t_j$ to $v_j$. Then

$$f(1 + t_1, 1 + t_2, \ldots, 1 + t_k) = (1 - a_1 - a_2 - \cdots - a_i - L(t_1, t_2, \ldots, t_k))^{-1}$$

where $L(t_1, t_2, \ldots, t_k)$ is the generating function for clusters.

For example, we want to count all words in $\{a, b, c\}^*$ by occurrences of $ab$ (weighted $t_1$) and occurrences of $bc$ (weighted $t_2$). Let $f(t_1, t_2)$ be the generating function

$$f(t_1, t_2) = \sum_w wt_1^i t_2^j$$

where $i$ and $j$ represent the number of occurrences of $ab$ and $bc$ in $w$. Consider

$$f(1 + t_1, 1 + t_2) = \sum_w w(1 + t_1)^i(1 + t_2)^j.$$ 

This counts all words in $\{a, b, c\}^*$ in which some occurrences of $ab$ may be marked and some occurrences of $bc$ may be marked. For example, a given word $w = ababc$ could contribute different marked words:

- $a\ b\ a\ b\ c$ together with empty set.
- $\overline{a\ b}\ a\ b\ c$ together with $\{(1, ab)\}$.
- $a\ b\ \overline{a\ b}\ c$ together with $\{(3, ab)\}$. 

Because \( ababc \) contains two occurrences of \( ab \) and one occurrence of \( bc \), its coefficient in the sum of the weights of the markings of \( ababc \) is \((1 + t_1)^2(1 + t_2)\), which corresponds to the eight marked words above.

On the other hand, all of the marked words are made of letters and clusters:

\[
a, \ b, \ c, \ \overline{a \ b}, \ \overline{b \ c}, \ \overline{a \ b \ c}
\]

Thus,

\[
f(1 + t_1, 1 + t_2) = (1 - a - b - c - abt_1 - bct_2 - abct_1 t_2)^{-1}. \tag{1}
\]

Therefore, we can get the real generating function \( f(t_1, t_2) \) by replacing \( t_1 \) with \( t_1 - 1 \) and \( t_2 \) with \( t_2 - 1 \) in equation (1). So

\[
f(t_1, t_2) = (1 - a - b - c - ab(t_1 - 1) - bc(t_2 - 1) - abc(t_1 - 1)(t_2 - 1))^{-1}.
\]

We can use the same method for counting more general sets of words. We can always reduce a problem of counting a set of words by occurrences of subwords to a problem of counting marked words. This approach will be useful whenever we can count the corresponding marked words.

We will apply the method to counting Dyck words. Suppose \( S = \{UU, UDU\} \), and \( w \) is restricted to Dyck words, then a marked Dyck word could be

\[
U(U(U)D(U)U) \ D U(U D(U)U)
\]
It is not true that every concatenation of clusters is a marked Dyck word, so we need to do more work to count marked Dyck words by replacing each cluster with a new step. In this example, there are single down steps $D$, single up steps $U$, and clusters consisting of $UU$ and $UDU$. So the problem reduces to counting paths with a more general set of steps that never go below the $x$-axis.

There is one complication. If, for example, $S$ contains the word $DU$ then we replace it with a flat step $F$, but this step cannot occur at height 0. This problem is no longer reduced to a problem of counting Motzkin paths. Instead, it reduces to a problem of counting Motzkin paths with no flat step at height 0. There are some special cases that are easier than the general case:

1. Each cluster is equivalent to an up step $U$, a down step $D$, or a flat step $F$, and there is no restriction on where they occur. So the problem is equivalent to counting Dyck paths or Motzkin paths. Examples are \{UUD\}, \{UD\}, \{UDU\}, etc.

2. Each cluster is equivalent to an up step $U$, a down step $D$, or a flat step $F$, but there are some restrictions on where they occur. We can derive generating functions for this case from some quadratic equations related to the Catalan, Motzkin or Narayana generating function. Examples are \{DU\} (cannot occur at height 0), \{DDU\} (cannot occur at height 1 or 0), etc.

3. The clusters are equivalent to steps that can go up by an arbitrary amount or down by at most 1 (or vice-versa). We can derive generating functions for this case from some functional equations which are sometimes quadratic or even of higher degree. So we may apply Lagrange inversion [20, Ch. 5, Page. 38] to solve them. Examples are \{UU\}, \{DD\}, \{UUU\}, etc.
CHAPTER 2

Examples of Counting Dyck Paths by Occurrences of Subwords

In the following examples, we compute a generating function $g$ for counting Dyck paths by occurrences of subwords in which each marked word is counted with a weight $1 + t$ and then we compute the real generating function $h$ with a weight $t$ by replacing $t$ with $t - 1$ in $g$.

2.1. Occurrences of UUD

Count Dyck paths by occurrences of $UUD$. For example, a marked Dyck word could be

\[ UD(UUD) U U D D D \]

We assign to such a marked word the weight $x^i t^j$, where the semilength is $i$ and there are $j$ marked occurrences of $UUD$. To count all these marked words, we replace each occurrences of $UUD$ by a new up step $U'$. So our example would be replaced with

\[ UD U' U U D D D \]

Note that the original word is a marked Dyck word if and only if the new word (when $U'$ is replaced by $U$) is a Dyck word. In this example, $UUD$ is the only cluster. Since the occurrence of $U$ in a Dyck path equals that of its semilength, we can count modified Dyck words where $U$ has the weight $x$, $D$ has the weight $1$, and $U'$ has the weight $x^2 t$. Equivalently, we can count ordinary Dyck paths where each up step is
weighted by $x + x^2t$. We set $u_1 = x + x^2t$ and $d_1 = 1$, where $u_1$ is the generating function for reducing up steps. Then we count Dyck paths with up steps weighted by $u_1$ and down steps weighted by $d_1$. To count these, we use a well-known decomposition called the first return decomposition.

Every nonempty Dyck path can be decomposed at the first return to $x$-axis. Every nonempty Dyck path can be factored as $UG_1DG_2$ where $U$ is an up step, $D$ is a down step, and $G_1, G_2$ are (possibly empty) Dyck paths. See Figure 2.1.

![Figure 2.1. The first return decomposition for Dyck paths](image)

So the generating function $g(x, t)$ with a weight $1 + t$ satisfies

$$g = 1 + u_1d_1g$$

where 1 represents the empty path, and $u_1d_1g$ represents the decomposition for nonempty Dyck paths. Replacing $u_1$ by $x + x^2t$, and $d_1$ by 1, we get

$$g = 1 + (x + x^2t)g^2.$$ 

Solving for $g$, we get

$$g = \frac{1 - \sqrt{1 - 4x(1 + xt)}}{2x(1 + xt)}$$

$$= C(x + x^2t).$$
As described in the introduction, \( g \) counts Dyck words where every occurrence of \( UUD \) is weighted by \( 1 + t \). So the generating function \( h \) for Dyck words weighted by \( t^j \) where \( j \) is the number of occurrences of \( UUD \), is obtained by replacing \( t \) with \( t - 1 \) in \( g \). So we get the real generating function \( h \) in which every occurrence of \( UUD \) is weighted by \( t \):

\[
h(x, t) = \frac{1 - \sqrt{1 - 4x(1 - x + xt)}}{2x(1 - x + xt)}
\]

\[
= 1 + x + (1 + t)x^2 + (1 + 4t)x^3
\]

\[
+ (1 + 11t + 2t^2)x^4 + (1 + 26t + 15t^2)x^5 + \cdots
\]

Here the coefficients are sequence A091156 in the Online Encyclopedia of Integer Sequences [18], where they are described as the number of Dyck paths of semilength \( n \), having \( k \) long ascents (i.e, ascents of length at least 2). It is easy to see that every Dyck path having \( k \) long ascents has exactly \( k \) occurrences of \( UUD \), since every long ascent must be followed by a down step \( D \).

In particular, for \( t = 0 \), we have

\[
h(x, 0) = \frac{1}{1 - x}
\]

\[
= 1 + x + x^2 + x^3 + \cdots
\]

This is the generating function of \( UUD \)-free Dyck paths (i.e. Dyck paths with no occurrences of \( UUD \)) with semilength weighted by \( x \). The only \( UUD \)-free Dyck paths are of the form \((UD)^n\). Therefore, the coefficients of powers of \( x \) in \( h(x, 0) \) are all 1.
2.2. Occurrences of UDU

Suppose we now count Dyck paths by occurrences of $UDU$. The occurrences of $UDU$ are weighted by $t$. In this example, we use the same approach as in the case of occurrences of $UUD$ in section 2.1. We need to find the clusters first, since the clusters are no longer trivial.

In this case, the clusters are marked Dyck words of the form $U(DU)^i$ for $i = 1, 2, 3, \ldots$:

$UDU, UDUDU, UDUDUDU, \ldots$

So the cluster generating function $L(t)$ is

$$udut + ududut^2 + udududut^3 + \cdots = \sum_{i \geq 1} u(du)^i t^i = \frac{u^2 dt}{1 - udt}$$

where $u$ and $d$ are commuting variables.

We count these modified Dyck words where $U$ has the weight $x$, $D$ has the weight 1, and the cluster generating function is $\frac{x^2 t}{1 - xt}$. Since these clusters reduce to up steps, we can set $u_1 = x + \frac{x^2 t}{1 - xt} = \frac{x}{1 - xt}$ to get that the generating function $g(x, t)$ with a weight $1 + t$ satisfies

$$g = 1 + u_1 g d_1 g$$

$$= 1 + \left( \frac{x}{1 - xt} \right) g^2$$

where $u_1$ represents reducing up steps and $d_1$ represents reducing down steps.
Notice that any functional equation of this form

\[ g = 1 + ag^2 \]

has the solution

\[ g = \frac{1 - \sqrt{1 - 4a}}{2a} = C(a). \]

Therefore, by equation (2), we get

\[
g(x, t) = C \left( \frac{x}{1 - xt} \right) \]

\[ = \frac{1 - \sqrt{1 - 4 \left( \frac{x}{1 - xt} \right)}}{2 \left( \frac{x}{1 - xt} \right)} \]

\[ = \frac{1 - xt - \sqrt{(1 - 4x - xt)(1 - xt)}}{2x}. \]

Using the cluster method, we replace \( t \) by \( t - 1 \), and we get the real generating function

\[
h(x, t) = \frac{1 + x - xt - \sqrt{(1 - 3x - xt)(1 + x - xt)}}{2x} \]

\[ = 1 + x + (t + 1)x^2 + (2 + 2t + t^2)x^3 + (4 + 6t + 3t^2 + t^3)x^4 + \cdots \]

Here the coefficients are sequence A091869 in the Online Encyclopedia of Integer Sequences [18]. Some related statistics have been studied by Sun [21].

In particular, for \( t = 0 \), we have

\[
h(x, 0) = \frac{1 + x - \sqrt{(1 - 3x)(1 + x)}}{2x} \]

\[ = \frac{1 + x - \sqrt{1 - 2x - 3x^2}}{2x} \]

\[ = 1 + x + x^2 + 2x^3 + 4x^4 + 9x^5 + 21x^6 + 51x^7 + 127x^8 + \cdots \]
This is the generating function of $UDU$-free Dyck paths with semilength weighted by $x$. Subtracting 1 from it and dividing by $x$ gives the generating function for Motzkin numbers as shown by Donaghey and Shapiro [11]. So the number of $UDU$-free Dyck paths with semilength $n$ is $m_{n-1}$, the $(n-1)$th Motzkin number.

We can also verify that

$$h(x, t) = 1 + \frac{x}{1 - xt} M \left( \frac{x}{1 - xt} \right)$$  \hspace{1cm} (3)$$

where $M(x) = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2}$, the Motzkin number generating function.

Expanding the right side of equation (3), we have

$$1 + \frac{x}{1 - xt} M \left( \frac{x}{1 - xt} \right) = 1 + \frac{1 - \left( \frac{x}{1 - xt} \right) - \sqrt{1 - 2 \left( \frac{x}{1 - xt} \right) - 3 \left( \frac{x}{1 - xt} \right)^2}}{2 \left( \frac{x}{1 - xt} \right)}$$

$$= 1 + \frac{1 - xt - x - \sqrt{(1 - xt)^2 - 2(1 - xt) - 3x^2}}{2x}$$

$$= 1 + \frac{x - xt - \sqrt{(1 - 3x - xt)(1 + x - xt)}}{2x}$$

$$= h(x, t).$$

Using equation (3), we can get an explicit formula for the coefficients of $h(x, t)$ so that

$$h(x, t) = 1 + \frac{x}{1 - xt} M \left( \frac{x}{1 - xt} \right)$$

$$= 1 + \sum_{i \geq 0} m_i \frac{x^{i+1}}{(1 - xt)^{i+1}}$$

$$= 1 + \sum_{i \geq 0} m_i x^{i+1} \sum_{k \geq 0} \binom{i+k}{k} (xt)^k$$

$$12$$
CHAPTER 2. OCCURRENCES OF SUBWORDS

\[= 1 + \sum_{i \geq 0, k \geq 0} \binom{i + k}{k} m_i x^{i+k+1} t^k\]

\[= 1 + \sum_{n \geq 1, 0 \leq k \leq n-1} \binom{n-1}{k} m_{n-k-1} x^n t^k\]

where \(m_i\) is the \(i\)th Motzkin number.

A bijective proof for this formula has been given by Callan [3].

2.3. Occurrences of UD

Count Dyck paths by occurrences of UD. For example, a marked Dyck word could be

\(\overline{UDUDUDD}\)

We assign to such a marked word the weight \(x^i t^j\) where the semilength is \(i\) and there are \(j\) marked occurrences of UD. To count all these marked words, we replace each occurrence of UD by a new flat step \(F\). So our example would be replaced with

\(FUDUFD\)

Note that the original word is a marked Dyck word if and only if the new word (when UD is replaced by \(F\)) is a Motzkin word. In this example, UD is the only cluster. We count these modified Motzkin words where \(U\) has the weight \(x\), \(D\) has the weight 1, and \(F\) has the weight \(xt\). Every nonempty Motzkin path can start with a flat step \(F\) or an up step \(U\). It can be decomposed into \(FG\) or \(UG_1DG_2\), where \(G, G_1, G_2\) are Motzkin paths. See Figure 2.2.

So the generating function \(g(x, t)\) with a weight \(1 + t\) satisfies

\[g = 1 + fg + ugdg\]
where \( f \) represents reducing flat steps, \( u \) represents reducing up steps, and \( d \) represents reducing down steps. In this case, \( f \) represents a \( UD \), \( u \) represents a single up step \( U \), and \( d \) represents a single down step \( D \).

Replacing \( f \) by \( xt \), \( u \) by 1, and \( d \) by \( x \), we solve equation (4) to get

\[
g = \frac{1 - xt - \sqrt{(1 - xt)^2 - 4x}}{2x}.
\]

Using the cluster method, we replace \( t \) by \( t - 1 \) in \( g \) and get the real generating function

\[
h(x, t) = \frac{1 + x - xt - \sqrt{(1 + x - xt)^2 - 4x}}{2x}.
\]  

(5)

In particular, for \( t = 0 \), we have \( h(x, 0) = 1 \). This is the generating function of Dyck paths with no peak, \( UD \). As we know, the empty path is the only Dyck path with no peak. Note that \( h \) is a generating function for the Narayana numbers \( N(n, k) = \frac{1}{n} \binom{n}{k} \binom{n}{k-1} \) satisfying

\[
h(x, t) = 1 + \sum_{n, k=1}^{\infty} N(n, k) x^n t^k.
\]  

(6)
2.4. Occurrences of DU

Now we look at an example of type 2. Suppose we count Dyck paths by occurrences of $DU$ weighted by $t$. We use the same approach as the case of occurrences of $UD$ in section 2.3. However, here we want to count Motzkin paths with no flat steps at height 0. So a generating function $g(x,t)$ with a weight $1+t$ for counting these modified Motzkin paths where $U$ has the weight $x$, $D$ has the weight 1, and $F$ (when $DU$ is replaced by $F$) has the weight $xt$, satisfies

$$g = 1 + fg + ugdg$$

where $f$, $u$, and $d$ are commuting variables. This is the same as equation (4) in section 2.3. In this case, $f$ represents $DU$, $u$ represents a single up step $U$, and $d$ represents a single down step $D$. However, $DU$ cannot occur at height 0. So the problem can reduce to one of counting Motzkin paths with no flat step at height 0.

Let $g$ be the generating function for counting Motzkin paths with no height restriction. Let $g_1$ be the generating function for counting Motzkin paths with no flat step at height 0. Every nonempty Motzkin path with no flat step at height 0 can be factored as $UGDG_1$ at the first return, where $G$ is a Motzkin path with no height restriction and $G_1$ is a Motzkin path with no flat step at height 0. See Figure 2.3.

![Figure 2.3](image-url)
Then \( g_1 \) satisfies
\[ g_1 = 1 + ugdg_1. \]

So we can write
\[ g_1 = \frac{1}{1 - ugd}. \quad (8) \]

Replacing \( f \) by \( tx \), \( u \) by 1, and \( d \) by \( x \), we solve equation (7) to get
\[ g = \frac{1 - xt - \sqrt{(1 - xt)^2 - 4x}}{2x}. \]

Then we substitute this for \( g \) in equation (8) and solve for \( g_1 \). We get
\[ g_1 = \frac{1}{1 - xg} = \frac{2}{1 + xt + \sqrt{(1 - xt)^2 - 4x}}. \]

Using the cluster method, we replace \( t \) by \( t - 1 \) in \( g_1 \) to get the real generating function
\[ h_1(x,t) = \frac{2}{1 + x(t - 1) + \sqrt{(1 - x(t - 1))^2 - 4x}} \]
\[ = \frac{1 + x(t - 1) - \sqrt{(1 - x(t - 1))^2 - 4x}}{2xt}. \quad (9) \]

In particular, for \( t = 0 \), we have
\[ h_1(x,0) = \frac{2}{1 - x + 1 - x} \]
\[ = \frac{1}{1 - x} \]
\[ = 1 + x + x^2 + x^3 + \cdots \]

This is the generating function of Dyck paths with no valley, \( DU \). The only possible Dyck paths with no valley are of the form \( U^nD^n \). Therefore, all coefficients
CHAPTER 2. OCCURRENCES OF SUBWORDS

of \(h_1(x,0)\) are 1. Note that \(h_1(x,t)\) is also a generating function for the Narayana numbers:

From equation (9), (5) and (6), we have

\[
th_1 - t + 1 = h = 1 + \sum_{n,k=1}^{\infty} N(n,k)x^nt^k.
\]

So,

\[
h_1 = 1 + \sum_{n,k=1}^{\infty} N(n,k)x^nt^k^{-1}.
\] (10)

We can see (10) directly, since a nonempty path with \(k\) peaks has \(k - 1\) valleys.

2.5. Occurrences of DDU

Suppose we now count Dyck paths by occurrences of DDU weighted by \(t\). In this example, the only cluster, DDU, can be reduced to a down step. Using the same approach as in the case of occurrences of UUD in section 2.1, we can start from a generating function \(g\) with no height restriction and with a weight 1+ which satisfies

\[
g = 1 + u_1gd_1g
\] (11)

where \(u_1 = u\) and \(d_1 = d + ddut\), a single down step \(D\) or a reducing down step DDU. Because there is a height restriction that DDU can only start from a height not less than 2, we can elevate \(g\) by using the same approach from equation (8) in section 2.4 to get a generating function \(g_1\) under the height restriction and with a weight 1+ which satisfies

\[
g_1 = \frac{1}{1 - ugd}.
\] (12)

In equation (11), we replace \(u\) by \(x\), and \(d\) by 1 to get

\[
g = 1 + x(1 + xt)g^2.
\]
Solving for $g$,

$$g = C(x(1+xt))$$

$$= \frac{1 - \sqrt{1 - 4x(1+xt)}}{2x(1+xt)}.$$ 

This is the same $g$ as in section 2.1. We can substitute this for $g$ in equation (12) to get

$$g_1 = \frac{1}{1-xg}$$

$$= \frac{2(1+xt)}{1 + 2xt + \sqrt{1-4x(1+xt)}}.$$ 

Replacing $t$ by $t - 1$ in $g_1$, we get the real generating function

$$h_1(x,t) = \frac{2(1 + x(t - 1))}{1 + 2x(t - 1) + \sqrt{1 - 4x(1 + x(t - 1))}}.$$ 

In particular, for $t = 0$, we have

$$h_1(x,0) = \frac{2 - 2x}{1 - 2x + \sqrt{1 - 4x + 4x^2}}$$

$$= \frac{1 - x}{1 - 2x}$$

$$= 1 + \frac{x}{1 - 2x}$$

$$= 1 + \sum_{n=1}^{\infty} 2^{n-1}x^n. \quad (13)$$

This is the generating function of $DDU$-free Dyck paths with semilength weighted by $x$. There are $2^{n-1}$ different $DDU$-free Dyck paths with semilength $n$. We can see equation (13) directly, since the $DDU$-free Dyck paths can be written as the form

$$U^{a_1} DU^{a_2} DU^{a_3} \ldots U^{a_{k-1}} DU^{a_k} D^{n-k-1}$$
where \( a_i \) are positive integers, and \( \sum_{i=1}^{k} a_i = n \). Therefore, the number of the \( DDU \)-free Dyck paths with semilength \( n \) is equal to \( 2^{n-1} \), the number of compositions of \( n \), if \( n \geq 1 \). See Callan [5].

We can also verify that

\[
 h_1(x, t) = 1 + \frac{x}{1 - 2x} C \left( \frac{x^2 t}{(1 - 2x)^2} \right) \tag{14}
\]

where \( C(x) = \frac{1 - \sqrt{1 - 4x}}{2x} \), the Catalan number generating function.

Expanding each side of equation (14), we have for the left-hand side

\[
 h_1(x, t) = \frac{2(1 + x(t - 1))}{1 + 2x(t - 1) + \sqrt{1 - 4x(1 + x(t - 1))}}
 = \frac{2(1 + x(t - 1)) \left( 1 + 2x(t - 1) - \sqrt{1 - 4x(1 + x(t - 1))} \right)}{(1 + 2x(t - 1))^2 - (1 - 4x(1 + x(t - 1)))}
 = \frac{2(1 + x(t - 1)) \left( 1 + 2x(t - 1) - \sqrt{1 - 4x(1 + x(t - 1))} \right)}{4xt(1 + x(t - 1))}
 = \frac{1 + 2x(t - 1) - \sqrt{1 - 4x(1 + x(t - 1))}}{2xt}
\]

and the right-hand side

\[
 1 + \frac{x}{1 - 2x} C \left( \frac{x^2 t}{(1 - 2x)^2} \right) = 1 + \frac{x}{1 - 2x} \left( 1 - \sqrt{1 - 4 \left( \frac{x^2 t}{(1 - 2x)^2} \right)} \right)
 = 1 + \frac{1 - \sqrt{1 - \frac{4x^2 t}{(1 - 2x)^2}}}{2xt} \left( \frac{1 - 2x}{2} \right)
 = 1 + \frac{1 - \sqrt{1 - \frac{4x^2 t}{(1 - 2x)^2}}}{2xt} \left( \frac{1 - 2x}{2} \right)
\]

19
CHAPTER 2. OCCURRENCES OF SUBWORDS

\[ = 1 + \frac{1 - 2x - \sqrt{(1 - 2x)^2 - 4x^2t}}{2xt} \]

\[ = 1 + 2x(t - 1) - \sqrt{1 - 4x(1 + x(t - 1))} \]

\[ = h_1(x,t). \]

We can get an explicit formula for the coefficients of \( h_1(x,t) \) so that

\[ h_1(x,t) = 1 + \frac{x}{1 - 2x} C \left( \frac{x^2t}{(1 - 2x)^2} \right) \]

\[ = 1 + \frac{x}{1 - 2x} \sum_{k \geq 0} c_k t^k \left( \frac{x^2}{(1 - 2x)^2} \right)^k \]

\[ = 1 + x \sum_{k \geq 0} c_k t^k 2^{-2k} \frac{(2x)^{2k}}{(1 - 2x)^{2k+1}} \]

\[ = 1 + x \sum_{k \geq 0} c_k t^k 2^{-2k} \sum_{n \geq 0} \binom{n}{2k} (2x)^n \]

\[ = 1 + x \sum_{n \geq 0, k \geq 0} 2^{n-2k} \binom{n}{2k} c_k x^n t^k \]

\[ = 1 + \sum_{n \geq 1, k \geq 0} 2^{n-2k-1} \binom{n-1}{2k} \frac{1}{k+1} \binom{2k}{k} x^n t^k \]

where \( c_k \) is the \( k \)th Catalan number.

There is a bijective proof for this formula given by Callan [3]. Some related problems have been studied by Deutsch [8] and Sun [22].

2.6. Occurrences of DD

Now we look at an example of type 3. In these problems, we need to count paths with steps that go up by 1 and down by any amount. We cannot use the approach of applying the first return decomposition (see Figure 2.1), so we use another decomposition.
We consider paths with steps $U$, $D_0$, $D_1$, ..., $D_i$ where $U$ is an up step and $D_j$ is a step that goes down by $j$. Every nonempty such path can be factored as $G_1UG_2UG_3\cdots UG_{i+1}D_i$, where each $G_j$ is a path which ends on the same height as the height of its starting point and never goes below the height. See Figure 2.4.

![Figure 2.4. Decomposition for such path](image)

So the generating function for such paths $g$ satisfies

$$g = 1 + \sum_{i=0}^{\infty} u^i g^{i+1} d_i = 1 + g \sum_{i=0}^{\infty} (ug)^i d_i$$

(15)

where $u$ represents a single up step and $d_i$ represents a step that goes down by $i$.

Now consider the special case of counting Dyck paths by occurrences of $DD$ weighted by $t$. Look at the generating function for all possible reducing down steps, including a single down step and clusters consisting of $DD$'s:

$$D, (DD), (DDD), \ldots$$

So, the generating function for reducing down steps $d_i$ is

$$d + dt + dt^2 + \cdots = \frac{d}{1 - dt}.$$  

This is a problem in which we count paths reduced to paths with steps up by 1, and down by any amount. We may keep track of the height of paths. Let

$$\phi(z) = \sum_{i=0}^{\infty} z^i d_i = \frac{dz}{1 - dzt}.$$
CHAPTER 2. OCCURRENCES OF SUBWORDS

Then, we have

\[ g = 1 + g\phi(ug). \]

Replacing \( u \) by \( x \) and \( d \) by 1, we have

\[ g = 1 + \frac{xg^2}{1-xgt}. \]

Solving for \( g \), we get the generating function \( g \) with a weight \( 1 + t \)

\[ g = \frac{1 + xt - \sqrt{(1-xt)^2 - 4x}}{2x(t+1)}. \]

Using the cluster method, we replace \( t \) by \( t-1 \), and we get the real generating function \( h \) with a weight \( t \)

\[ h(x, t) = \frac{1 + x(t-1) - \sqrt{(1-x(t-1))^2 - 4x}}{2xt}. \] (16)

From equations (9) and (16), we can see that the generating functions are the same for counting paths by occurrences of \( DD \) and \( DU \), as is well known [7, 8].

2.7. Occurrences of DDD

Now we can look at another example of type 3. Suppose we count Dyck paths by occurrences of \( DDD \) weighted by \( t \). We can use the same approach as in the case of occurrences of \( DD \) in section 2.6. The generating function \( g \) with a weight \( 1 + t \) satisfies

\[ g = 1 + \sum_{i=0}^{\infty} u^i g^{i+1} d_i = 1 + g \sum_{i=0}^{\infty} (ug)^i d_i. \] (17)

The reducing down steps are a single down step and clusters consisting of \( DDD \)'s

\[ D, (DDD), (DDDD), (DDDDD), ... \]
So, the generating function for reducing down steps \( d_i \) is

\[
d + \frac{d^3 t}{1 - dt - d^2 t}.
\]

Let

\[
\phi(z) = dz + \frac{d^3 z^3 t}{1 - dzt - d^2 z^2 t} = \sum_{i=0}^{\infty} z^i d_i
\]

where \( d_i \) is the contribution from reducing steps that go down by \( i \).

From equation (17), we have

\[
g = 1 + g\phi(ug).
\]

Replacing \( u \) by \( x \) and \( d \) by 1, we have

\[
\phi(ug) = xg + \frac{x^3 g^3 t}{1 - xgt - x^2 g^2 t}.
\]

So,

\[
g = 1 + g \left( xg + \frac{x^3 g^3 t}{1 - xgt - x^2 g^2 t} \right).
\]

This looks cubic, but turns out to be quadratic.

Solving for \( g \), we get the generating function \( g \) with a weight \( 1 + t \)

\[
g = \frac{1 + xt - \sqrt{1 - 2xt + x^2 t^2 - 4x + 4x^2 t}}{2x(t + 1 - xt)}.
\]

Using the cluster method, we replace \( t \) by \( t - 1 \), and we get the real generating function \( h \) with a weight \( t \)

\[
h(x, t) = \frac{1 + xt - x - \sqrt{1 - 2xt + x^2 t^2 - 2x + 2x^2 t - 3x^2}}{2x(t - xt + x)}. \tag{18}
\]
CHAPTER 2. OCCURRENCES OF SUBWORDS

In particular, for $t = 0$, we have

$$h(x, 0) = \frac{1 - x - \sqrt{1 - 2x - 3x^2}}{2x^2}$$

$$= 1 + x + 2x^2 + 4x^3 + 9x^4 + 21x^5 + 51x^6 + 127x^7 + \cdots$$

This is the generating function of $DDD$-free Dyck paths with semilength weighted by $x$. This is also the generating function for the Motzkin numbers. There is a bijection from $UUU$-free Dyck $n$-paths to Motzkin $n$-paths given by Callan [3]. It is easy to see that $DDD$-free Dyck $n$-paths have the same distribution as $UUU$-free Dyck $n$-paths when writing a path in reverse order.

2.8. Occurrences of UUD and UDD

Now we will give an example of counting Dyck paths according to the occurrences of two subwords. Suppose they are $UUD$ weighted by $s$ and $UDD$ weighted by $t$. In this case, the only nontrivial cluster consisting of $UUD$ and $UDD$ is $UUDD$. This is an example of type 2. So we use the same approach as in the case of $UD$ in section 2.3. The generating function $g$ with weights $1 + s$ and $1 + t$ satisfies

$$g = 1 + fg + u_1gd_1g$$

$$= 1 + u^2d^2stg + (u + uuds)g(d + uddt)g$$

where $f = u^2d^2st$ represents a flat step, $u_1 = u + uuds$ represents up steps, $d_1 = d + uddt$ represents down steps.

Replacing $u$ by $x$ and $d$ by 1, and solving for $g$, we get

$$g = \frac{1 - x^2st - \sqrt{1 - 2x^2st + x^4s^2t^2 - 4x - 4x^2s - 4x^2t - 4x^3st}}{2x(1 + xs)(1 + xt)}$$
CHAPTER 2. OCCURRENCES OF SUBWORDS

\[
\begin{align*}
&= 1 - x^2 st - \sqrt{(1 - x^2 st)^2 - 4x(1 + xs)(1 + xt)} \\
&= \frac{2x(1 + xs)(1 + xt)}{1 - x^2 st - \sqrt{(1 - x^2 st)^2 - 4x(1 + xs)(1 + xt)}}.
\end{align*}
\]

Using the cluster method, we replace \(s\) by \(s - 1\) and \(t\) by \(t - 1\), and get the real generating function \(h\) with weights \(s\) and \(t\)

\[
h = \frac{1 - x^2(s - 1)(t - 1) - \sqrt{(1 - x^2(s - 1)(t - 1))^2 - 4x(1 - x + xs)(1 - x + xt)}}{2x(1 - x + xs)(1 - x + xt)}.
\]

In particular, letting \(s = t\) gives

\[
h(x, t, t) = \frac{1 + x - xt - \sqrt{(1 + x - xt)^2 - 4x}}{2x(1 - x + xt)}
\]

\[
= 1 + x + (1 + t^2)x^2 + (1 + 4t^2)x^3 + (1 + 10t^2 + 2t^3 + t^4)x^4 + \cdots
\]

Here the coefficients are sequence A127155 in the Online Encyclopedia of Integer Sequences [18], where they are described as the number of Dyck paths of semilength \(n\) having a total of \(k\) long ascents and long descents. It is easy to see that every Dyck path having a total of \(k\) long ascents and long descents has a total of \(k\) occurrences of \(UUD\) and \(UDD\), since every long ascent is followed by a down step \(D\) and every long descent is preceded with an up step \(U\).

2.9. Occurrences of UU and UDD

Now we give another example to count Dyck paths according to the occurrences of the given subwords \(UU\) weighted by \(s\) and \(UDD\) weighted by \(t\). The semilength is weighted by \(x\). This is an example in which we count paths reduced to paths with steps down by 1 or up by any amount. See Figure 2.5.

We consider paths with steps \(D_1, U_0, U_1, \ldots, U_i\) where \(D_1\) is the reducing down step that goes down by 1 and \(U_j\) is the step that goes up by \(j\). Every nonempty such path can be factored as \(U_iG_1D_1G_2D_1G_3\cdots G_iD_1G_{i+1}\), where each \(G_j\) is a path which
CHAPTER 2. OCCURRENCES OF SUBWORDS

ends on the same height as that of its starting point and never goes below the height. See Figure 2.5.

So the generating function \( g \) with weights \( 1+s \) for \( UU \) and \( 1+t \) for \( UDD \) satisfies

\[
g = 1 + \sum_{i=0}^{\infty} u_i g^{i+1} d_1^i
\]

where \( u_i \) represents reducing steps that go up by \( i \), and \( d_1 \) represents reducing down steps that go down by 1. The generating function for reducing down steps, including only a single down step and a \( UDD \) is

\[
d_1 = d + uddt.
\]

Therefore, we can see that the clusters consisting of \( UU \) and \( UDD \) are of the form \( U^j \)
for \( j = 2, 3, 4, \ldots \) or \( U^k DD \) for \( k = 1, 2, 3, \ldots \) So the generating function for reducing up steps, including a single up step and the clusters consisting of \( UU \) and \( UDD \), is obtained by subtracting \( d_1 \) from the generating function of all possible reducing steps:

\[
\begin{align*}
  u + d + uddt + \frac{u^2 s}{1 - us} (1 + ddt) - d_1 &= u + \frac{u^2 s}{1 - us} (1 + ddt) \\
  &= u + u^2 d^2 st \\
  &= \frac{1}{1 - us}.
\end{align*}
\]

So, let

\[
\phi(z) = \sum_{i=0}^{\infty} u_i z^i = \frac{uz + u^2 d^2 st}{1 - uz s}
\]
where $z$ is the weight for height.

Then we have

$$g = 1 + g\phi(gd_1).$$

Replacing $u$ by $x$, $d$ by 1, and $d_1$ by $1 + xt$, we get

$$\phi(gd_1) = \phi((1 + xt)g) = \frac{x(1 + xt)g + x^2st}{1 - x(1 + xt)gs}.$$ 

So,

$$g = 1 + \left(\frac{x(1 + xt)g + x^2st}{1 - x(1 + xt)gs}\right)g.$$ 

Simplifying, we get a quadratic equation

$$x(1 + xt)(s + 1)g^2 - (1 + xs)g + 1 = 0.$$ 

Solving for $g$, we get the generating function $g$ with weights $1 + s$ and $1 + t$

$$g = \frac{1 + xs - \sqrt{(1 + xs)^2 - 4x(1 + xt)(s + 1)}}{2x(1 + xt)(s + 1)}.$$ 

Using the cluster method, we replace $s$ by $s - 1$ and $t$ by $t - 1$, and get the real generating function $h$ with weights $s$ and $t$

$$h(x, s, t) = \frac{1 - x + xs - \sqrt{1 - 2xs - 2x + x^2s^2 + 2x^2s + x^2 - 4x^2st}}{2xs(1 - x + xt)}.$$ 

Notice that $h$ satisfies

$$1 - (1 - x + xs)h + xs(1 - x + xt)h^2 = 0.$$ 

Then, we have

$$h - xsh^2 = 1 + xh - xsh - x^2sh^2 + x^2sth^2.$$
CHAPTER 2. OCCURRENCES OF SUBWORDS

\[ h(1 - xsh) = (1 - xsh)(1 + xh) + x^2sth^2. \]

So, \( h \) also satisfies

\[ h = 1 + xh + \frac{x^2sth^2}{1 - xsh}. \]  

(20)

It would be interesting to find a direct proof of (20). We can do this using a decomposition of Deutsch \[6\]. See Figure 2.6.

\[ \begin{align*}
G_1 & \quad U \quad G_2 \quad U \quad \cdots \quad U \quad G_k \quad U \quad D \quad \cdots \\
& \quad U \quad \cdots \quad D \\
& \quad \cdots \\
& \quad D
\end{align*} \]

**Figure 2.6.** Deutsch’s decomposition for Dyck paths

Let \( G \) be a nonempty Dyck path. Suppose there are exactly \( k \) consecutive down steps after the last up step in \( G \). Then \( G \) can be factored uniquely as \( G_1UG_2U \cdots G_kUD^k \), where each \( G_j \) is a Dyck path. Let \( h \) be the generating function for counting all Dyck paths. Then this decomposition shows that \( h \) satisfies

\[ h = 1 + \sum_{k=1}^{\infty} (hu)^k d^k. \]

Now we count Dyck paths according to the occurrences of \(UU\) weighted by \( s\) and \(UDD\) weighted by \( t\). If \( k = 1 \), the occurrences of \(UU\) and \(UDD\) in \( G \) are the same as those in \( G_1 \). If \( k \geq 2 \), then every \(UU\) or \(UDD\) in each \(G_j\) occurs in \( G \). Moreover, for \(1 \leq j \leq k-1\) the \(U\) between \(G_j\) and \(G_{j+1}\) is followed by another \(U\), giving additional \(k - 1\) occurrences of \(UU\) and there is one extra \(UDD\) from the last up step followed by at least two down steps.
So the generating function $h$ with weights $s$ for $UU$ and $t$ for $UDD$ satisfies

$$h = 1 + hud + \sum_{k=2}^{\infty} s^{k-1} t(hu)^{k}d^{k}$$

Replacing $u$ by $x$ and $d$ by 1, we get

$$h = 1 + xh + \frac{x^2sth^2}{1-xsh}.$$ We can apply Lagrange inversion [20, Ch. 5, Page. 38] to get an explicit formula for the coefficients of $h(x, s, t)$.

**Theorem 2.** Let $h(x, s, t)$ be the generating function for counting Dyck paths by occurrences of $UU$ (weighted $s$) and $UDD$ (weighted $t$). Then

$$h(x, s, t) = \sum_{n,i,j} \frac{1}{n+1} \binom{n+1}{i+1,j,n-i-j} (i-1)^{n} s^{i} t^{j}$$

where the sum runs over all nonnegative integers for $n$, $i$, and $j$.

**Proof.** In order to apply Lagrange inversion, we can add a dummy variable $z$ to equation (20) getting

$$h = z \left( 1 + xh + \frac{x^2sth^2}{1-xsh} \right).$$ By Lagrange inversion [20, Ch. 5, Page. 38], we have

$$[z^n]h^k = \frac{k}{n} [y^{n-k}] \left( 1 + xy + \frac{x^2st}{1-xsy} \right)^n$$

$$= \frac{k}{n} [y^{n-k}] \sum_{i,j,m} (xy)^{i} \left( \frac{x^2st}{1-xsy} \right)^{j}$$

$$= \frac{k}{n} [y^{n-k}] \sum_{i,j,m,l} (xy)^{i} (x^2st)^{j} \left( \frac{x^2st}{1-xsy} \right)^{l} (j + l - 1) (xsy)^{l}$$
\[ = \frac{k}{n} \sum_{i,j,m,l} \binom{n}{i,j,m} x^m (x^2 st)^j \left( \frac{j + l - 1}{l} \right) (xs)^l. \]

Replacing \( n \) with \( i + j + m \) and \( l \) with \( n - k - 2j - m = i - k - j \), we get a power series for \( h^k \) in \( z \),

\[ h(x, s, t, z)^k = \sum_{i,j,m} \frac{k}{i + j + m} \binom{i + j + m}{i, j, m} \binom{i - k - 1}{i - k - j} x^m (x^2 st)^j (xs)^{i-k-j} z^n. \]

If we set \( z = 1 \),

\[ h(x, s, t)^k = \sum_{i,j,m} \frac{k}{i + j + m} \binom{i + j + m}{i, j, m} \binom{i - k - 1}{i - k - j} x^m (x^2 st)^j (xs)^{i-k-j}. \]

Replacing the variables, \( i \) by \( i + k \), and \( m \) by \( n - i - j \), we get

\[ h(x, s, t)^k = \sum_{n,i,j} \frac{k}{n+k} \binom{n+k}{i+k, j, n-i-j} \binom{i-1}{i-j} x^n s^j t^j. \]

In particular for \( k = 1 \), we have

\[ h(x, s, t) = \sum_{n,i,j} \frac{1}{n+1} \binom{n+1}{i+1, j, n-i-j} \binom{i-1}{i-j} x^n s^j t^j. \]  \hspace{1cm} (21)

Here, for all nonnegative integers \( n, i, \) and \( j \), the coefficient of \( x^n s^i t^j \) is nonzero and equal to

\[ \frac{1}{n+1} \binom{n+1}{i+1, j, n-i-j} \binom{i-1}{i-j} \]

for \( n \geq i + j \) and \( i \geq j \) and is 0 otherwise. \( \Box \)
CHAPTER 2. OCCURRENCES OF SUBWORDS

In equation (19), setting \( s = t \) gives

\[
h(x, t, t) = \frac{1 - x + xt - \sqrt{1 - 2xt - 2x + 2x^2t + x^2 - 3x^2t^2}}{2xt(1 - x + xt)}
= 1 + x + (1 + t^2)x^2 + (1 + 3t^2 + t^3)x^3 + (1 + 6t^2 + 4t^3 + 3t^4)x^4 + \cdots
\]

This is counting Dyck paths by the sum of occurrences of \( UU \) and \( UDD \) (weighted \( t \)). Here the coefficients are sequence A124926 in the Online Encyclopedia of Integer Sequences [18].

In particular, we can find the generating function for Dyck paths with the same semilength and sum of occurrences of \( UU \) and \( UDD \) by setting \( s = t \) and letting \( i = n - j \) in equation (21). We obtain

\[
\sum_{n,j} \frac{1}{n+1} \binom{n+1}{j} \binom{n-j-1}{n-2j} x^n t^n
\]

Here the coefficients are sequence A005043 in the Online Encyclopedia of Integer Sequences [18]. They are called the Riordan numbers (or ring numbers). Let \( r_n \) be the \( n \)-th Riordan number defined by

\[
r_n = \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \frac{1}{n+1} \binom{n+1}{j} \binom{n-j-1}{n-2j}
\]

with \( r_0 = 1 \).

The Riordan numbers count Motzkin paths containing no flatsteps at ground level. Deutsch [18] gave the interpretation that the Riordan number is equal to the number of Dyck paths of semilength \( n \) with no ascents of length 1 (an ascent in a Dyck path is a maximal string of up steps).

It is known that

\[
r_n + r_{n+1} = m_n \tag{22}
\]
where \( m_n \) is the \( n \)th Motzkin number. We can give a combinatorial interpretation to equation (22).

We know that \( m_n \) counts \( UDU \)-free Dyck paths with semilength \( n + 1 \) by the result in section 2.2. We can also see that a Dyck path with the same semilength and sum of occurrences of \( UU \) and \( UDD \) is equivalent to a Dyck path with no \( UDU \) that does not end in \( UD \), because every \( U \) must be followed by \( U \) or \( DD \).

Consider \( UDU \)-free Dyck paths with semilength \( n + 1 \). We can separate them into two cases.

1. Paths that end with \( UD \): Removing the \( UD \) at the end gives Dyck \( n \)-paths with no \( UDU \) that do not end with \( UD \). These are counted by \( r_n \). It is impossible to get Dyck \( n \)-paths with no \( UDU \) that do end with \( UD \), since we start from \( UDU \)-free Dyck paths with semilength \( n + 1 \). They are not allowed to end with \( UDUD \).
2. Paths that don’t end with \( UD \): These are counted by \( r_{n+1} \).

### 2.10. Occurrences of UUU and UDD

Now count Dyck paths according to the occurrences of two subwords, \( UUU \) weighted by \( s \) and \( UDD \) weighted by \( t \). The semilength is weighted by \( x \). We use the same approach as in the case in section 2.9. The generating function \( g \) with weights \( 1 + s \) for \( UUU \) and \( 1 + t \) for \( UDD \) satisfies

\[
g = 1 + \sum_{i=0}^{\infty} u_i g^{i+1} d_1^i
\]

where \( u_i \) represents reducing steps that go up by \( i \), and \( d_1 \) represents reducing down steps that go down by 1. The generating function for reducing down steps, including
CHAPTER 2. OCCURRENCES OF SUBWORDS

only a single down step and a \( UDD \) is

\[
d_1 = d + uddt.
\]

Therefore, we can see that the clusters consisting of \( UUU \) and \( UDD \) are of the form \( U^j \) for \( j = 3, 4, 5, \ldots \) or \( U^kDD \) for \( k = 1, 3, 4, 5, \ldots \). So, the generating function for reducing up steps, including a single up step and the clusters consist of \( UUU \) and \( UDD \), is obtained by subtracting \( d_1 \) from the generating function of all possible reducing steps:

\[
u + d + uddt + \frac{u^3s}{1 - us - u^2s}(1 + ddt) - d_1 = \frac{u^3s(1 + ddt)}{1 - us - u^2s} = \frac{u - u^2s + u^3d^2t}{1 - us - u^2s}.
\]

So, let

\[
\phi(z) = \sum_{i=0}^{\infty} u_iz^i = \frac{uz - u^2z^2s + u^3d^2zst}{1 - uz - u^2z^2s}.
\]

Then we have

\[
g = 1 + g\phi(gd_1).
\]

Replacing \( u \) by \( x \), \( d \) by 1, and \( d_1 \) by \( 1 + xt \), we get

\[
\phi(gd_1) = \phi((1 + xt)g) = \frac{x(1 + xt)g - x^2(1 + xt)^2g^2s + x^3(1 + xt)gst}{1 - x(1 + xt)gs - x^2(1 + xt)^2g^2s}.
\]

So,

\[
g = 1 + \left( \frac{x(1 + xt)g - x^2(1 + xt)^2g^2s + x^3(1 + xt)gst}{1 - x(1 + xt)gs - x^2(1 + xt)^2g^2s} \right) g.
\]

Simplifying, we get a quadratic equation

\[
x(1 + xt)(xs - s - 1)g^2 - (1 + xs + x^2st)g + 1 = 0.
\]
Solving for $g$, we get the generating function $g$ with weights $1 + s$ and $1 + t$
\[
g = \frac{1 + xs + x^2st - \sqrt{(1 + xs + x^2st)^2 - 4x(1 + xt)(xs - s - 1)}}{2x(1 + xt)(xs - s - 1)}.
\]
Using the cluster method, we replace $s$ by $s - 1$ and $t$ by $t - 1$, and get the real generating function $h$ with weights $s$ and $t$
\[
h(x, s, t) = \frac{1 - x + xs + x^2(s - 1)(t - 1) - \sqrt{(1 - x + xs + x^2(s - 1)(t - 1))^2 - 4x(1 - x + xt)(xs - x - s)}}{2x(1 - x + xt)(xs - x - s)}.
\]

2.11. Occurrences of $DU^k$DU

Barnabei, Bonetti, and Silimbani [1, Proposition 7] showed that the two statistics, number of occurrences of $DDD$ and number of occurrences of $DU^k$DU, where $k$ is any positive integer, are equidistributed on Dyck $n$-paths.

We have already found the generating function for the number of occurrences of $DDD$ in equation (18) in section 2.7. Now we use the cluster method to count Dyck paths according to the occurrences of $DU^k$DU weighted by $t$, where $k$ is any positive integer. The semilength is weighted by $x$. This is also an example in which we count paths reduced to paths with steps down by 1, and up by any amount. We use the same decomposition as used in Figure 2.5.

The clusters consisting of $DU^k$DU are

$DU^{k_1}DU^{k_2} \cdots DU^{k_j} \cdots DU$ where $k_j$ is any positive integer and $j = 1, 2, 3, \ldots$
CHAPTER 2. OCCURRENCES OF SUBWORDS

So the generating function for reducing up steps, including a single up step and the clusters consisting of $DU^kDU$ is

\[
\begin{align*}
  u + \frac{d \frac{u}{1-u} du}{1-d \frac{u}{1-u} t} &= u + \frac{u^2 d^2 t}{1-u - u dt}.
\end{align*}
\]

So, let

\[
\phi(z) = \infty \sum_{i=0} u_i z^i = uz + \frac{u^2 d^2 t}{1-uz - u dt}
\]

where $z$ is the weight for height. The generating function $g(x, t)$ with a weight $1 + t$ for $DU^kDU$ satisfies

\[
\begin{align*}
g &= 1 + \infty \sum_{i=0} u_i g^{i+1} d^i
\end{align*}
\]

where $u_i$ represents reducing steps that go up by $i$, and $d$ represents a single down step. Then we have

\[
g = 1 + g \phi(gd).
\]

Replacing $u$ by $x$, and $d$ by 1, we get

\[
\phi(gd) = \phi(g) = xg + \frac{x^2 t}{1-xg - xt}.
\]

So,

\[
g = 1 + g \left( xg + \frac{x^2 t}{1-xg - xt} \right).
\]

Solving for $g$, we get the generating function $g$ with weights $1 + t$

\[
g = \frac{1 - xt - \sqrt{1 - 2xt + x^2 t^2 + 4x^2t - 4x}}{2x}.
\quad (23)
\]

However, there is a height restriction whereby the clusters consisting of $DU^kDU$ cannot occur at height 0. By elevating $g$, we get a generating function $g_1(x, t)$ under
CHAPTER 2. OCCURRENCES OF SUBWORDS

the restriction which satisfies

\[ g_1 = \frac{1}{1 - gud}. \]  

Replacing \( u \) by \( x \) and \( d \) by 1 in equation (24), we substitute for \( g \) to get \( g_1 \). We get

\[ g_1 = \frac{1}{1 - xg} = \frac{2}{1 + xt + \sqrt{1 - 2xt + x^2t^2} + 4x^2t - 4x}. \]

Using the cluster method, we replace \( t \) by \( t - 1 \) and get the real generating function \( h_1(x, t) \) for occurrences of \( DU^k DU \)

\[ h_1(x, t) = \frac{2}{1 + xt - x - \sqrt{1 - 2xt + x^2t^2} - 2x + 2x^2t - 3x^2} \]

\[ = \frac{2}{1 + xt - x - \sqrt{1 - 2xt + x^2t^2 - 2x + 2x^2t - 3x^2} + 4xt - 4x^2t + 4x^2} \]

\[ = \frac{1 + xt - x - \sqrt{1 - 2xt + x^2t^2 - 2x + 2x^2t - 3x^2}}{2x(t - xt + x)}. \]  

From equation (18) and (25), we can see that the generating functions for the number of occurrences of \( DDD \) and the number of occurrences of \( DU^k DU \), where \( k \) is any positive integer, are the same.

2.12. Occurrences of UUU and UD

Now we want to count Dyck paths according to the occurrences of two subwords, \( UUU \) weighted by \( s \) and \( UD \) weighted by \( t \). The semilength is weighted by \( x \). This is also an example of type 3. In this problem, we need to count paths with steps that go down by 1 or up by any amount. We use the same approach as applied in section 2.9. We can see that the clusters consisting of \( UUU \) and \( UD \) are of the form \( U^j \) for \( j = 3, 4, 5, \ldots \) or \( U^k D \) for \( k = 1, 3, 4, 5, \ldots \) So, the generating function for reducing
up steps, including a single up step and the clusters consist of $UUU$ and $UD$, is
\[ u + \frac{u^3s}{1 - us - u^2s} + udt + \frac{u^3dst}{1 - us - u^2s} = (u + udt) \left( 1 + \frac{u^2s}{1 - us - u^2s} \right) = \frac{(u + udt)(1-us)}{1-us-u^2s}. \]

So the generating function $g$ for such path with weights $1+s$ for $UUU$ and $1+t$ for $UD$ satisfies
\[ g = 1 + \sum_{i=0}^{\infty} u_i g^{i+1} d^i \]
where $u_i$ represents reducing steps that go up by $i$, and $d$ represents down steps that go down by $1$. So, let
\[ \phi(z) = \sum_{i=0}^{\infty} u_i z^i = \frac{(uz + udt)(1 - uzs)}{1 - uzs - u^2z^2s}. \] (26)

Then we have
\[ g = 1 + g\phi(gd). \]

Replacing $u$ by $x$, $d$ by $1$, we get
\[ g = 1 + g \frac{(xg + xt)(1 - xgs)}{1 - xgs - x^2g^2s}. \]

Simplifying, we get a quadratic equation
\[ x(1 + s - xs - xst)g^2 - (1 + xs - xt)g + 1 = 0. \]

Solving for $g$, we get the generating function $g$ with weights $1+s$ and $1+t$
\[ g = \frac{1 + xs - xt - \sqrt{(1 + xs - xt)^2 - 4x(1 + s - xs - xst)}}{2x(1 + s - xs - xst)} = \frac{1 + xs - xt - \sqrt{1 - 2xs - 2xt + x^2s^2 + 2x^2st + x^2t^2 + 4x^2s - 4x}}{2x(1 + s - xs - xst)}. \] (27)
We replace $s$ by $s-1$, $t$ by $t-1$ in equation (27) and get the real generating function $h$ with weights $s$ for $UUU$ and $t$ for $UD$

$$h(x, s, t) = \frac{1 + xs - xt - \sqrt{1 - 2x^2s + 2x^2st + 4x^2t + xt^2}}{2x(s + xt - xst)}.$$

In particular, letting $s = t$ gives

$$h(x, t, t) = \frac{1 - \sqrt{1 - 4xt - 4x^2t^2 + 4x^2t^2}}{2x(t + xt - xt^2)}$$

$$= 1 + tx + (t + t^2)x^2 + (4t^2 + t^3)x^3 + (2t^2 + 11t^3 + t^4)x^4$$

$$+ (15t^3 + 26t^4 + t^5)x^5 + \cdots$$

Here the coefficients are sequence A091156 with rows reversed in the Online Encyclopedia of Integer Sequences [18]. In section 2.1 we counted Dyck paths by $UUD$ and got these coefficients in reverse order. We can see that the number of Dyck $n$-paths having exactly $k$ occurrences of $UUU$ is equal to the number of Dyck $n$-paths having a total of $n-k$ $UUU$ and $UD$, since the sum of the numbers of $UUD$, $UUU$, and $UD$ is always equal to the semilength $n$ in a Dyck path.

### 2.13. Occurrences of UUU and DU

Now count Dyck paths according to the occurrences of two subwords, $UUU$ weighted by $s$ and $DU$ weighted by $t$. This is an example of counting paths with steps that go down by 1 or up by any amount.

The clusters are almost the same as those in section 2.12. We can see that the clusters consisting of $UUU$ and $DU$ are of the form $U^j$ for $j = 3, 4, 5, \ldots$ or $DU^k$ for $k = 1, 3, 4, 5, \ldots$. However, we cannot use the same decomposition as the one in section 2.12. It is not allowed to have a cluster of the form $DU^k$ for $k = 1, 3, 4, 5, \ldots$ at height 0.
We discuss a more general path–counting problem as presented in section 2.9. We count paths reduced to paths with steps down by 1 or up by any amount. Let us consider the path with a step that goes up by \( i \) is weighted \( v_i \) if it starts on the \( x \)-axis and \( u_i \) if it starts at height \( > 0 \). Every nonempty such path with these weights can be factored as \( U_i G_1 D G_2 D \cdots G_i D G' \), where \( G' \) is a path with these weights and each \( G_j \) is a path with no height restriction (\( U_i \) is only weighted by \( u_i \)). See Figure 2.7.

![Figure 2.7. Decomposition for such path](image)

Let \( g' \) be the generating function for these paths. Then

\[
g' = 1 + \sum_{i=0}^{\infty} v_i (gd)^i g'.
\]

Let

\[
\psi(z) = \sum_{i=0}^{\infty} v_i z^i.
\]

Then

\[
g' = 1 + \psi(gd) g'.
\]

So

\[
g' = \frac{1}{1 - \psi(gd)}. \tag{28}
\]

Now we apply these formulas to the problem of counting such paths by \( UUU \) and \( DU \). We can see that the clusters consisting of \( UUU \) and \( DU \) are of the form \( U^j \) for \( j = 3, 4, 5, \ldots \) or \( DU^k \) for \( k = 1, 3, 4, 5, \ldots \) So applying the results of previous
paragraph, we have

\[ \phi(z) = \sum_{i=0}^{\infty} u_i z^i = \frac{(uz + dut)(1 - uz s)}{1 - uz s - u^2 z^2 s} \]

and the generating function \( g \) for each \( G_j \) which satisfies

\[ g = 1 + g\phi(gd). \]

From equation (26), we can see that the generating function \( g \) is the same as the \( g \) in section 2.12 (\( UUU \) and \( UD \)). However, it is not allowed that \( DU^k \) occur at height 0. So \( \psi(z) \) counts only clusters whose underlying is of the form \( U^j \) for \( j = 3, 4, 5, \ldots \) as applied in section 2.7 with \( D \) replaced by \( U \). So, the generating function for these reducing up steps is

\[ u + \frac{u^3 s}{1 - us - u^2 s}. \]

Then

\[ \psi(z) = \sum_{i=0}^{\infty} v_i z^i = uz + \frac{u^3 z^3 s}{1 - uz s - u^2 z^2 s}. \]

By equation (28) and replacing \( u \) by \( x \) and \( d \) by 1, we have

\[ g' = \frac{1}{1 - \psi(gd)} = \frac{1}{1 - (xg + \frac{x^3 g^3 s}{1 - xgs - x^2 g^2 s})} = \frac{1 - xgs - x^2 g^2 s}{1 - xg - xgs}. \]
We substitute $g$ in equation (27) for $g$ in equation (29) and simplify to get $g'$:

$$g' = 1 + xs + xt + 2xst - 2x^2st - 2x^2st^2$$

We replace $s$ by $s - 1$ and $t$ by $t - 1$, and get the real generating function $h$ with weights $s$ for $UUU$ and $t$ for $DU$:

$$h(x, s, t) = \frac{-\sqrt{1 - 2xs - 2xt + x^2s^2 + 2x^2st + x^2t^2 + 4x^2s - 4x}}{2x(1 + t)(1 + s - xs - xst)}.$$
CHAPTER 3

Counting Dyck Paths with Bounded Height

In the following examples, we apply the cluster method to count paths with bounded height by occurrences of subwords.

3.1. Occurrences of UD

We want to count Dyck paths of height at most $k$, by occurrences of peaks $UD$ weighted by $t$. We can use the same approach as in section 2.3. We can replace each peak by a new flat step $F$. Then this problem is equivalent to the problem of counting modified Motzkin paths of height at most $k$ with no flat step at height $k$. It is not allowed to have a flat step at height $k$ in these modified Motzkin paths, otherwise there will be a peak reaching height $k + 1$ in the original Dyck paths.

Let $g_k(x,t)$ be the generating function for such Motzkin paths with flat steps weighted $t$ which correspond to peaks weighted $1 + t$. We can use the same decomposition as in Figure 2.2 even with bounded heights. Then $g_k$ satisfies

$$g_{k+1} = 1 + fg_{k+1} + ug_kdg_{k+1}$$

and $g_0 = 1$

where $f$ represents a flat step $F$, $u$ represents a single up step $U$, and $d$ represents a single down step $D$. This can be written as

$$g_{k+1} = \frac{1}{1 - f - ug_kd}.$$  \hspace{1cm} (30)
Then $g_k$ can be written as a continued fraction

$$g_k = \frac{1}{1 - f - u - \frac{1}{1 - f - u - \frac{1}{\ddots - \frac{1}{1 - f - u - \frac{1}{1 - f - u}d}}}}$$

We define $p_k$ by the linear recurrence equation. We will show that $g_k = p_k/p_{k+1}$.

$$p_{k+2} = (1 - f)p_{k+1} - udp_k \quad \text{with} \quad p_0 = 1, p_1 = 1 \quad (31)$$

Let

$$G_k = \frac{p_k}{p_{k+1}}.$$ 

Dividing both sides of equation (31) by $p_{k+1}$, we get

$$\frac{1}{G_{k+1}} = 1 - f - udp_k \quad \text{with} \quad G_0 = 1.$$ 

Comparing this with equation (30) shows that $G_k = g_k$. Then let

$$P(z) = \sum_{k=0}^{\infty} p_k z^k.$$ 

We can solve equation (31) to get

$$P(z) = \frac{1 + z - (1 - f)z}{1 - (1 - f)z + udz^2}. \quad (32)$$
CHAPTER 3. COUNTING DYCK PATHS WITH BOUNDED HEIGHT

We replace \( f \) by \( xt \), \( u \) by \( x \), and \( d \) by 1 to get

\[
P(z) = \frac{1 + z - (1 - xt)z}{1 - (1 - xt)z + xz^2}.\]

Using the cluster method, we replace \( t \) by \( t - 1 \) to get

\[
\hat{P}(z) = \frac{1 + z - (1 - x(t - 1))z}{1 - (1 - x(t - 1))z + xz^2} = 1 + z + (1 - xt)z^2 + (1 - 2xt - x^2t + x^2t^2)z^3 + \cdots
\]

\[
= \sum_{k=0}^{\infty} \hat{p}_k z^k.
\]

Then the real generating function \( h_k \) for Dyck paths with heights at most \( k \) and weight \( t \) for \( UD \) is given by

\[
h_k(x, t) = \frac{\hat{p}_k}{\hat{p}_{k+1}}.
\]

The formula for counting Dyck paths of bounded height as a quotient of these polynomials is well known.

The first few values for \( \hat{p}_k \) are

\[
\begin{align*}
\hat{p}_0 &= 1 \\
\hat{p}_1 &= 1 \\
\hat{p}_2 &= 1 - xt \\
\hat{p}_3 &= 1 - 2xt - x^2t + x^2t^2 \\
\hat{p}_4 &= 1 - 3xt - 2x^2t - x^2t - x^3t^2 + 2x^3t^2 - x^3t^3 \\
\hat{p}_5 &= 1 - 4xt - 3x^2t - 6x^2t^2 - 2x^3t - 6x^3t^2 - 4x^3t^3 - x^4t + 3x^4t^2 - 3x^4t^3 + x^4t^4.
\end{align*}
\]
We can find an explicit formula for \( \hat{p}_k \). We start from subtracting 1 from \( \hat{P}(z) \), then we get

\[
\hat{P}(z) - 1 = \frac{1 - xz + xtz}{1 - z - xz + xtz + xz^2} - 1
= \frac{z - xz^2}{1 - z - xz + xz^2 + xtz}
= \frac{z(1 - xz)}{(1 - z)(1 - xz) + xtz}
= \frac{z}{(1 - z)} \cdot \frac{1}{1 + \frac{xtz}{(1 - z)(1 - xz)}}
= \frac{z}{1 - z} \sum_{i=0}^{\infty} (-1)^i \frac{(xtz)^i}{(1 - z)^i(1 - xz)^i}
= \sum_{i=0}^{\infty} (-1)^i \frac{x^i t^i z^{i+1}}{(1 - z)^{i+1}(1 - xz)^i}
= \sum_{i,l,m} (-1)^i x^i t^i z^{i+1} \binom{i + l}{l} z^l \binom{i + m - 1}{m} (xz)^m
= \sum_{i,l,m} (-1)^i \binom{i + l}{l} \binom{i + m - 1}{m} x^{i+m} t^i z^{i+1+l+m}.
\]

Replacing the variables, \( m \) by \( n - i \) and \( l \) by \( k - n - 1 \), we get

\[
\hat{P}(z) - 1 = \sum_{n,i,k} (-1)^i \binom{i + k - n - 1}{i} \binom{n - 1}{n - i} x^n t^i z^k.
\]

So, for \( k \geq 1 \), we have

\[
\hat{p}_k = \sum_{n=0}^{k-1} \sum_{i=0}^{n} (-1)^i \binom{i + k - n - 1}{i} \binom{n - 1}{n - i} x^n t^i.
\]
Note that setting $t = 1$ gives

$$\hat{p}_k = \sum_{n=0}^{k-1} \sum_{i=0}^{n} (-1)^i \binom{i + k - n - 1}{i} \binom{n - 1}{n - i} x^n$$

$$= \sum_{n=0}^{k-1} \sum_{i=0}^{n} \binom{n - k}{i} \binom{n - 1}{n - i} x^n$$

$$= \sum_{n=0}^{k-1} \binom{2n - k - 1}{n} x^n$$

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

$$= \sum_{n=0}^{k-1} (-1)^n \binom{k - n}{n} x^n$$

$$= \sum_{n=0}^{\lfloor \frac{k}{2} \rfloor} (-1)^n \binom{k - n}{n} x^n.$$  

Let us look more closely at the case of Dyck paths of height at most 2. Here we have

$$h_2(x, t) = \frac{\hat{p}_2}{\hat{p}_3}$$

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

$$= \frac{1}{1 - xt} \cdot \frac{1}{1 - \frac{x^2t}{(1 - xt)^2}}$$

$$= \sum_{n=0}^{\infty} \frac{(x^2t)^n}{(1 - xt)^{2n+1}}$$

$$= \sum_{n=0}^{\infty} x^{2n} t^n \sum_{i=0}^{\infty} \binom{2n + i}{i} x^i t^i$$

$$= \sum_{n, i} \binom{2n + i}{i} x^{2n + i} t^{n+i}.$$
Replacing the variables $l$ by $2n + i$ and $m$ by $n + i$, we get

$$h_2(x, t) = \sum_{l, m} \left( \frac{l}{2l - 2m} \right) x^l t^m.$$ (33)

We can give a combinatorial interpretation for equation (33). We can start from a Dyck path with semilength $l$ and exactly $m$ peaks and height at most 2. We add an extra down step in front of it and an extra up step after it. The modified path can be decomposed as

$$(DU)^{i_1} (UD)^{i_2} (DU)^{i_3} (UD)^{i_4} \cdots (DU)^{i_{2k+1}}$$

where each $i_s$ is a positive integer and $i_1 + i_2 + \cdots + i_{2k+1} = l + 1$. Conversely, any path with such a decomposition is a modified path from a Dyck path with semilength $l$ and height at most 2. See Figure 3.1. We call the components $(DU)^{i_{2j+1}}$ odd components and components $(UD)^{i_{2j}}$ even components. We can think of counting these paths as counting compositions of $l + 1$ with $2k + 1$ parts. We know that the number of such paths is the number of compositions of $l + 1$ with $2k + 1$ parts, which is $\binom{l}{2k}$.

For example, the corresponding composition for the path in Figure 3.1 is

$$9 = 3 + 2 + 2 + 1 + 1.$$

We want to find the connection between $k, l$ and the number $m$ of peaks. An odd component $(DU)^i$ contributes $i$ to the semilength and $i - 1$ to the number of peaks. On the other hand, an even component $(UD)^j$ contributes $j$ to the semilength and also $j$ to the number of peaks. So the difference between the sum $i_1 + i_2 + \cdots + i_{2k+1}$ and the number of peaks is $k + 1$, the number of odd components. Therefore, we have $l + 1 - m = k + 1$. So $k = l - m$ and thus, the number of Dyck path with semilength
CHAPTER 3. COUNTING DYCK PATHS WITH BOUNDED HEIGHT

Figure 3.1. Decomposition for modified paths

Let $g_k(x, t)$ be the generating function for such paths with bounded height $k$. The clusters are of the form $U(DU)^i$ for $i = 1, 2, 3, \ldots$. 

3.2. Occurrences of UDU

Count Dyck paths with bounded height by occurrences of $UDU$ weighted by $t$. We can use the same approach as in section 2.2. This problem is equivalent to the problem of counting modified Dyck paths of height at most $k$.

Let $g_k(x, t)$ be the generating function for such paths with bounded height $k$. The clusters are of the form $U(DU)^i$ for $i = 1, 2, 3, \ldots$. 

$l$ and $m$ peaks and height at most 2 is

$$\binom{l}{2k} = \binom{l}{2l - 2m}.$$
CHAPTER 3. COUNTING DYCK PATHS WITH BOUNDED HEIGHT

So, the cluster generating function is

\[ udut + ududut^2 + udududut^3 + \cdots = \frac{u^2 dt}{1 - u dt}. \]

Since these clusters reduce to up steps, we can set \( u_1 = u + \frac{u^2 dt}{1 - u dt} = \frac{u}{1 - u dt} \) to get the generating function \( g_k(x, t) \) with a weight \( 1 + t \) which satisfies

\[ g_{k+1} = 1 + u_1 g_k d g_{k+1} \quad \text{and} \quad g_0 = 1. \]

This can be written as

\[ g_{k+1} = \frac{1}{1 - u_1 g_k d}. \quad (34) \]

Then \( g_k \) can be written as a continued fraction

\[ g_k = \frac{1}{1 - u_1 \frac{1}{1 - u_1 \frac{1}{1 - u_1 \ddots \frac{1}{1 - u_1 \frac{1}{1 - u_1 g_0 d}}}}} \]

We want to find \( g_k \) and we can use a similar approach as in section 3.1. Define \( p_k \) by the linear recurrence equation as equation (31). So setting \( f = 0 \) and replacing \( u \) with \( u_1 \) in equation (32), we get \( g_k = p_k / p_{k+1} \) where

\[ P(z) := \sum p_k z^k = \frac{1}{1 - z + \frac{x}{1 - xt} z^2}. \]
This continued fraction is the same as the previous one in section 3.1 with $f = 0$ and $u$ replace with $u_1$. So

$$g_k = \frac{p_k}{p_{k+1}}$$

and

$$P(z) = \sum_{k=0}^{\infty} p_k z^k.$$ 

Then we can solve a linear recurrence equation by substituting 0 for $f$ and $u_1$ for $u$ in equation (32) to get

$$P(z) = \frac{1}{1 - z + u_1 dz^2}.$$ 

We replace $u_1$ by $\frac{x}{1 - xt}$ and $d$ by 1 to get

$$P(z) = \frac{1}{1 - z + \frac{x}{1 - xt} z^2}.$$ 

Using the cluster method, we replace $t$ by $t - 1$ in $P(z)$ to get

$$\hat{P}(z) = \sum_{k=0}^{\infty} \hat{p}_k z^k$$

$$= \frac{1}{1 - z + \frac{x}{1 - xt + x^2} z^2}$$

$$= 1 + z + \frac{1 - xt}{1 - xt + x^2} z^2 + \frac{1 - xt - x}{1 - xt + x^3} z^3 + \ldots$$

Then the real generating function $h_k$ for Dyck paths with heights at most $k$ and weight $t$ for $UDU$ is given by

$$h_k(x, t) = \frac{\hat{p}_k}{\hat{p}_{k+1}}$$

where $\hat{p}_k$ is given by the coefficient of $\hat{P}(z)$. 

50
The first few values for $\hat{p}_k$ are

$\hat{p}_0 = 1$

$\hat{p}_1 = 1$

$\hat{p}_2 = \frac{1 - xt}{1 - xt + x} = 1 - \frac{x}{1 - xt + x}$

$\hat{p}_3 = \frac{1 - xt - x}{1 - xt + x} = 1 - 2 \left( \frac{x}{1 - xt + x} \right)$

$\hat{p}_4 = \frac{1 - 2xt + x^2t^2 - x + x^2t - x^2}{(1 - xt + x)^2} = 1 - 3 \left( \frac{x}{1 - xt + x} \right) + \left( \frac{x}{1 - xt + x} \right)^2$

$\hat{p}_5 = \frac{1 - 2xt + x^2t^2 - 2x + 2x^2t}{(1 - xt + x)^2} = 1 - 4 \left( \frac{x}{1 - xt + x} \right) + 3 \left( \frac{x}{1 - xt + x} \right)^2$

We can find an explicit formula for $\hat{p}_k$:

$\hat{P}(z) = \frac{1}{1 - z + \frac{x}{1 - xt + x} z^2}$

$= \sum_{l=0}^{\infty} \left( \frac{-x z^2}{1 - xt + x} + z \right)^l$

$= \sum_{l,m} \binom{m}{l} (-1)^l z^{2l} \left( \frac{x}{1 - xt + x} \right)^l z^{m-l}$

$= \sum_{l,m} (-1)^l \binom{m}{l} \left( \frac{x}{1 - xt + x} \right)^l z^{m+l}$

Replacing the variables, $m$ by $k - l$, we get

$\hat{P}(z) = \sum_{l,k} (-1)^l \binom{k-l}{l} \left( \frac{x}{1 - xt + x} \right)^l z^k.$
For the coefficients of $\hat{P}(z)$, if $l > \frac{k}{2}$, we have

$$\binom{k - l}{l} = 0.$$  

So, we have

$$\hat{p}_k = \sum_{l=0}^{\left\lfloor \frac{k}{2} \right\rfloor} (-1)^l \binom{k - l}{l} \left( \frac{x}{1 - xt + x} \right)^l.$$  

In particular, for $k = 2$, we have

$$h_2(x,t) = \frac{\hat{p}_2}{\hat{p}_3} \hat{p}_3$$

$$= \frac{1 - xt}{1 - xt - x}$$

$$= 1 + \frac{x}{1 - x(t + 1)}$$

$$= 1 + \sum_{n=1}^{\infty} x^n(t + 1)^{n-1}$$

$$= 1 + \sum_{n=1}^{\infty} \sum_{m=0}^{n-1} \binom{n - 1}{m} x^n t^m. \quad (35)$$

We can give a combinatorial interpretation for equation (35). We can start from a Dyck path with semilength $n$ and exactly $m$ occurrences of $UDU$ and height at most 2. We can use the same decomposition as Figure 3.1. We add an extra down step in front of the path and an extra up step after it. The modified path can be decomposed as

$$(DU)^{i_1}(UD)^{i_2}(DU)^{i_3}(UD)^{i_4} \cdots (DU)^{i_{2k+1}}$$

where each $i_s$ is a positive integer and $i_1 + i_2 + \cdots + i_{2k+1} = n + 1$.

Conversely, any path with such a decomposition is a modified path from a Dyck path with semilength $n$ and height at most 2. We can think of counting these paths
as counting compositions of \( n + 1 \) with \( 2k + 1 \) parts. We know that the number of such paths is the number of compositions of \( n + 1 \) with \( 2k + 1 \) parts, which is \( \binom{n}{2k} \).

Then, we want to find the connection between \( k, n \) and the number \( m \) of occurrences of \( UDU \). An even component \((UD)^j\) contributes \( j \) to the semilength and \( j - 1 \) to the occurrences of \( UDU \).

On the other hand, an odd component \((DU)^i\) contributes \( i \) to the semilength and \( i - 1 \) to the occurrences of \( UDU \) except the last odd component. Therefore, we can separate into two cases.

If \( i_{2k+1} = 1 \), the last odd component \((DU)^{i_{2k+1}}\) contributes \( i_{2k+1} \) to the semilength and \( i_{2k+1} - 1 \) to the occurrences of \( UDU \). So the difference between the sum \( i_1 + i_2 + \cdots + i_{2k+1} \) and the number of occurrences of \( UDU \) is \( 2k + 1 \). Therefore, we have \( n + 1 - m = 2k + 1 \). The compositions we are counting are compositions of \( n + 1 \) with \( 2k + 1 \) parts in which the last part is 1. Deleting the last part gives a composition of \( n \) with \( 2k \) parts, and here are \( \binom{n-1}{2k-1} \) of them.

So \( 2k = n - m \) and

\[
\binom{n - 1}{2k - 1} = \binom{n - 1}{n - m - 1} = \binom{n - 1}{m}.
\]

Note that this only applies when \( n - m \) is even.

If \( i_{2k+1} \geq 2 \), the last odd component \((DU)^{i_{2k+1}}\) contributes \( i_{2k+1} \) to the semilength but \( i_{2k+1} - 2 \) to the occurrences of \( UDU \), since the last \( UDU \) in the modified path is not in the original path. So the difference between the sum \( i_1 + i_2 + \cdots + i_{2k+1} \) and the number of \( UDU \) is \( 2k + 2 \). Therefore we have \( n + 1 - m = 2k + 2 \). The compositions we are counting are compositions of \( n + 1 \) with \( 2k + 1 \) parts in which
the last part is at least 2. Subtracting 1 from the last part gives a composition of \( n \) with \( 2k + 1 \) parts, and here are \( \binom{n - 1}{2k} \) of them.

So \( 2k = n - m - 1 \) and

\[
\binom{n - 1}{2k} = \binom{n - 1}{n - m - 1} = \binom{n - 1}{m}.
\]

Note that this only applies when \( n - m \) is odd.

Thus, combining the results of these two cases, we get that the number of Dyck path with semilength \( n \) and \( m \) occurrences of \( UDU \) and height at most 2 is

\[
\binom{n - 1}{m}.
\]
CHAPTER 4

Applications to $r$-Dyck paths

In the following examples, we apply the cluster method to count paths with up steps $U$ that go up by 1 and down steps $D$ that go down by an arbitrary number, $r$. We define an $r$-Dyck path to be a path with up steps $U$ that go up by 1 and down steps $D$ that go down by $r$.

4.1. Occurrences of UD

We count $r$-Dyck paths by occurrences of $UD$ (weighted $t$). In this case, the only cluster is $UD$. We may use the same approach as section 2.6. We consider paths with steps that go up by 1, $U$, down by $r$, $D_r$, and down by $r - 1$, $D_{r-1}$. Here a $U$ in such a path corresponds to a $U$ in a $r$-Dyck path, weighted by $x$, a $D_r$ in such a path corresponds to a $D$ in a $r$-Dyck path, weighted by 1, and a $D_{r-1}$ in such a path corresponds to a $UD$ in a $r$-Dyck path, weighted by $t$. So, by equation (15) that we obtained in section 2.6, the generating function for such paths $g$ satisfies

$$g = 1 + u^{r-1}g^rd_{r-1} + u^r g^{r+1}d_r.$$ 

Replacing $u$ by $x$, $d_r$ by 1, and $d_{r-1}$ by $xt$, we get

$$g = 1 + x^{r-1}g^rxt + x^rg^{r+1}$$

$$= 1 + x^rg^rt + x^rg^{r+1}.$$
CHAPTER 4. APPLICATIONS TO $r$-DYCK PATHS

Using the cluster method, we replace $t$ by $t - 1$ and get the real generating function $h(x, t)$ which satisfies

$$h = 1 + x^r h^r (t - 1) + x^r h^{r+1}. \quad (36)$$

Let

$$h = 1 + tH.$$ 

We substitute this in equation (36) to get

$$1 + tH = 1 + x^r (1 + tH)^r (t - 1) + x^r (1 + tH)^{r+1}$$

$$tH = x^r (1 + tH)^r (t - 1 + 1 + tH)$$

$$H = x^r (1 + tH)^r (1 + H).$$

Set

$$x^r = z.$$ 

By Lagrange inversion [20, Ch. 5, Page. 38], we have

$$[z^n] H^k = \frac{k}{n} \left[ y^{n-k} \right] \left( (1 + ty)^r (1 + y) \right)^n$$

$$= \frac{k}{n} \left[ y^{n-k} \right] (1 + ty)^{nr} (1 + y)^n$$

$$= \frac{k}{n} \left[ y^{n-k} \right] \sum_{i,j} \binom{nr}{i} t^i y^j \binom{n}{n-j} y^j$$

$$= \frac{k}{n} \left[ y^{n-k} \right] \sum_{i+j=n-k} \binom{nr}{i} \binom{n}{n-j} t^i y^{i+j}$$

$$= \frac{k}{n} \sum_{i} \binom{nr}{i} \binom{n}{i+k} t^i.$$
CHAPTER 4. APPLICATIONS TO $r$-DYCK PATHS

So, we get

$$H^k = \sum_{n,i} k \binom{nr}{i} \binom{n}{i+k} t^i z^n$$

$$= \sum_{n,i} k \binom{nr}{i} \binom{n}{i+k} x^{nr} t^i.$$ 

For $k = 1$, we have

$$H(x, t) = \sum_{n,i} \frac{1}{n} \binom{nr}{i} \binom{n}{i+1} x^{nr} t^i.$$ 

Then we have

$$h(x, t) = 1 + tH$$

$$= 1 + \sum_{n \geq 1} \sum_{i \geq 0} \frac{1}{n} \binom{nr}{i} \binom{n}{i+1} x^{nr} t^{i+1}.$$ 

For $r = 2$, these numbers are A108767 or A120986 in the Online Encyclopedia of Integer Sequences [18]

4.2. Occurrences of UU and UDD

Count $r$-Dyck paths by occurrences of $UU$ (weighted $s$) and $UDD$ (weighted $t$), since we got an interesting result for $r = 1$ in section 2.9. We may use the same approach as section 2.9 to find the generating function. However, we can find a similar equation to equation (20) for the generating function by using Deutsch’s decomposition. See Figure 4.1.

Figure 4.1. Deutsch’s decomposition for $r$-Dyck paths

57
Let $G$ be a nonempty $r$-Dyck path. Suppose there are exactly $k$ consecutive down steps after the last up step in $G$. Then $G$ can be factored uniquely as $G_1 U G_2 U \cdots G_k U D^k$, where each $G_j$ is a $r$-Dyck path. Let $h$ be the generating function for counting all $r$-Dyck paths. Then this decomposition shows that $h$ satisfies
\[ h = 1 + \sum_{k=1}^{\infty} (hu)^{kr} d^k. \]

Now we assign weights $s$ for occurrences of $UDD$ and $t$ for occurrences of $UDD$. If $k = 1$, the occurrences of $UU$ in $G$ are the same as those in $G_1$ to $G_r$, but for the occurrences of $UU$ in $G$, the $U$ between $G_j$ and $G_{j+1}$ is followed by another $U$ for $1 \leq j \leq r$, giving additional $r - 1$ occurrences of $UU$. If $k \geq 2$, then every $UU$ or $UDD$ in each $G_j$ occurs in $G$. Moreover, for $1 \leq j \leq kr - 1$, the $U$ between $G_j$ and $G_{j+1}$ is followed by another $U$, giving additional $kr - 1$ occurrences of $UU$. There is also one extra $UDD$ from the last up step followed by at least two down steps. So the generating function $h$ with weights $s$ for $UU$ and $t$ for $UDD$ satisfies
\[ h = 1 + s^{r-1}(hu)^r d + \sum_{k=2}^{\infty} s^{kr-1} t(hu)^{kr} d^k. \]

Replacing $u$ by $x$ and $d$ by $1$, we get
\[ h = 1 + s^{r-1}(xh)^r + \sum_{k=2}^{\infty} s^{kr-1} t(xh)^{kr}. \]

Let
\[ H = sxh. \]

We multiply both sides of equation (37) by $sx$ and substitute $H$ for $sxh$
\[ H = x \left( s + H^r + t \frac{H^{2r}}{1 - H^r} \right). \]
CHAPTER 4. APPLICATIONS TO r-DYCK PATHS

By Lagrange inversion [20, Ch. 5, Page. 38], we have

\[ [x^n] H^k = \frac{k}{n} [y^{n-k}] \left( s + y^r + t \frac{y^{2r}}{1 - y^r} \right)^n \]

\[ = \frac{k}{n} [y^{n-k}] \sum_{i,j,l \atop n=i+j+l} \left( \frac{n}{i,j,l} \right) s^i \left( \frac{t \frac{y^{2r}}{1 - y^r}}{y^l} \right)^j (y^r)^l \]

\[ = \frac{k}{n} [y^{n-k}] \sum_{i,j,l,m \atop n=i+j+l} \left( \frac{n}{i,j,l} \right) s^i t^j y^l y^{2jr} \left( \frac{j + m - 1}{m} \right) (y^r)^m \]

For this to be nonzero, \( n - k \) must be a multiple of \( r \). Let \( n - k = pr \). Thus

\[ [x^n] H^k = \frac{k}{pr + k} [y^{pr}] \sum_{i,j,l,m \atop pr+k=i+j+l} \left( \frac{pr+k}{i,j,l} \right) \left( \frac{j + m - 1}{m} \right) s^i t^j y^l y^{2jr+2j+m} \]

\[ = \frac{k}{pr + k} \sum_{i,j,l,m \atop pr+k=i+j+l} \left( \frac{pr+k}{i,j,l} \right) \left( \frac{j + m - 1}{m} \right) s^i t^j \]

\[ = \frac{k}{pr + k} \sum_{i,j,l,m \atop pr+k=i+j+l} \left( \frac{pr+k}{i,j,l} \right) \left( \frac{p - j - l - 1}{p - l - 2j} \right) s^i t^j \]

\[ = \frac{k}{pr + k} \sum_{i,j,l \atop pr+k=i+j+l} \left( \frac{pr+k}{i,j} \right) \left( \frac{p - pr - k + i - 1}{p - pr - k + i - j} \right) s^i t^j. \]

So

\[ H^k = \sum_{i,j,p \atop n=pr+k} \frac{k}{pr + k} \left( \frac{pr+k}{i,j,pr+k-i-j} \right) \left( \frac{p - pr - k + i - 1}{p - pr - k + i - j} \right) x^n s^i t^j. \]
For $k=1$, we have

$$H = \sum_{i,j,p \atop n=pr+1} \frac{1}{pr+1} \left( \binom{pr+1}{i,j,pr+1-i-j} \binom{p-pr+i-2}{p-pr+i-j-1} \right) x^{pr+1} s^i t^j.$$  

Therefore, the generating function for $r$-Dyck paths counting by occurrences of $UU$ (weighted $s$) and $UDD$ (weighted $t$) is

$$h(x, s, t) = \sum_{i,j,p \atop pr+1} \frac{1}{pr+1} \left( \binom{pr+1}{i,j,pr+1-i-j} \binom{p-pr+i-2}{p-pr+i-j-1} \right) x^{pr+1} s^i t^j$$

$$= \sum_{i,j,p \atop pr+1} \frac{1}{pr+1} \left( \binom{pr+1}{i+1,j,pr-i-j} \binom{p-pr+i-1}{p-pr+i-j} \right) x^{pr} s^i t^j$$  \hspace{1cm} (38)$$

Here, for all nonnegative integers $p$, $i$, and $j$, the coefficient of $x^{pr} s^i t^j$ is nonzero and equal to

$$\frac{1}{pr+1} \left( \binom{pr+1}{i+1,j,pr-i-j} \binom{p-pr+i-1}{p-pr+i-j} \right)$$

for $pr \geq i + j$ and $i \geq j$ and is 0 otherwise.

In particular for $r = 1$, equation (38) reduces to equation (21) which counts for Dyck paths by occurrences of $UU$ and $UDD$. 

60
Bibliography