Studying the Area Under Generalized Dyck Paths

AJ Bu

Rutgers University

March 7, 2024

Introduction to Dyck and Motzkin Paths

Two well-known types of paths:

- A **Motzkin path** of length n is a path in the xy-plane from the origin to $(n, 0)$ with steps in $\{(1,1), (1,0), (1,-1)\}$ that never goes below the x-axis.

  We call

$$\begin{aligned} U &:= (1,1) \text{ an up step}, \\ F &:= (1,0) \text{ a flat step, and} \\ D &:= (1,-1) \text{ a down step} \end{aligned}$$

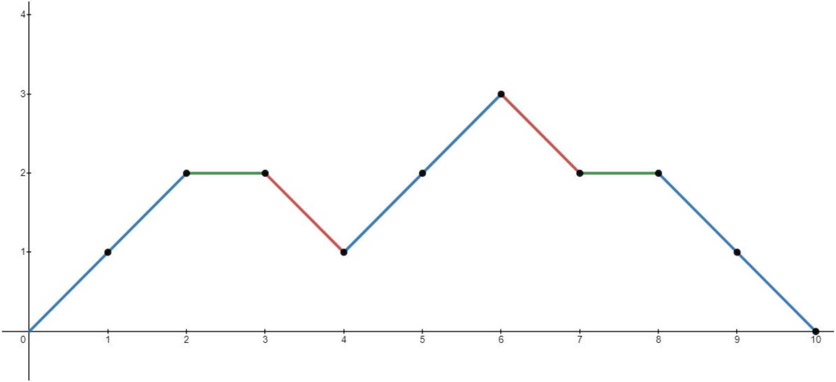
- A **Dyck path** is a Motzkin path that avoids flat steps.

Example

The following is a Motzkin path of length 10

$$UUFDUUDFDD$$

Example:



Natural Question: How do we enumerate Motzkin paths?

Use **weight enumerator**:

$$P(t) = \sum_{W \in \mathcal{P}} t^{Length(W)}$$

$$P(t) = 1 + tP(t) + t^2\left[P(t)\right]^2.$$

Let $\mathcal{P}$ denote the set of all Motzkin paths.
Then $\mathcal{P}$ is generated by

$$\mathcal{P} = \{EmptyPath\} \cup F\mathcal{P} \cup U\mathcal{P}D\mathcal{P}.$$

Therefore, the enumerator of each of these gives us

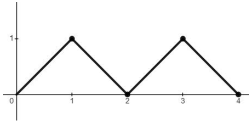
$$P = 1 + tP + t^2P^2.$$

Area Under Motzkin Paths

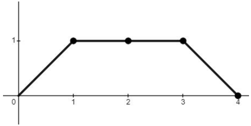
To keep track of area under the paths in $\mathcal{P}$, as well as the number of paths, we use the following bi-variate weight enumerator:

$$P(t,q) = \sum_{W \in \mathcal{P}} t^{Length(W)} q^{AreaUnder(W)}$$

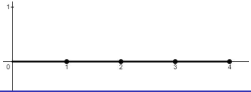
E.g.



UDUD has weight $t^4 q^2$



UFFD has weight $t^4 q^3$



FFFF has weight t^4

Area Under Motzkin Paths

$$\mathcal{P} = \{EmptyPath\} \cup F\mathcal{P} \cup U\mathcal{P}D\mathcal{P}$$

Note that for

$$M = FM_0,$$

both M and M_0 have the same area.

We, however, need to make adjustments for

$$M = UM_1DM_0.$$

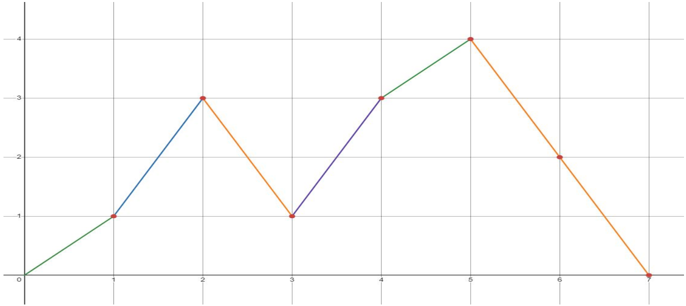
1. The total area under the steps U and D is 1
2. The area under the Motzkin path M_0 is equal to the area under the portion of M that it represents
3. Since M_1 is shifted to height 1, however, every step in M_1 has one more unit block below it.

$$\implies M(t,q) = 1 + tM(t,q) + t^2qM(t,q)M(qt,q).$$

A **generalized Dyck path** is a path in the $xy-$plane from the origin $(0,0)$ to $(n,0)$ with an arbitrary set of atomic steps and that never go below the $x-$axis.

E.g. A generalized Dyck path with steps in $S = \{1, 2, -1, -2\}$.

$$[1, 2, -2, 2, 1, -2, -2]$$



Joint Work with Doron Zeilberger:

Paper: Using Symbolic Computation to Explore Generalized Dyck Paths and Their Areas (posted on arXiv)

Accompanying Maple Package: `GDW.txt`
(Link is found in paper. Also posted on both of our websites)

1. Use symbolic programming to generate $F(t, X)$ s.t. $F(t, P) = 0$, where $P(t)$ is the weight-enumerator for the generalized Dyck paths with steps in a given set S.
2. Make an analogous method that keeps track of area as well

E.g. Generalized Dyck paths with steps in $S = \{1, 2, -1, -2\}$

Using our Maple procedure,

```
EqGFt({1,2,-1,-2},P,t)
```

outputs

$$1 + (-2t - 1)P + t(3t + 2)P^2 - t^2(2t + 1)P^3 + P^4 t^4.$$

Notation

First let's introduce the following notation:

$$\mathcal{P}_{a,b} = \text{the set of generalized Dyck paths with a set of steps given by } S \text{ that start at } (0, a) \text{ and end at height } b,$$

$$P_{a,b}(t) = \text{the desired weight-enumerator for the paths in } \mathcal{P}_{a,b}.$$

$$\mathcal{Q}_{a,b} = \text{the subset of } \mathcal{P}_{a,b} \text{ that contains all non-empty paths that stay strictly above the } x-\text{axis, except at an endpoint if } a = 0 \text{ or } b = 0,$$

$$Q_{a,b}(t) = \text{the desired weight-enumerator for the paths in } \mathcal{Q}_{a,b}.$$

General Idea

- Begin with $\mathcal{P}_{0,0}$
- Get new equations and variables by breaking the paths down into a concatenation of legal steps and sub-paths with various starting and ending heights
  - Use the enumerating function for the "children" to get the enumerator for the original set
  - Sometimes, we will replace a child set with one that has the same number of elements but is easier to work with.
- Repeat this whole process with each child set until no more children are produced.
- Assigning different variables to each of these sets gives us our system of equations
- We can then use Gröbner bases to get $P(t)$

A Brief Summary of Gröbner Bases

A **Göbner basis** of an ideal $I \subset k[x_1, ..., x_n]$ is a finite subset $G = \{g_1, ..., g_t\}$ of I such that, for every nonzero polynomial f in I, f is divisible by the leading term of g_i for some i.

The Gröbner basis simplifies solving the ideal membership problem and finding solutions to a system of polynomial equations.

A polynomial f lies in the ideal $I \subset k[x_1, ..., x_n]$ with Gröbner basis G if and only if the remainder on division of f by G is zero.

Example of Process: $P_{0,0}(t)$

Suppose $0 \in S$. We want to find $P_{0,0}(t)$.

- $EmptyPath \in \mathcal{P}_{0,0}$
- If the path begins with the flat step, then we have

$$F\mathcal{P}_{0,0}$$

- Otherwise, we begin with a positive step, and the path must return to the x-axis for a first time. We will split our path into two sub-paths at this point

$$\mathcal{Q}_{0,0}\mathcal{P}_{0,0}$$

$$\implies \qquad \mathcal{P}_{0,0} = \{EmptyPath\} \cup F\mathcal{P}_{0,0} \cup \mathcal{Q}_{0,0}\mathcal{P}_{0,0}$$

$$\implies \qquad P_{0,0} = 1 + t \cdot P_{0,0} + Q_{0,0} \cdot P_{0,0}$$

Example of Process: $Q_{0,0}(t)$

Now we want to find $Q_{0,0}(t)$

First, let us introduce the following notation:

Let the set U give the legal upward steps and D give the legal downward steps

e.g. For $S = \{1, 2, -1, -2\}$, our legal steps are

Up steps: $u_1 =$ up 1 unit and $u_2 =$ up 2 units

$\implies U = \{1, 2\}$

Down steps: $d_1 =$ down 1 unit and $d_2 =$ down 2 units

$\implies D = \{1, 2\}$

Example of Process: $Q_{0,0}(t)$

- Let the set U give the legal upward steps and D give the legal downward steps
- **Legal Initial Steps:** u_k s.t. $k \in U$

  Separating this step leaves a path that starts at height k
- **Legal final steps:** d_ℓ s.t. $\ell \in D$

  Separating this step leaves a path that ends at height ℓ

$$\implies \mathcal{Q}_{0,0} = \bigcup_{k \in U} \bigcup_{\ell \in D} u_k \left[\mathcal{Q}_{k,\ell} \right] d_\ell$$

- Shifting the paths in $\mathcal{Q}_{k,\ell}$ down by 1 unit creates a bijection with $\mathcal{P}_{k-1,\ell-1}$

$$\implies Q_{0,0}(t) = t^2 \sum_{k \in U} \sum_{\ell \in D} P_{k-1,\ell-1}(t)$$

E.g. Generalized Dyck paths with steps in $\{1, 2, -1, -2\}$

$$U = \{1, 2\} \qquad\qquad D = \{1, 2\}$$

Legal initial steps: $u_1 = 1$ and $u_2 = 2$

$$\Longrightarrow \qquad \mathcal{Q}_{0,0} = u_1 \ \mathcal{Q}_{1,0} \quad \cup \quad u_2 \ \mathcal{Q}_{2,0}$$

Legal final steps: $d_1 := -1$ and $d_2 := -2$

$$\mathcal{Q}_{0,0} = u_1 \ \mathcal{Q}_{1,1} \ d_1 \ \cup \ u_1 \ \mathcal{Q}_{1,2} \ d_2 \ \cup \ u_2 \ \mathcal{Q}_{2,1} \ d_1 \ \cup \ u_2 \ \mathcal{Q}_{2,2} \ d_2$$

$$\Longrightarrow \qquad Q_{0,0} = t^2 \cdot Q_{1,1} + t^2 \cdot Q_{1,2} + t^2 \cdot Q_{2,1} + t^2 \cdot Q_{2,2}$$

Bijections:

$$\mathcal{Q}_{1,1} \longleftrightarrow \mathcal{P}_{0,0} \qquad \Longrightarrow \qquad Q_{1,1}(t) = P_{0,0}(t)$$

$$\mathcal{Q}_{1,2} \longleftrightarrow \mathcal{P}_{0,1} \qquad \Longrightarrow \qquad Q_{1,2}(t) = P_{0,1}(t)$$

$$\mathcal{Q}_{2,1} \longleftrightarrow \mathcal{P}_{1,0} \qquad \Longrightarrow \qquad Q_{2,1}(t) = P_{1,0}(t)$$

$$\mathcal{Q}_{2,2} \longleftrightarrow \mathcal{P}_{1,1} \qquad \Longrightarrow \qquad Q_{2,2}(t) = P_{1,1}(t)$$

$$\Longrightarrow \qquad Q_{0,0} = t^2 \cdot P_{0,0} + t^2 \cdot P_{0,1} + t^2 \cdot P_{1,0} + t^2 \cdot P_{1,1}$$

E.g. Generalized Dyck paths with steps in $\{1, 2, -1, -2\}$

Keep doing this until no more new "children" are produced

E.g. Our procedure

```
MakeSyst(P,Q,t, { 1,2,-1,-2 } )
```

gives the system of equations:

$$
\begin{aligned}
P_{0,0} &= P_{0,0}Q_{0,0} + 1, \\
P_{0,1} &= P_{0,0}Q_{0,1}, \\
P_{1,0} &= Q_{1,0}P_{0,0}, \\
P_{1,1} &= P_{0,1}Q_{1,0} + P_{0,0}, \\
Q_{0,0} &= t^2P_{0,0} + t^2P_{0,1} + t^2P_{1,0} + t^2P_{1,1}, \\
Q_{0,1} &= tP_{0,0} + tP_{1,0}, \\
Q_{1,0} &= tP_{0,0} + tP_{0,1}
\end{aligned}
$$

with variables $\{P_{0,0}, P_{0,1}, P_{1,0}, P_{1,1}, Q_{0,0}, Q_{0,1}, Q_{1,0}\}$

Area Under Generalized Dyck Paths

We can modify our method of enumerating generalized Dyck paths to keep track of the total area.

e.g. Before we had

$$\mathcal{Q}_{0,0} = \bigcup_{k\in U} \bigcup_{\ell \in D} u_k \ \mathcal{Q}_{k,\ell} \ d_\ell$$

$$\implies \qquad Q_{0,0}(t) = t^2 \sum_{k\in U} \sum_{\ell\in D} P_{k-1,\ell-1}(t)$$

Now, considering area, we have...

$$Q_{0,0}(t,q) = t^2 \sum_{k\in U} \sum_{\ell\in D} q^{k/2+\ell/2} P_{k-1,\ell-1}(qt,q).$$

E.g. Generalized Dyck paths with steps in $\{1, 2, -1, -2\}$

We showed that

$$\mathcal{Q}_{0,0} = u_1\ \mathcal{Q}_{1,1}\ d_1\ \cup\ u_1\ \mathcal{Q}_{1,2}\ d_2\ \cup\ u_2\ \mathcal{Q}_{2,1}\ d_1\ \cup\ u_2\ \mathcal{Q}_{2,2}\ d_2$$
$$Q_{0,0}(t) = t^2 \cdot P_{0,0}(t) + t^2 \cdot P_{0,1}(t) + t^2 \cdot P_{1,0}(t) + t^2 \cdot P_{1,1}(t)$$

Area under steps:

- Area under u_1 = Area under $d_1 = \frac{1}{2}$
- Area under u_2 = Area under $d_2 = 1$

$$Q_{0,0}(t,q) = qt^2 P_{0,0}(qt,q) + q^{3/2} t^2 P_{0,1}(qt,q) + q^{3/2} t^2 P_{1,0}(qt,q) + q^2 t^2 P_{1,1}(qt,q)$$

E.g. Our procedure

```
qMakeSyst(P,Q,t,q,{1,2,-1,-2} )
```

gives the following system of **functional** equations.

$$
\begin{aligned}
P_{0,0}(t,q) &= P_{0,0}(t,q)Q_{0,0}(t,q) + 1,\\
P_{0,1}(t,q) &= P_{0,0}(t,q)Q_{0,1}(t,q),\\
P_{1,0}(t,q) &= Q_{1,0}(t,q)P_{0,0}(t,q),\\
P_{1,1}(t,q) &= P_{0,1}(t,q)Q_{1,0}(t,q) + P_{0,0}(t,q),\\
Q_{0,0}(t,q) &= qt^2 \cdot P_{0,0}(qt,q) + q^{3/2}t^2 \cdot P_{0,1}(qt,q)\\
&\qquad + q^{3/2}t^2 \cdot P_{1,0}(qt,q) + q^2t^2 \cdot P_{1,1}(qt,q),\\
Q_{0,1}(t,q) &= q^{1/2}t \cdot P_{0,0}(qt,q) + qt \cdot P_{1,0}(qt,q),\\
Q_{1,0}(t,q) &= q^{1/2}t \cdot P_{0,0}(qt,q) + qt \cdot P_{0,1}(qt,q)
\end{aligned}
$$

Solving the System of Functional Equations

After the computer finds the system of *functional* equations described above, we instruct it to find a system *algebraic* equations for the 'components' of the $P_{a,b}(t, q)$ (and we also need $Q_{a,b}(t, q)$).

To do this, use:

- The Taylor Series expansions about $q = 1$:

$$P_{a,b}(t, q) = \sum_{n=0}^{\infty} \frac{1}{n!} \left[\frac{d^n}{dq^n} P_{a,b}(t, q) \right]_{q=1} (q-1)^n.$$

- Use the following **Lemma:**
  If $f(t)$ is the formal power series of a single variable t, and q is another variable, then

$$f(qt) = \sum_{n=0}^{\infty} \frac{1}{n!} t^n \left[\frac{d^n}{dt^n} f(t) \right] (q-1)^n.$$

Solving the System of Functional Equations

Let $P'_{a,b}(t,1)$ denote $\frac{d}{dq}P_{a,b}(t,q)\Big|_{q=1}$.

The generating function for the sum of the areas of all legal walks of length n is

$$P'(t,1)$$

- Rewrite all our $P_{a,b}(t,q)$ and $Q_{a,b}(t)$ as

$$P_{a,b}(t,q) = P_{a,b}(t,1) + (q-1) \ \cdot \ P'_{a,b}(t,1) + O((q-1)^2), \quad \text{and}$$
$$Q_{a,b}(t,q) = Q_{a,b}(t,1) + (q-1) \ \cdot \ Q'_{a,b}(t,1) + O((q-1)^2)$$

- We expand in powers of $q-1$ then collect terms
- Use lemma on previous slide and get more equations by differentiating with respect to t each of these equations using implicit differentiation

E.g. Generalized Dyck paths with steps in $\{1, 2, -1, -2\}$

E.g. Our procedure

```
qEqGFt({1,2,-1,-2} ,P,t )
```

gives

$$
\begin{aligned}
20736P^4t^{10} &- 2304P^4t^9 - 6560P^4t^8 - 10368P^3t^9 + 1520P^4t^7 \\
&- 23328P^3t^8 + 465P^4t^6 + 3848P^3t^7 + 3888P^2t^8 \\
&- 184P^4t^5 + 9530P^3t^6 + 17352P^2t^7 + 16P^4t^4 \\
&- 2290P^3t^5 - 429P^2t^6 - 648Pt^7 - 878P^3t^4 \\
&- 8914P^2t^5 - 2214Pt^6 + 352P^3t^3 + 2289P^2t^4 \\
&- 970Pt^5 + 81t^6 - 32P^3t^2 + 704P^2t^3 + 2295Pt^4 \\
&- 144t^5 - 324P^2t^2 - 628Pt^3 + 358t^4 + 32P^2t \\
&- 122Pt^2 - 168t^3 + 72Pt + 24t^2 - 8P
\end{aligned}
$$

Area Under Generalized Dyck Paths

Say we know bi-variate polynomials $f(t,q)$, $g(t,q)$, and $h(t,q)$ s.t.

$$P(t,q) = f(t,q) + g(t,q) \cdot P(t,q) + h(t,q) \cdot P(t,q) \cdot P(qt,q).$$

We can solve for $P'(t,1)$, which gives the total area under the paths of length n.

Note: Rather than outputting algebraic equations, as seen earlier, we now produce closed-form expressions in terms of radicals

We can also solve for higher order derivatives:

$$P^{(k)}(t,1) = \frac{d^k}{dq^k} P(t,q)\Big|_{q=1}$$

Area Under Generalized Dyck Paths

Paper: Explicit Generating Functions for the Sum of the Areas Under Dyck and Motzkin Paths (and for Their Powers)

(Posted on arXiv as well as my website)

Accompanying Maple Package: `qEW.txt`

(Link in paper as well as on my website)

Brief Description of Process:

1. Plug in $q = 1$
2. Solve for $P(t, 1)$
3. Using Taylor series about $q = 1$ and comparing the coefficients of $(q-1)^n$, we can solve for $P^{(n)}(t, 1)$
   - Express $P^{(n)}(t, 1)$ as the sum of derivatives $P^{(k)}(t, 1)$ where $k < n$ and derivatives of functions of t with respect to t
   - Since we have $P(t, 1)$, we can simply compute any order derivative with respect to t as well as $P'(t, 1)$
   - To find $P^{(n)}(t, 1)$, repeat this process with the coefficient of $(q-1)^k$ to get $P^{(k)}(t, 1)$ for $k = 1, \ldots n$

Demonstrate this Process with the Motzkin Paths

$$M(t,q) = 1 + t\,M(t,q) + t^2\,q\,M(qt,q)\,M(t,q).$$

1. Plugging in $q = 1$, we get

$$M(t,1) = 1 + t\,M(t,1) + t^2\;[M(t,1)]^2.$$

2. Solving for $M(t,1)$:

$$M(t,1) = \frac{1 - t \pm \sqrt{-3t^2 - 2t + 1}}{2t^2}$$

3. $M(t,1)$ is the enumerator for Motzkin paths of length n and has a Taylor series expansion about $t = 0$. Thus

$$M(t,1) = \frac{1 - t - \sqrt{-3t^2 - 2t + 1}}{2t^2}$$

Area Under Motzkin Paths: Finding $M_q(t,1)$

$$\sum_{k=0}^{n} \frac{(q-1)^k}{k!} M^{(k)}(t,1)$$

$$= 1 + t \sum_{k=0}^{n} \frac{(q-1)^k}{k!} M^{(k)}(t,1)$$

$$+ qt^2 \sum_{k=0}^{n} \frac{(q-1)^k}{k!} M^{(k)}(t,1) \sum_{k=0}^{n} \frac{(q-1)^k}{k!} M^{(k)}(qt,1).$$

The coefficient of $(q-1)$ on both sides gives:

$$M_q(t,1) = t\, M_q(t,1) + t^2\, M(t,1)\Big(t\, M_t(t,1) + 2M_q(t,1) + M(t,1) \Big).$$

Area Under Motzkin Paths

$$M_q(t,1) = \frac{t^3 M(t,1) M_t(t,1) + t^2 M^2(t,1)}{1 - t - 2t^2 M(t,1)}.$$

We know $M(t,1)$ and can solve for $M_t(t,1)$ by taking the derivative.

Plugging these in, we get:

$$M_q(t,1) = \frac{\left(t - 1 + \sqrt{-3t^2 - 2t + 1}\right)^2}{4t^2(-3t^2 - 2t - 1)}$$

To find $M^{(n)}(t,1)$, we can repeat this process with the coefficient of $M^{(k)}(t,1)$ for $k \leq n$.

Now that we have the derivatives...

We can then look at the Maclaurin series of these function to get some pretty interesting information! For example:

1. $M(t,1)$ is the weight enumerator of Motzkin paths of length n

   $$1 + t + 2t^2 + 4t^3 + 9t^4 + 21t^5 + 51t^6 + 127t^7 + 323t^8 + O(t^9)$$

2. $M_q(t,1)$ is the weight enumerator of the total area under all Motzkin paths of length n

   $$t^2 + 4t^3 + 16t^4 + 56t^5 + 190t^6 + 624t^7 + 2014t^8 + 6412t^9 + O(t^{10})$$

3. $M_{qq}(t,1) + M_q(t,1)$ is the weight enumerator for the sum of the squares of the areas of Motzkin paths of length n

   $$t^2 + 6t^3 + 40t^4 + 198t^5 + 910t^6 + 3848t^7 + 15492t^8 + 59920t^9 + O(t^{10})$$

   We can also do this for higher powers

Further Study

Look at average areas and the variance.

- Given a family of paths let

$$\begin{aligned} a_0(n) =& \text{ the number of such paths of length } n, \\ a_1(n) =& \text{ the total area under such paths of length } n, \\ a_2(n) =& \text{ the sum of the squares of the areas under such paths} \\ & \text{ of length } n \end{aligned}$$

- Using `qEW.txt`, we can generate 10,000 (or more) terms of the sequences of:
  - The average areas $\left\{ \frac{a_1(n)}{a_0(n)} \right\}$
  - The variances $\left\{ \frac{a_2(n)}{a_0(n)} - \left(\frac{a_1(n)}{a_0(n)} \right)^2 \right\}$