# Symbolic Computation to Study Explicit Gröbner Bases and Lattice Path Enumeration

AJ Bu

Rutgers University

February 29, 2024

# Experimental Mathematics

Experimental mathematics involves using computation and algorithms to study mathematical objects, typically with computer-assisted proving.

The two methods in my dissertation:

1. Writing a code that allows the computer to automatically generate the desired mathematical object
2. Using such codes to produce sufficiently many examples, studying these examples, and identify (then proving) patterns among them

# Projects in Dissertation

My dissertation contains the following projects:

1. Finding the Gröbner bases of ideals generated by elementary symmetric polynomials
2. Enumerating infinite families of restricted Dyck paths
3. Enumerating restricted Motzkin paths
4. Enumerating generalized Dyck paths
5. Enumerating generalized Dyck paths with respect to area

# A Brief Summary of Gröbner Bases

Many of my projects use Gröbner bases.

A **Gröbner 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.

# Introduction to Dyck and Motzkin 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

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

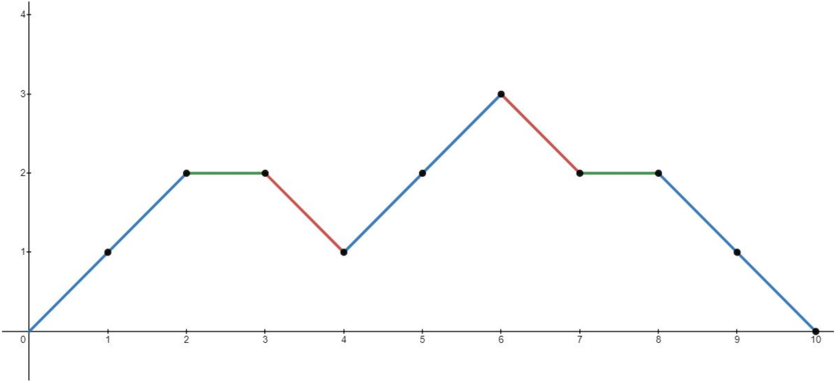
A **Dyck path** is a Motzkin path that avoids flat steps.

# Example

The following is a Motzkin path of length 10

$$UUFDUUDFDD$$

Example:



# Natural Question: How do we enumerate Motzkin paths?

Use **weight enumerator**:

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

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

Let $\mathcal{P}$ denote the set of all Motzkin paths.
Then $\mathcal{P}$ is generated by the unambiguous, context-free grammar

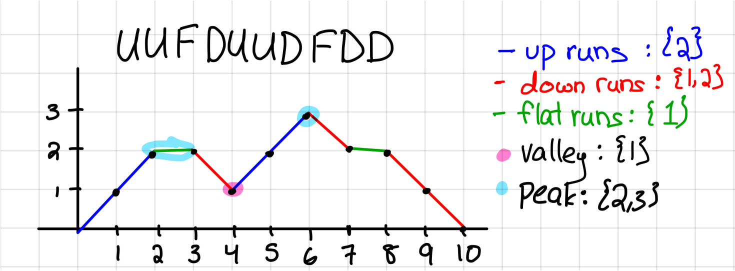
$$\mathcal{P} = \{EmptyPath\} \cup F\mathcal{P} \cup U\mathcal{P}D\mathcal{P}.$$

Therefore, the enumerator of each of these gives us

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

What if we apply restrictions on run lengths, peak and valley heights?

Going back to our example:



# Results from Zeilberger

Created Maple package to find $F(t, X)$ s.t. $F(t, P) = 0$, where $P$ is the weight enumerator of the set of Dyck paths avoiding

- peak heights in $A$
- valley heights in $B$
- upward runs in $C$
- downward runs in $D$,

where $A$, $B$, $C$, and $D$ are given finite sets containing integers and/or arithmetic progressions.

Programs limited to finding generating functions for one specific set of restrictions

**Goals:**

1. Can we generalize these findings to Motzkin paths?
2. Can we find identities on the weight enumerator functions for an infinite family of such $\{A, B, C, D\}$?
   - Eg: $C = \{ar + b : r \in \mathbb{N}\}$ for arbitrary $a, b \in \mathbb{N}$

# Automatically Enumerating Restricted Motzkin Paths

Generalized Zeilberger's procedures to Motzkin paths:

Find $F(t, P)$ directly by generating a finite system of algebraic equation using symbolic dynamic programming and solving it for $F(t, P)$.

- Get new equations and variables by recursively breaking paths down into a concatenation of steps and sub-paths with simpler restrictions
- Use Gröbner bases to get $F(t, P)$

# Example Results

The generating function $P(t)$ of the sequence of Motzkin paths with the following restrictions satisfies the given algebraic equation:

**Avoiding up-runs of length 1, 2, or 3**

$$1 - (t^2 - t + 1)P - t^2(t - 1)P^2 + t^8P^4 + t^9P^5 = 0$$

**Avoiding odd peak heights and valley heights**

$$(t - 1)^2 + (t - 1)^3P + t^4P^2 = 0$$

Note: Can simultaneously restrict multiple characteristics, where each forbids values given in finite sets containing integers and/or arithmetic progressions

# e.g. Avoiding Peak Heights in $A$ and Valley Heights in $B$

Let $A, B \subset \mathbb{N}$ be finite. Consider the set $\mathcal{P}_{A,B}$ of Motzkin paths avoiding

- peak-heights in A, and
- valley-heights in B.

Let $P_{A,B}(t)$ denote the enumerator for $\mathcal{P}_{A,B}$

Consider the following three cases:

1. $0 \in A$
2. $0 \notin A$ and $0 \in B$
3. $0 \notin A$ and $0 \notin B$

# Peak at Height 0

Note: The set $\mathcal{F}$ of walks containing only flat-steps contains

$$\{\text{Empty Path}\}, F, FF, FFF, FFFF, \ldots$$

Thus, the weight-enumerator of $\mathcal{F}$ is

$$\sum_{n=0}^{\infty} t^n = \frac{1}{1-t}.$$

Say a path has a peak at height 0 if and only if it is a flat run.

# Case 1: $0 \in A$

Let $A_1 := A \backslash \{0\}$.

Then

$$\mathcal{P}_{A,B} \cup \mathcal{F} = \mathcal{P}_{A_1,B}.$$

$$\implies P_{A,B}(t) = P_{A_1,B}(t) - \frac{1}{1-t}.$$

# Case 2: $0 \notin A$ and $0 \in B$

Let $A_1 := \{a - 1 \mid a \in A\}$ and $B_1 := \{b - 1 | b \in B \backslash \{0\}\}$.

Then

$$\mathcal{P}_{A,B} = \mathcal{F} \cup \mathcal{F} U \mathcal{P}_{A_1,B_1} D \mathcal{F}$$

$$\implies P_{A,B}(t) = \frac{1}{1-t} + \frac{t^2}{(1-t)^2} P_{A_1,B_1}(t).$$

# Case 3: $0 \notin A$ and $0 \notin B$

Let $A_1 := \{a - 1 \mid a \in A\}$ and $B_1 := \{b - 1 \mid b \in B\}$.

Then

$$\mathcal{P}_{A,B} = \mathcal{F} \cup \mathcal{F} U \mathcal{P}_{A_1,B_1} D \mathcal{P}_{A,B}.$$

$$\implies P_{A,B}(t) = \frac{1}{1-t} + \frac{t^2}{1-t} P_{A,B}(t) P_{A_1,B_1}(t)$$

# Avoiding Peak Heights in $A$ and Valley Heights in $B$

Applying recursively we will:

- Remove all elements of $A$ and $B$
- Get a finite system of equations
- Eliminate all variables except the ones representing our original $P_{A,B}$

Easy to extend to arithmetic progressions $ar + b$ where $a, b \in \mathbb{N}$.

Just use same method on the constant term of the arithmetic progression.

Also wrote programs for avoiding up-runs, down-runs, and flat-runs. Due to time constraints, I will not discuss them here.

# Context-Free Grammars

**Joint work with Robert Dougherty-Bliss:**

Given a family of Dyck paths $\mathcal{P}$, how do we find an equation satisfied by the weight enumerator $P(t)$?

**The Basic Strategy:**

1. Identify a context-free grammar that generates $\mathcal{P}$
   - Often begin with attempt to subdivide elements of $\mathcal{P}$ into multiple shorter elements of $\mathcal{P}$
   - Add additional terms to accommodate corner cases
2. Translate grammar into equation by replacing every set of $2k$ steps with $t^k$, and every instance of $\mathcal{P}$ with $P(t)$

# Grammatical Family Example

**Theorem 1**

Let $b < a$ be non-negative integers, and let $\mathcal{P}$ be the set of Dyck paths whose up-run lengths avoid $\{ar + b \mid r \geq 0\}$. Then

$$P(t) = \sum_{\substack{0 \leq k \leq a-1 \\ k \neq b}} t^k \left[P(t)\right]^k + t^a P^{a+1}(t),$$

where $P(t)$ is the weight-enumerator of $\mathcal{P}$.

# Constructing the Bijection

**First look at simple cases to get an idea -**

Avoid upwards run of length $2r+1$:

$$P = t^2P^3 + 1$$
$$\implies \mathcal{P} = \{EmptyPath\} \cup U^2(\mathcal{P}D)^2\mathcal{P}$$

Avoid upwards run of length $3r+1$:

$$P = P^4t^3 + P^2t^2 + 1$$
$$\implies \mathcal{P} = \{EmptyPath\} \cup U^3(\mathcal{P}D)^3\mathcal{P} \cup U^2(D\mathcal{P})^2$$

# Constructing the Bijection

Divide $\mathcal{P}$ into subfamilies based on length of initial up-run

- Length $k$ for some $0 \leq k \leq a-1$:
  - Always contains empty path
  - If $k = b$, set of paths is $\emptyset$
  - Otherwise, we have: $U^k(D\mathcal{P})^k$
    - $i$th $D$: first time path returns to height $k-i$ after initial up-run
    - Every up-run in $\mathcal{P}$-subpath is up-run in full path
- Length $\geq a$: $U^a\mathcal{P}(D\mathcal{P})^a$
  - Initial up-run is $U^a$ plus initial up-run in first subpath

Add together subfamilies:

$$\mathcal{P} = \bigcup_{\substack{0 \leq k \leq a-1 \\ k \neq b}} U^k(D\mathcal{P})^k \cup U^a(\mathcal{P}D)^a\mathcal{P}$$

$$\implies \qquad P = \sum_{k=0, k\neq b}^{a-1} t^k P^k + t^a P^{a+1}$$

# Summary of Results and Conclusions

Our results establish:

- Grammatical proofs of combinatorial results, which provide structural information not given by automated proof
- Found the equations for these infinite families of Dyck paths:
  1. Avoid up-runs of lengths in $\{ar + b : r \in \mathbb{N}\}$
     - Separate cases when $a \leq b$ and when $a > b$
     - Did the same thing for down-runs
  2. Avoid up and down runs of lengths $\{1, \ldots, r\}$
  3. Avoid up-runs of lengths $\{1, \ldots, m\}$ and down-runs of lengths $\{1, \ldots, n\}$
  4. Avoid up-runs of lengths $\{1, \ldots, r\}$ and down-runs of lengths $\{k + 1, \ldots, r\}$

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.

**Joint work with Doron Zeilberger:**
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 $S$.

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

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

# 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 $P(t)$

# 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

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

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

$$\implies \quad P[0,0](t) = 1 + t \cdot P[0,0](t) + Q[0,0](t) \cdot P[0,0](t)$$

# Example of Process: $Q[0,0](t)$

Now we want to find $Q[0,0](t)$

Let $A$ denote the positive legal steps and $B$ denote the negative legal steps.

- The legal initial steps are $s_k \in A$

  Separating this step leaves a path that starts at height $k$

- The legal final steps are $s_\ell \in B$

  Separating this step leaves a path that ends at height $-\ell$

$$\implies \mathcal{Q}_{0,0} = \bigcup_{k \in A} \bigcup_{\ell \in B} s_k \left[ \mathcal{Q}_{k,-\ell} \right] s_\ell$$

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

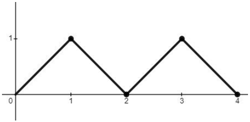
$$\implies Q[0,0](t) = t^2 \sum_{k \in A} \sum_{\ell \in B} P[k-1,-\ell-1](t)$$

# Area Under Generalized Dyck Paths

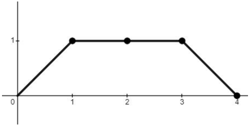
To keep track of area as well as the number of paths, we use the following bi-variate weight enumerator:

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

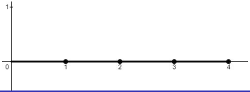
E.g.




*UDUD* has weight $t^4 q^2$




*UFFD* has weight $t^4 q^3$




*FFFF* has weight $t^4$

# E.g. Area Under Motzkin Paths

Recall from earlier the grammar

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

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.

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

# Area Under Generalized Dyck Paths

Using similar adjustments to account for area, we can modify our method of enumerating restricted Dyck paths to keep track of the total area.

e.g. Before we had

$$\mathcal{Q}_{0,0} = \bigcup_{k \in A} \bigcup_{\ell \in B} s_k \mathcal{Q}_{k,-\ell} s_\ell$$

$$\implies \qquad Q[0,0](t) = t^2 \sum_{k \in A} \sum_{\ell \in B} P[k-1, -\ell-1](t)$$

Now, considering area, we have...

$$Q[0,0](t,q) = t^2 \sum_{k \in A} \sum_{\ell \in B} q^{k/2 - \ell/2} P[k-1, -\ell-1](qt, q).$$

# 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 $P_q(t,1)$, which gives the total area under the paths of length $n$.

We can also solve for higher order derivatives:

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

**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 $(q-1)^k$, we can solve for $P^{(k)}(t,1)$

# 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

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

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

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

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

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

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

# 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

From these, we can also get the average area and the variances.

# Take Away

Experimental mathematics is an incredible tool for mathematical research and more people should use it!

# Thank You!

Thank you all for attending my defense and for your support!

Thank you to my advisor, Dr. Z, for all of your help and advice throughout my time at Rutgers

Thank you to Angela Gibney for helping me learn about Gröbner bases and for your support

Thank you to my other Committee Members: Prof. Retakh and Bhargav