# Gessel's Lattice Path Conjecture and Dyck Paths

Arvind Ayyer

Institut de Physique Théorique

CEA Saclay

91191 Gif-sur-Yvette Cedex, France

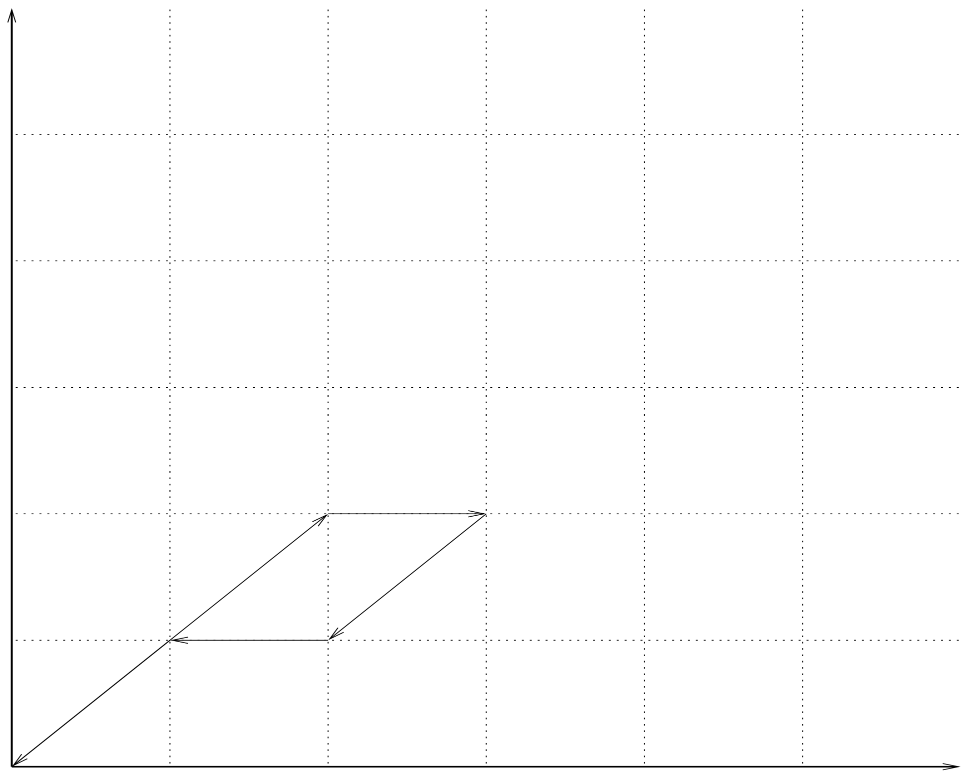
Séminaire Lotharingien de Combinatoire,
Bertinoro, September 29, 2009

**Abstract**

Ira Gessel's conjecture (circa 2001) about the number of certain lattice paths in the quarter plane starting and ending at the origin has generated a lot of interest in the enumerative combinatorics community. Even though it has been proved by Kauers, Koutschan and Zeilberger in 2008 using computer algebra techniques, the structure behind the problem and the reason it is nontrivial are not known. We attempt to understand these facets by formulating an explicit bijection between these paths and certain subsets of Dyck paths, which we then use to count a specialized subset of these quarter plane walks.

## Gessel Walks

Walks with steps $\{(1,1),(1,0),(-1,0),(-1,-1)\}$ in the quarter plane of the integer lattice in two dimensions.



An example of a Gessel walk which begins and ends at the origin in six steps.

## Gessel Conjecture → KKZ Theorem

Ira Gessel conjectured that the number of such walks is given by

$$16^n \frac{(5/6)_n (1/2)_n}{(2)_n (5/3)_n}, \tag{1}$$

where

$$(a)_n = a(a+1)\ldots(a+n-1) \tag{2}$$

is the Pochhammer symbol or rising factorial. This has been proven by Kauers, Koutschan and Zeilberger, `arXiv:0806.4300`.

The main idea is as follows: Let $f(n;i,j)$ be the number of steps from the origin to $(i,j)$ using these steps. They use Maple to compute a homogeneous linear recurrence for $f(n;0,0)$. The smallest recurrence they find is of order 32, polynomial coefficients of degree 172, involves integers with up to 385 decimal digits and takes about 250 pages to be printed! They then verify that the function in (1) satisfies the recurrence and i.c's.

## Gessel Words

We define words analogous to these walks. Although these ideas can be extended to arbitrary dimensions, we restrict ourselves to $d = 2$.

The *Gessel alphabet in two letters* consists of a set of letters $S = \{1, 2\}$ with an order $<$ $(1 < 2)$ along with their complements which we denote $\bar{S} = \{\bar{1}, \bar{2}\}$. Denote by $N_\alpha(w)$ the number of occurences of the letter $\alpha \in S \cup \bar{S}$ in the word $w$. (For example, $N_2(2\bar{2}) = N_{\bar{2}}(2\bar{2}) = 1, N_1(2\bar{2}) = 0$.) Then a *Gessel word* $w$ is a word consisting of letters in $S \cup \bar{S}$ such that every prefix $p$ of the word satisfies

$$N_2(p) \geq N_{\bar{2}}(p), N_1(p) + N_2(p) \geq N_{\bar{1}}(p) + N_{\bar{2}}(p). \tag{3}$$

A *complete Gessel word* is a Gessel word $w$ where $N_i(w) = N_{\bar{i}}(w)$ for all letters $i \in S$.

As an example, both $2\bar{1}1\bar{2}$ and $12\bar{1}\bar{2}1$ are Gessel words but $1\bar{2}21\bar{1}\bar{1}$ is not because the prefix consisting of two letters fails the criterion (3). Among the other two, the first one is a complete Gessel word.

## Triangle of Gessel words

Complete Gessel words are clearly in bijection with Gessel walks. We form the number triangle $T(n,k)$ of the number of walks with $2n$ steps and $k$ occurences of steps $(1,1)$ and $(-1,-1)$. Equivalently, the number of complete Gessel words of length $2n$ with $k$ 2 and $\bar{2}$'s.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 2 & & 7 & & 2 & & \\ & 5 & & 37 & & 38 & & 5 & \\ 14 & & 177 & & 390 & & 187 & & 14 \end{array} \tag{4}$$

The extreme entries are Catalan numbers because they correspond exactly to Dyck paths.

The reflection asymmetry is exactly because of (3).

## OEIS to the Rescue

The next-to-rightmost sequence $1, 7, 38, 187, \ldots$ turns out to be present in the Sloane database! The $n$th term in the sequence is the degree of the polynomial satisfied by $16K^2$, where $K$ is the area of a *cyclic polygon* of $2n+1$ sides. This was originally conjectured by Dave Robbins.

We are thus led to the following conjecture: The number of complete Gessel words of length $2n$ with one 1 and $\bar{1}$, is given by

$$G_1(n) = \frac{(2n+1)}{2}\binom{2n}{n} - 2^{2n-1}. \tag{5}$$

## Connection with Dyck Paths

Suppose that the first non-$\{2,\bar{2}\}$ letter in a Gessel word is 1 at position $i+1$. Then the prefix of $i$ letters is exactly a Dyck path, with $2 \rightarrow \nearrow, \bar{2} \rightarrow \searrow$, and no other conditions. On the other hand if the letter at position $i+1$ is $\bar{1}$, then the prefix is a Dyck path with the condition that the height at the $i$th position is at least 1.

For example, if $i = 2$, then both 22 and $2\bar{2}$ are allowed in the former case, but only 22 in the latter. We generalize this idea to a definition.

Let $P = (P_1, \ldots, P_m)$ be an increasing list of positive integers and $H = (H_1, \ldots, H_m)$ be a list of nonnegative integers of the same length. We define a $(P,H)$-*Dyck path* to be a Dyck path which satisfies the constraint that between positions $P_i$ and $P_{i+1}$ (both inclusive), the ordinate of the path is greater than or equal to $H_i$ for $i = 1, \ldots, m-1$.

## Bijection Time

Consider a complete Gessel word with $n_1$ $1,\bar{1}$'s and $n_2$ $2,\bar{2}$'s. We will construct lists $P$ and $H$ purely from the information about $1,\bar{1}$'s in the word.

Let $S$ be the list of 1 or $\bar{1}$ of length $2n_1$ as they occur in the word and $T$ be the same list with $\bar{1}$ replaced by $-1$. Let $\tilde{P}$ be the list of positions of 1 and $\bar{1}$ and $P$ be the list such that $P_i = \tilde{P}_i - i$. Lastly, define $H$ via the formula $H_i = \max\left\{ -\sum_{k=1}^{i} T_k, 0\right\}$.

Fix the list $S$ of length $2n_1$ in letters 1 and $\bar{1}$ beforehand. Complete Gessel words of length $2(n_1+n_2)$ in two letters with the positions of the letters in $S$ given by the list $\tilde{P}$ are in bijection with $(P,H)$-Dyck paths of length $2n_2$ where the pairs of lists $(P,H)$ are constructed by the algorithm described above.

## Example

Suppose $n = 4$ with $n_1 = 1, n_2 = 3$ and we let $S = (\bar{1}, 1)$. Thus, $T = (-1, 1)$ and $H = (1, 0)$. Fix the positions of the letters 1 and $\bar{1}$ at 4 and 7, say. So, the word looks like

$$\_ \quad \_ \quad \_ \quad \bar{1} \quad \_ \quad \_ \quad 1 \quad \_.$$

Thus $\tilde{P} = (4, 7)$ and $P = (3, 5)$.

We thus have to construct Dyck paths of length 6 which stay above the line $y = 1$ between $x = 3$ and $x = 5$. There are exactly three such possibilities:

- $\nearrow\nearrow\nearrow\searrow\searrow\searrow$ corresponds to $222\bar{1}\bar{2}\bar{2}1\bar{2}$
- $\nearrow\nearrow\searrow\nearrow\searrow\searrow$ corresponds to $22\bar{2}\bar{1}2\bar{2}1\bar{2}$
- $\nearrow\searrow\nearrow\nearrow\searrow\searrow$ corresponds to $2\bar{2}2\bar{1}2\bar{2}1\bar{2}$

## A Classical Result

The number of Dyck paths $a_{i,j}(k)$ that stay above the $x$-axis starting at the position $(0, i)$ and end at position $(k, j)$ is given by

$$a_{i,j}(k) = \binom{k}{(k+i-j)/2} - \binom{k}{(k+i+j)/2+1} \tag{6}$$

if $(k+i+j) \equiv 0 \mod 2$ and 0 otherwise.

The proof uses the reflection principle and is a simple generalization of the proof that the Catalan numbers count the number of Dyck paths starting and ending on the $x$-axis.

## A General Formula

Using the bijection and the classical formula (6), we can count the number of complete Gessel words with fixed lists of the order and position of $1, \bar{1}$'s given by $S$ and $\tilde{P}$ respectively.

Calculate the lists $T$ and $H$ by the algorithm above and let the number of such complete Gessel words be denoted by $G_{n_1}(\tilde{P}, S; 2n)$. Then

$$G_{n_1}(\tilde{P}, S; 2n) = \sum_{k_1=H_1}^{\tilde{P}_1-1} \sum_{k_2=H_2}^{\tilde{P}_2-1} \cdots \sum_{k_i=H_i}^{\tilde{P}_i-1} \cdots \sum_{k_{2n_1}=H_{2n_1}}^{\tilde{P}_{2n_1}-1} a_{0,k_1}(\tilde{P}_1 - 1) \\ a_{k_{2n_1},0}(2n - \tilde{P}_{2n_1}) \prod_{i=2}^{2n_1} a_{k_{i-1}-H_{i-1},k_i-H_{i-1}}(\tilde{P}_i - \tilde{P}_{i-1} - 1). \quad (7)$$

Here $k_i$ is the height of the Dyck path between position $P_{i-1}$ and $P_i$. The reason we take $a_{k_{i-1}-H_{i-1},k_i-H_{i-1}}()$ is that the Dyck path cannot fall below $H_{i-1}$ between $P_{i-1}$ and $P_i$.

## The Grand Formula

Form the set

$$\mathcal{S} = \left\{ (\tilde{P}, S) \,\middle|\, \begin{array}{l} S \text{ is a list of } n_1 \text{ 1's and } n_1 \ \bar{1}\text{'s.} \\ \tilde{P} \text{ is an increasing list of} \\ \quad 2n_1 \text{ positions between 1 and } 2n, \end{array} \right\}, \tag{8}$$

Therefore the number of complete Gessel words with exactly $n_1$ $1, \bar{1}$'s is given by

$$G_{n_1}(n) = \sum_{(\tilde{P},S)\in\mathcal{S}} G_{n_1}(\tilde{P}, S; 2n), \tag{9}$$

and the number of complete Gessel words in $2n$ letters is

$$G(n) = \sum_{n_1=0}^{n} G_{n_1}(n). \tag{10}$$

Unfortunately, this involves terms with $2n$ summands of binomial coefficients! So, even numerical calculation of $G(n)$ takes as much time as the recursive formula.

## Complete Gessel words with one $1, \bar{1}$

Let $d_{i,j}$ be the number of possibilities where a 1 or a $\bar{1}$ is at position $i$ and its counterpart at position $j$. Then we draw the following triangle for a specific $n$,

$$\begin{array}{ccccccccc} & & & & d_{1,2n} & & & & \\ & & & d_{1,2n-1} & & d_{2,2n} & & & \\ & & \cdot^{\cdot^{\cdot}} & & & & \ddots & & \\ d_{1,2} & & \cdots & & \cdots & & \cdots & & d_{2n-1,2n}. \end{array} \tag{11}$$

For $n = 3$, the triangle is

$$\begin{array}{ccccccccc} & & & & 2 & & & & \\ & & & 2 & & 2 & & & \\ & & 2 & & 3 & & 2 & & \\ & 2 & & 3 & & 3 & & 2 & \\ 2 & & 4 & & 3 & & 4 & & 2, \end{array} \tag{12}$$

and for $n = 4$, the triangle is

$$
\begin{array}{ccccccccccccc}
 & & & & & & 5 & & & & & & \\
 & & & & & 5 & & 5 & & & & & \\
 & & & & 5 & & 7 & & 5 & & & & \\
 & & & 5 & & 7 & & 7 & & 5 & & & \\
 & & 5 & & 8 & & 7 & & \underline{8} & & 5 & & \\
 & 5 & & 8 & & 8 & & 8 & & 8 & & 5 & \\
5 & & 10 & & 8 & & 10 & & 8 & & 10 & & 5,
\end{array} \qquad (13)
$$

where the underlined case is the one done before. Notice that

1. The triangles are reflection symmetric.

2. The extremal diagonal entries are Catalan numbers.

3. Alternate entries on the bottom row are double Catalan numbers.

4. Each element in the interior repeats $4k$ times.

**Proof of** (5)

We want to calculate the sum of $G_1([i,j],[1,\bar{1}];2n)$ and $G_1([i,j],[\bar{1},1];2n)$ over all $i,j$ such that $1 \leq i < j \leq 2n$. The former summand is just the Catalan number $C_{n-1}$. Therefore, we have the contribution

$$(2n-1)\binom{2n-2}{n-1}. \tag{14}$$

The latter summand is given by

$$G_1([i,j],[\bar{1},1];2n) = \sum_{k_1=1}^{i-1} \sum_{k_2=0}^{j-1} a_{0,k_1}(i-1)\, a_{k_1-1,k_2-1}(2j-2i-1)\, a_{k_2,0}(2n-j), \tag{15}$$

We now use the classical result to write

$$\begin{aligned} G_1([2i,2j],[\bar{1},1];2n) &= G_1([2i,2j+1],[\bar{1},1];2n) \\ = G_1([2i+1,2j],[\bar{1},1];2n) &= G_1([2i+1,2j+1],[\bar{1},1];2n). \end{aligned} \tag{16}$$

Then the total number of such Gessel words with a $\bar{1}$ preceding a 1 is given by

$$\begin{aligned}\sum_{i=1}^{2n-1}\sum_{j=i+1}^{2n} G_1([i,j],[\bar{1},1];2n) =& 4\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} G_1([2i,2j],[\bar{1},1];2n) \\ &+ \sum_{i=1}^{n-1} G_1([2i,2i+1],[\bar{1},1];2n) \\ &= 4(S_2 - S_3) + S_1.\end{aligned} \tag{17}$$

where we have split the sum in three parts, with

$$S_1 = \sum_{i=1}^{n-1} G_1([2i,2i+1],[\bar{1},1];2n). \tag{18}$$

To define $S_2, S_3$, we need the Catalan triangle numbers $C_n^m$, given by $\frac{(m-n+1)}{(m+1)}\binom{m+n}{n}$ for $0 \leq n \leq m$. This is useful because the factors $a_{0,k_1}(i-1)$ and $a_{k_2,0}(2n-j)$ are these numbers.

We split $a_{k_1-1,k_2-1}(2j-2i-1)$ and get

$$S_2 = \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{r=0}^{i-1}\sum_{s=0}^{n-j-1} C_r^{2i-r-1} C_s^{2n-2j-s} \binom{2j-2i-1}{2j+s-n-r-1},$$
$$S_3 = \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{r=0}^{i-1}\sum_{s=0}^{n-j-1} C_r^{2i-r-1} C_s^{2n-2j-s} \binom{2j-2i-1}{n-s-r-1}. \tag{19}$$

$S_1$ is the simplest,

$$S_1 = \sum_{i=1}^{n-1}\sum_{r=1}^{i}\sum_{s=1}^{i} C_{i-r}^{i-1+r} C_{n-i-s}^{n-i+s-1} \left[\binom{0}{r-s} - \binom{0}{r+s-1}\right]. \tag{20}$$

The first term forces $r = s$ and the second term is identically zero because $r + s \geq 2$. This means we are left with

$$S_1 = \sum_{i=1}^{n-1}\sum_{r=1}^{i} C_{i-r}^{i-1+r} C_{n-i-r}^{n-i+r-1} = (n-1)C_{n-1}. \tag{21}$$

The sums $S_2$ and $S_3$ should be computable using known computer algorithms, we needed to massage these in order to compute them.

Using variants of the Chu-Vandermonde identities and certain Catalan triangle identities, we were able to show

$$S_2 = \frac{n+2}{4}\binom{2n}{n} - 3 \cdot 2^{2n-3}, \tag{22}$$

and

$$S_3 = \binom{2n-2}{n-4} - 2^{2n-2} + \frac{(2n)!(3n^2+n+2)}{2n!(n+2)!}. \tag{23}$$

Combining these proves the result. ■

We computed these using somewhat cumbersome manipulations and it would be nice if this could be automated.