A REMARKABLE FORMULA FOR COUNTERING NONINTERSECTING LATTICE PATHS IN A LADDER WITH RESPECT TO TURNS

C. KRATTENTHALER AND M. PROHASKA

Abstract. We prove a formula, conjectured by Conca and Herzog, for the number of all families of nonintersecting lattice paths with certain starting and end points in a region that is bounded by an upper ladder. Thus we are able to compute explicitly the Hilbert series for certain one-sided ladder determinantal rings.

1. Introduction

Recent work of Abhyankar and Kulkarni [1], [2], [18], [19], Bruns, Conca, Herzog, and Trung [3], [5], [6], [10] showed that the computation of the Hilbert series for ladder determinantal rings (see section 3 for the definition) boils down to counting families of \( n \) nonintersecting lattice paths with a given total number of turns in a certain ladder-shaped region. For one-sided ladders, Conca and Herzog [6, last paragraph] conjectured a remarkable formula that reduces the problem to counting the number of all single lattice paths with a given number of turns in the ladder region. We state this conjecture in section 2 (our Theorem 1), and prove it in section 4. The crucial idea in the proof of the conjecture is to encode lattice paths by two-rowed arrays, and then play with these two-rowed arrays. This idea was introduced in [11], [16] and developed to full power in [13], [14], [15]. The implication of Theorem 1 to the computation of Hilbert series of ladder determinantal rings is developed in section 3, see Theorem 2. An extension of Theorem 1 can be found in section 5. In section 6 we address the problem of counting single lattice paths with a given number of turns in a ladder region. Finally, in section 7, we comment on some more general problems and related results.

2. Statement of the formula

To be able to state Conca and Herzog’s conjecture, we have to introduce a few terms. When we say lattice path we always mean a lattice path in the plane consisting of unit horizontal and vertical steps in the positive direction, see Figure 1.
We shall frequently abbreviate the fact that a lattice path $P$ goes from $A$ to $E$ by $P : A \rightarrow E$.

Also, given lattice points $A$ and $E$, we denote the set of all lattice paths from $A$ to $E$ by $\mathcal{P}(A \rightarrow E)$. If $A = (A_1, \ldots, A_n)$ and $E = (E_1, \ldots, E_n)$ are $n$-tuples of lattice points, we denote the set of all families $(P_1, \ldots, P_n)$ of lattice paths, where $P_i$ runs from $A_i$ to $E_i$, $i = 1, \ldots, n$, by $\mathcal{P}(A \rightarrow E)$. A family of lattice paths is said to be nonintersecting if no two lattice paths of this family have a point in common, otherwise it is called intersecting. We denote the set of all families $(P_1, P_2, \ldots, P_n)$ of nonintersecting lattice paths, $P_i : A_i \rightarrow E_i$, by $\mathcal{P}^+(A \rightarrow E)$.

Our lattice paths will be restricted to a certain ladder shaped region, $L$ say, which is described next. Let $a, b$ be positive integers, and let $f$ be a weakly increasing function from $[0, a]$ to $[1, b + 1]$. Here, by $[c, d]$ we mean the set of all integers $\geq c$ and $\leq d$. Then, by definition,

$$L := \{(x, y) : x \in [0, a] \text{ and } 0 \leq y < f(x)\}. \tag{2.1}$$

An example with $a = 13$, $b = 15$, and $f$ given by $f(0) = 7$, $f(1) = 7$, $f(2) = 7$, $f(3) = 7$, $f(4) = 10$, $f(5) = 11$, $f(6) = 12$, $f(7) = 13$, $f(8) = 16$, $f(9) = 16$, $f(10) = 16$, $f(11) = 16$, $f(12) = 16$, $f(13) = 16$, is displayed in Figure 2.

We extend our lattice path notation in the following way. By $\mathcal{P}_L(A \rightarrow E)$ we mean the set of all lattice paths $P$ from $A$ to $E$ all of whose lattice points lie in the ladder region $L$. Similarly, by $\mathcal{P}_L(A \rightarrow E)$ (respectively $\mathcal{P}^+_L(A \rightarrow E)$) we mean the set of all families $(P_1, P_2, \ldots, P_n)$ of lattice paths (respectively nonintersecting lattice paths), where $P_i$ runs from $A_i$ to $E_i$ and where all the lattice points of $P_i$ lie in the ladder region $L$.

A point in a lattice path $P$ which is the end point of a vertical step and at the same time the starting point of a horizontal step will be called a north-east turn (NE-turn for short) of the lattice path $P$. The NE-turns of the lattice path in Figure 1 are $(1, 1)$, $(2, 3)$, and $(5, 4)$. We write $\text{NE}(P)$ for the number of NE-turns of $P$. Also, given a family $P = (P_1, P_2, \ldots, P_n)$ of paths $P_i$, we write $\text{NE}(P)$ for the number $\sum_{i=1}^n \text{NE}(P_i)$ of all NE-turns in the family. Finally, given any weight function $w$ defined on a set $\mathcal{M}$, by the generating function $\text{GF}(\mathcal{M}; w)$ we mean $\sum_{x \in \mathcal{M}} w(x)$.
Now we are in the position to state Conca and Herzog’s conjecture which becomes a theorem by our proof in section 4.

**Theorem 1.** Let $n, a, b$ be positive integers. Furthermore, let $L$ be a ladder-shaped region determined by the weakly increasing function $f : [0, a] \to [1, b+1]$ by means of (2.1), with $f$ satisfying $f(x) \geq n$ for all $x$, and $f(x) = b+1$ for $a-n < x \leq a$. Also, let $A_i = (0, n-i)$ and $E_i = (a-n+i, b)$, $i = 1, 2, \ldots, n$, be lattice points. Then for the generating function $\sum z^{\sum NE(P)}$, where the sum is over all families $P = (P_1, P_2, \ldots, P_n)$ of nonintersecting lattice paths lying in the region $L$, $P_i : A_i \to E_i$, $i = 1, 2, \ldots, n$, there holds

$$GF(P^+_L(A \to E); z^{\sum NE(P)}) = z^{-\binom{n}{2}} \det_{1 \leq s, t \leq n} \left( GF(P_L(A_i \to E_s); z^{NE(P)}) \right).$$

**Remarks.** (1) Note that the starting points $A_i$ are lined up vertically along the $y$-axis, top to bottom, while the end points $E_i$ are lined up horizontally along the horizontal line $y = b$, from left to right. The assumptions about the function $f$ are needed to make the enumeration problem non-degenerate, i.e., to guarantee that there exists at least one family of nonintersecting lattice paths in the corresponding region $L$ that connects the starting points $A$ with the end points $E$.

(2) Conca and Herzog state their conjecture in a different way, but it is easily seen that the above statement is equivalent to theirs. (It is actually simply the translation into generating functions.)

(3) Conca and Herzog proved Theorem 1 for $n = 2$ (private communication), but for generic $n$ only if $f(x) \equiv b + 1$ (i.e., if there is no ladder restriction; see [6, next-to-last paragraph]).

(4) The formula (2.2) clearly reduces the problem of enumerating families of nonintersecting lattice paths in the ladder region $L$ with respect to NE-turns to the problem of enumerating single lattice paths in $L$ with respect to NE-turns. We are going to address this problem in section 6.
3. Ladder determinantal rings

As mentioned in the Introduction, the above enumeration problem occurs in the
determination of the Hilbert series of certain ladder determinantal rings. Ladder
determinantal rings arise for instance in Abhyankar’s study on the singularities
of Schubert varieties of flag manifolds. Let us recall the definition of a ladder
determinantal ring. Let $X = (X_{i,j})_{0 \leq i \leq b, 0 \leq j \leq a}$ be a $(b+1) \times (a+1)$ matrix of
indeterminates. Let $Y = (Y_{i,j})_{0 \leq i \leq b, 0 \leq j \leq a}$ be another $(b+1) \times (a+1)$ matrix with
the property that $Y_{i,j} = X_{i,j}$ or 0, and if $Y_{i,j} = X_{i,j}$ and
$Y_{i',j'} = X_{i',j'}$, where $i \leq i'$ and $j \leq j'$, then $Y_{s,t} = X_{s,t}$ for all $s, t$ with $i \leq s \leq i'$ and $j \leq t \leq j'$.
An example for such a matrix $Y$, with $b = 15$ and $a = 13$ is displayed in Figure 3.

\[
\begin{pmatrix}
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{0,8} & X_{0,9} & X_{0,10} & X_{0,11} & X_{0,12} & X_{0,13} \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{0,8} & X_{0,9} & X_{0,10} & X_{0,11} & X_{0,12} & X_{0,13} \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{0,8} & X_{0,9} & X_{0,10} & X_{0,11} & X_{0,12} & X_{0,13} \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_{0,7} & X_{0,8} & X_{0,9} & X_{0,10} & X_{0,11} & X_{0,12} & X_{0,13} \\
  0 & 0 & 0 & 0 & 0 & 0 & X_{4,6} & X_{4,7} & X_{4,8} & X_{4,9} & X_{4,10} & X_{4,11} & X_{4,12} & X_{4,13} \\
  0 & 0 & 0 & 0 & 0 & X_{6,5} & X_{6,6} & X_{6,7} & X_{6,8} & X_{6,9} & X_{6,10} & X_{6,11} & X_{6,12} & X_{6,13} \\
  0 & 0 & 0 & 0 & X_{6,4} & X_{6,5} & X_{6,6} & X_{6,7} & X_{6,8} & X_{6,9} & X_{6,10} & X_{6,11} & X_{6,12} & X_{6,13} \\
  0 & 0 & 0 & X_{7,4} & X_{7,5} & X_{7,6} & X_{7,7} & X_{7,8} & X_{7,9} & X_{7,10} & X_{7,11} & X_{7,12} & X_{7,13} \\
  0 & 0 & X_{8,4} & X_{8,5} & X_{8,6} & X_{8,7} & X_{8,8} & X_{8,9} & X_{8,10} & X_{8,11} & X_{8,12} & X_{8,13} \\
  X_{9,0} & X_{9,1} & X_{9,2} & X_{9,3} & X_{9,4} & X_{9,5} & X_{9,6} & X_{9,7} & X_{9,8} & X_{9,9} & X_{9,10} & X_{9,11} & X_{9,12} & X_{9,13} \\
  X_{10,0} & X_{10,1} & X_{10,2} & X_{10,3} & X_{10,4} & X_{10,5} & X_{10,6} & X_{10,7} & X_{10,8} & X_{10,9} & X_{10,10} & X_{10,11} & X_{10,12} & X_{10,13} \\
  X_{11,0} & X_{11,1} & X_{11,2} & X_{11,3} & X_{11,4} & X_{11,5} & X_{11,6} & X_{11,7} & X_{11,8} & X_{11,9} & X_{11,10} & X_{11,11} & X_{11,12} & X_{11,13} \\
  X_{12,0} & X_{12,1} & X_{12,2} & X_{12,3} & X_{12,4} & X_{12,5} & X_{12,6} & X_{12,7} & X_{12,8} & X_{12,9} & X_{12,10} & X_{12,11} & X_{12,12} & X_{12,13} \\
  X_{13,0} & X_{13,1} & X_{13,2} & X_{13,3} & X_{13,4} & X_{13,5} & X_{13,6} & X_{13,7} & X_{13,8} & X_{13,9} & X_{13,10} & X_{13,11} & X_{13,12} & X_{13,13} \\
  X_{14,0} & X_{14,1} & X_{14,2} & X_{14,3} & X_{14,4} & X_{14,5} & X_{14,6} & X_{14,7} & X_{14,8} & X_{14,9} & X_{14,10} & X_{14,11} & X_{14,12} & X_{14,13} \\
  X_{15,0} & X_{15,1} & X_{15,2} & X_{15,3} & X_{15,4} & X_{15,5} & X_{15,6} & X_{15,7} & X_{15,8} & X_{15,9} & X_{15,10} & X_{15,11} & X_{15,12} & X_{15,13}
\end{pmatrix}
\]

Figure 3

(Note that there could be 0’s in the bottom-right corner of the matrix also.)

Such a “submatrix” $Y$ of $X$ is called a ladder. This terminology is motivated by
the identification of such a matrix $Y$ with the set of all points $(j, b-i)$ in the plane
for which $Y_{i,j} = X_{i,j}$. Thus in particular, the set of all such points for the special
matrix in Figure 3 is exactly the ladder region of Figure 2. (It should be apparent
from comparison of Figures 3 and 2 that the reason for taking $(j, b-i)$ instead of
$(i, j)$ is to take care of the difference in “orientation” of row and column indexing.
of a matrix versus coordinates in the plane.) In general, this set of points looks like a (two-sided) ladder-shaped region.

Let $K[Y]$ denote the ring of all polynomials over some field $K$ in the $Y_{i,j}$’s, $0 \leq i \leq b$, $0 \leq j \leq a$, and let $I_{n+1}(Y)$ be the ideal in $K[Y]$ that is generated by all $(n+1) \times (n+1)$ minors of $Y$ that contain only nonzero entries. The ideal $I_{n+1}(Y)$ is called a ladder determinantal ideal. The associated ladder determinantal ring is $R_{n+1}(Y) := K[Y]/I_{n+1}(Y)$. (We remark that the definition of ladder is more general in [1], [2], [5], [10]. However, there is in effect no loss of generality since the ladders of [1], [2], [5], [10] can always be reduced to our definition by discarding superfluous 0’s.) Combining results of Abhyankar [1], [2] or Herzog and Trung [10] with our Theorem 1, we are able to give an explicit formula for the Hilbert series of the ladder determinantal ring $R_{n+1}(Y)$ in the case of one-sided ladders, which can be effectively used computationally. By a one-sided ladder, we mean a ladder $Y$ where either $Y_{0,0} = X_{0,0}$ or $Y_{b,a} = X_{b,a}$. In the first case we call $Y$ a lower ladder, in the second an upper ladder. Thus, the matrix in Figure 3 is an upper ladder. Without loss of generality we state the result for the case of upper ladders only.

**Theorem 2.** Let $Y = (Y_{i,j})_{0 \leq i \leq b, 0 \leq j \leq a}$ be an upper ladder, let $L$ be the associated ladder region, i.e., $Y_{i,j} = X_{i,j}$ if and only if $(j, b-i) \in L$, and let $f: [0, a] \rightarrow [1, b+1]$ be the function that describes this ladder region by means of (2.1), i.e., $Y_{i,j} = X_{i,j}$ if and only if $b - i < f(j)$. Furthermore, as in Theorem 1 we let $A_i = (0, n-i)$ and $E_i = (a-n + i, b)$, $i = 1, 2, \ldots, n$. Then the Hilbert series of the ladder determinantal ring $R_{n+1}(Y) = K[Y]/I_{n+1}(Y)$ equals

$$
\sum_{\ell=0}^{\infty} \dim_K R_{n+1}(Y)_\ell z^\ell = \frac{z^{-\binom{a}{2}} \det_{1 \leq s,t \leq n} (\text{GF}(P_L(A_s \rightarrow E_s); z^{\text{NE}(s)}))}{(1 - z)^{(a+b+1)n-2\binom{a}{2}}},
$$

where $R_{n+1}(Y)_\ell$ denotes the homogeneous component of degree $\ell$ in $R_{n+1}(Y)$.

**Proof.** We describe two ways to establish the theorem. One is by use of work of Abhyankar [1] (see also [2]), the other is by use of work of Herzog and Trung [10].

**First proof.** We follow basically the considerations of Kulkarni [18], [19] (see also [8]). In Theorem (20.10) of [1] (see also Theorem (6.7) in [7]; for our purposes it is sufficient to consider the special case $a = 1, 2, \ldots, n \mid 1, 2, \ldots, n$) in these theorems, cf. Example (1) on p. 10 of [10]) Abhyankar proves that the monomials $\prod_{(i,j) \in S} X_{b-j,i}$ form a basis of $R_{n+1}(Y)$, where $S$ ranges over the following multisets:

1. $Y_{b-j,i} = X_{b-j,i}$ for all $(i,j) \in S$ (i.e., $(i,j)$ is a point in the ladder region $L$ associated to $Y$).
2. Define a partial order on the points in the plane by

$$(x_1, y_1) \leq (x_2, y_2) \quad \text{if and only if} \quad x_1 \leq x_2 \text{ and } y_1 \geq y_2,$$

i.e., if and only if $(x_2, y_2)$ is located weakly to the right and below of $(x_1, y_1)$. Then the length of a chain $(i_1, j_1) < (i_2, j_2) < \cdots < (i_k, j_k)$ of elements of $S$, in this partial order, is at most $n$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
An example of such a set $S$, with $a = 13$, $b = 15$, $n = 4$, the ladder region being the one from Figure 2, is

$$
\{(0,1),(0,5),(2,4),(2,6),(3,0),(3,3),(4,4),(4,6),(4,6),(4,6),(5,8),(5,8),(5,8),(6,1),(6,6),
(7,2),(7,2),(8,6),(8,6),(8,6),(8,10),(8,13),(8,13),(8,15),(9,2),(9,5),(9,8),(9,11),(9,11),
(10,2),(10,7),(10,9),(10,13),(10,15),(11,4),(12,4),(12,4),(12,7),(13,10),(13,10),(13,11),(13,11),(13,15)\}.
$$

A geometric realization of this set is contained in Figure 4, the elements of the multiset $S$ being indicated by bold dots, numbers in parentheses attached to dots indicating multiplicities. (For the moment the reader should ignore the lattice paths in Figure 4.)

By Abhyankar’s theorem, the dimension $\dim_K R_{n+1}(Y)_\ell$ of the homogeneous component of degree $\ell$ in $R_{n+1}(Y)$ equals the number of all such multisets $S$ of size $\ell$.

Let $S$ be such a multiset. By Viennot’s “light and shadow procedure” with the sun in the top-left corner (see [10], [22], [23]) we can associate a family $(P_1, P_2, \ldots, P_n)$ of $n$ nonintersecting lattice paths to $S$, $P_i$ running from $A_i = (0, n-i)$ to $E_i = (a - n + i, b)$ and staying in the ladder region $L$, $i = 1, 2, \ldots, n$. The result of “light and shadow” when applied to our example in (3.2) is displayed in Figure 4.

Each path $P_i$ contains a few (possibly multiple) points of $S$. However, in each NE-turn of $P_i$ there has to be at least one element of $S$, $i = 1, 2, \ldots, n$. Therefore, given a family $(P_1, P_2, \ldots, P_n)$ of $n$ nonintersecting lattice paths, $P_i$ running from $A_i$ to $E_i$ and staying in the ladder region $L$, $i = 1, 2, \ldots, n$, with a total number of exactly $m$ NE-turns, there are exactly $\binom{d+m-1}{\ell-m}$ multisets $S$ of cardinality $\ell$ that
reduce to \((P_1, P_2, \ldots, P_n)\) under light and shadow, where

\[(3.3) \quad d = (a + b + 1)n - 2\binom{n}{2}\]

is the total number of lattice points on the lattice paths \(P_1, P_2, \ldots, P_n\).

Hence, if we let \(h_m\) denote the number of all families \((P_1, P_2, \ldots, P_n)\) of \(n\) nonintersecting lattice paths, \(P_i\) running from \(A_i\) to \(E_i\) and staying in the ladder region \(L, i = 1, 2, \ldots, n\), with a total number of exactly \(m\) NE-turns, we obtain for the Hilbert series,

\[
\sum_{\ell=0}^{\infty} \dim_K R_{n+1}(Y)^\ell z^\ell = \sum_{\ell=0}^{\infty} \left( \sum_{m=0}^{\ell} \binom{d + \ell - m - 1}{\ell - m} h_m \right) z^\ell \\
= \sum_{m=0}^{\infty} h_m \sum_{\ell=m}^{\infty} \binom{d + \ell - m - 1}{\ell - m} z^\ell \\
= \sum_{m=0}^{\infty} h_m z^m \sum_{\ell=0}^{\infty} \binom{d + \ell - 1}{\ell} z^\ell \\
= \sum_{m=0}^{\infty} h_m z^m \frac{1}{(1 - z)^d}.
\]

Now, the generating function \(\sum_{m=0}^{\infty} h_m z^m\) is exactly the left-hand side of (2.2). Finally, if we substitute the right-hand side of (2.2) for \(\sum h_m z^m\) in the last line of the above computation we obtain exactly (3.1).

Second proof. We use results of Herzog and Trung [10]. In section 4 of [10], ladder determinantal rings are introduced and investigated. The situation that we consider here is the special case \(M = (1, 2, \ldots, n \mid 1, 2, \ldots, n)\) of the considerations in [10, section 4] (cf. Example (1) on p. 10 of [10]). It is shown in [10, section 4] that for a suitable term order (order of monomials) the ideal \(I_{n+1}(Y)^\ast\) of leading monomials of \(I_{n+1}(Y)\) is generated by square-free monomials. Thus \(K[Y]/I_{n+1}(Y)^\ast\) may be viewed as a Stanley-Reisner ring of a certain simplicial complex \(\Delta\). The faces of this simplicial complex \(\Delta\) are described in Lemma 4.4 of [10]. Namely, translated into a less formal language, the faces are sets (not multisets!) \(S\) of points, such that (1) and (2) at the beginning of the first proof hold. (Note that the pictures of [10] have to be reflected in a horizontal line to be conformable to ours.)

As usual, let \(f_i\) be the number of \(i\)-dimensional faces of \(\Delta\), i.e., the number of such sets \(S\) of cardinality \(i + 1\). By Corollary 4.3 of [10], there holds

\[
\dim_K R_{n+1}(Y)^\ell = \sum_{i \geq 0} \binom{\ell - 1}{i} f_i.
\]

As in the first proof, light and shadow can be applied to any such set \(S\). Here, we infer that, given a a family \((P_1, P_2, \ldots, P_n)\) of \(n\) nonintersecting lattice paths, \(P_i\) running from \(A_i\) to \(E_i\) and staying in the ladder region \(L, i = 1, 2, \ldots, n\), with a total number of exactly \(m\) NE-turns, there are exactly \(\binom{d - m}{i + 1 - m}\) sets \(S\) of cardinality \(i + 1\) that reduce to \((P_1, P_2, \ldots, P_n)\) under light and shadow, where, as before, \(d\) is the total number of lattice points one the lattice paths \(P_1, P_2, \ldots, P_n\), and thus given by (3.3).

Hence, if again \(h_m\) denotes the number of all families \((P_1, P_2, \ldots, P_n)\) of \(n\) nonintersecting lattice paths, \(P_i\) running from \(A_i\) to \(E_i\) and staying in the ladder region
Although, there are no ladders in [3], [6], everything goes through word for word. Arguments are described in detail in the proof of Theorem 2.4 in [3] (see also [6]).

For the proof of Theorem 1 we rely on the Lattice paths and two-rowed arrays. For the proof of Theorem 1 we rely on the encoding of lattice paths in terms of two-rowed arrays, which was introduced and used in [11], [13], [14], [15], [16]. Obviously, given the starting and the end point of a path, the NE-turns uniquely determine the path. Suppose that \( P \) is a path from \( A = (\alpha_1, \alpha_2) \) to \( E = (\eta_1, \eta_2) \) and let the NE-turns of \( P \) be \((a_1, b_1), (a_2, b_2), \ldots, \)

\[
L, \ i = 1, 2, \ldots, n, \text{ with a total number of exactly } m \text{ NE-turns, we obtain for the Hilbert series, }
\[
\sum_{\ell=0}^\infty \dim_K R_{n+1}(Y)_\ell z^\ell = \sum_{\ell=0}^\infty \left( \sum_{i \geq 0} \binom{\ell-1}{i} f_i \right) z^\ell = \sum_{\ell=0}^\infty \sum_{i \geq 0} \binom{\ell-1}{i} \left( \sum_{m=0}^{i+1} \binom{d-m}{i+1-m} h_m \right) z^\ell = \sum_{m=0}^\infty h_m \sum_{\ell=0}^\infty \binom{\ell-1}{i} \left( \frac{d-m}{d-i-1} \right),
\]

and if we sum the inner sum by means of the Vandermonde summation (see e.g. [9, sec. 5.1, (5.27)]),

\[
\sum_{\ell=0}^\infty \dim_K R_{n+1}(Y)_\ell z^\ell = \sum_{m=0}^\infty h_m \sum_{\ell=0}^\infty z^\ell \left( \frac{d + \ell - m - 1}{d-1} \right) = \frac{\sum_{m=0}^\infty h_m z^m}{(1-z)^d}.
\]

The rest of the arguments is as in the first proof.

There is also a different argument that uses the shellability of the simplicial complex \( \Delta \) and the McMullen–Walkup formula (cf. [4, ch. 5, Def. 5.1.11, Cor. 5.1.14] for information on these topics). It avoids computations such as the above ones. Instead, one concludes by properties of facets of \( \Delta \) alone, that for the Hilbert series of \( R_{n+1}(Y) \), where \( h_m \) has the same definition as before. These arguments are described in detail in the proof of Theorem 2.4 in [3] (see also [6]). Although, there are no ladders in [3], [6], everything goes through word for word.

\[\Box\]

**Example.** Let \( Y = (Y_{i,j}) \) be the matrix of Figure 3. Our formula (3.1) gives for the Hilbert series of \( R_2(Y) = K[Y]/I_2(Y) \), using (6.10) for determining the generating function \( \sum z^{\text{NE}(P)} \) for all paths \( P \) in the corresponding ladder region \( L \) of Figure 2,

\[
1+66z+2211z^2+50116z^3+864501z^4+11928939z^5+134672548z^6+1260746851z^7+9885072697z^8
\]

\[
+65397111347z^9+36707499427z^{10}+1755530258028z^{11}+7177561951667z^{12}+25154724107849z^{13}
\]

\[
+7573145361773z^{14}+196211938959224z^{15}+438134985445560z^{16}+84417618468558z^{17}
\]

\[
+1404677025713433z^{18}+201961901866910z^{19}+2509471858669986z^{20}+2694052295102592z^{21}
\]

\[
+2497084208937285z^{22}+1995888618034374z^{23}+1373221471734161z^{24}+811310035353064z^{25}
\]

\[
+41027358338378z^{26}+176833645375538z^{27}+64602317617558z^{28}+1985645822365z^{29}
\]

\[
+508285928078z^{30}+106809067468z^{31}+180378873471z^{32}+23682607189z^{33}+2283692607z^{34}
\]

\[
+144671625z^{35}+4539150z^{36}/(1-z)^{104}.
\]

4. Proof of Theorem 1

**Lattice paths and two-rowed arrays.** For the proof of Theorem 1 we rely on the encoding of lattice paths in terms of two-rowed arrays, which was introduced and used in [11], [13], [14], [15], [16]. Obviously, given the starting and the end point of a path, the NE-turns uniquely determine the path. Suppose that \( P \) is a path from \( A = (\alpha_1, \alpha_2) \) to \( E = (\eta_1, \eta_2) \) and let the NE-turns of \( P \) be \((a_1, b_1), (a_2, b_2), \ldots, \)
(\(a_k, b_k\)), where we assume that the \((a_i, b_i)\) are ordered from left to right, which is equivalent with \(\alpha_1 \leq a_1 < a_2 < \cdots < a_k < \eta_1\), and \(\alpha_2 < b_1 < b_2 < \cdots < b_k \leq \eta_2\). Then \(P\) can be represented by the two-rowed array
\[
\begin{array}{c}
a_{1} & a_{2} & \cdots & a_{k} \\
b_{1} & b_{2} & \cdots & b_{k}
\end{array}
\tag{4.1}
\]
or, if we wish to make the bounds which are caused by the starting and the end point transparent,
\[
\begin{array}{c}
\alpha_{1} \leq a_{1} \ a_{2} \ \cdots \ a_{k} < \eta_{1} \\
\alpha_{2} < b_{1} \ b_{2} \ \cdots \ b_{k} \leq \eta_{2} 
\end{array}
\tag{4.2}
\]
For a given starting point and a given end point, by definition the empty array is the representation for the only path that has no NE-turn. For the path in Figure 1 we obtain the array representation
\[
\begin{array}{c}
1 & 2 & 5 \\
1 & 3 & 4
\end{array}
\tag{4.4a}
\]
or with bounds included,
\[
\begin{array}{c}
1 \leq 1 & 2 & 5 < 6 \\
-1 < 1 & 3 & 4 \leq 6
\end{array}
\tag{4.4b}
\]
It is obvious that under this correspondence there holds the following relation for any path \(P\):
\[
\text{NE}(P) = \frac{1}{2} \text{(number of entries in the two-rowed array corresponding to } P) \tag{4.3}
\]
Later, also two-rowed arrays of different “shape” will be considered. But these arrays will also have the property that the rows are strictly increasing. So, by convention, whenever we speak of two-rowed arrays we mean two-rowed arrays with strictly increasing rows.

**Nonintersecting two-rowed arrays.** Since we consider nonintersecting lattice paths in Theorem 1, we have to figure out what to be nonintersecting means in terms of the corresponding two-rowed arrays. We do this for more general starting and end points then we actually need.

Let \(P_1, P_2\) be two paths, \(P_1 : A \to E, \ P_2 : B \to F\), where \(A = (\alpha_1, \alpha_2), \ B = (\beta_1, \beta_2), \ E = (\eta_1, \eta_2), \ F = (\phi_1, \phi_2)\) with
\[
\begin{align*}
\alpha_1 & \leq \beta_1, \ \alpha_2 \geq \beta_2, \ \eta_1 < \phi_1, \ \eta_2 \geq \phi_2.
\end{align*}
\]
Roughly speaking, these inequalities mean that \(A\) is located in the north-west of \(B\) (strictly in direction north and weakly in direction west), and \(E\) is located in the north-west of \(F\) (weakly in direction north and strictly in direction west), which will always be true for the cases we consider. Let the array representations of \(P_1\) and \(P_2\) be
\[
\begin{align*}
P_1 : \quad & \alpha_1 \leq a_1 \ \cdots \ a_k < \eta_1 \\
& \alpha_2 < b_1 \ \cdots \ b_k \leq \eta_2
\end{align*}
\tag{4.4a}
\]
and
\[
\begin{align*}
P_2 : \quad & \beta_1 \leq c_1 \ \cdots \ c_l < \phi_1 \\
& \beta_2 < d_1 \ \cdots \ d_l \leq \phi_2
\end{align*}
\tag{4.4b}
\]
respectively.
Suppose that $P_1$ and $P_2$ intersect, i.e., have a point in common. Let $S$ be a meeting point of $P_1$ and $P_2$. By definition set $a_{k+1} := \eta_1$ and $b_0 := \alpha_2$. (Note that the thereby augmented sequences $a$ and $b$ remain strictly increasing.) Considering the east-north turn $(a_I, b_{I-1})$ in $P_1$ immediately preceding $S$ (and being allowed to be equal to $S$) and the north-east turn $(c_J, d_J)$ in $P_2$ immediately preceding $S$ (and being allowed to be equal to $S$), we get the inequalities (cf. Figure 5)

\begin{align}
  c_J &\leq a_I, \\
  b_{I-1} &\leq d_J,
\end{align}

where

\begin{align}
  1 \leq I \leq k + 1, \quad 1 \leq J \leq l.
\end{align}

Of course, $k, l, a_I, b_{I-1}, c_J, d_J$, etc., refer to the array representations (4.4a,b) of $P_1$ and $P_2$. It now becomes apparent that the above assignments for $a_{k+1}$ and $b_0$ are needed for the inequalities (4.4c,d) to make sense for $I = 1$ or $I = k + 1$. Note that $S = (a_I, d_J)$. Vice versa, if (4.4c,d,e) is satisfied then there must be a meeting point between $P_1$ and $P_2$ (because of the particular location of the starting and end points $A, B, E, F$).

Summarizing, the existence of $I, J$ satisfying (4.4c,d,e) characterize the array representations of intersecting pairs of paths. Equivalently, $P_1$ and $P_2$ are nonintersecting if and only if there does not exist $I, J$ such that (4.4c,d,e) is satisfied for the corresponding two-rowed arrays.

The two-rowed array interpretation of the left-hand side of (2.2). Let $P = (P_1, P_2, \ldots, P_n)$ be an element of $\mathcal{P}^+_L(A \to B)$, i.e., $P = (P_1, P_2, \ldots, P_n)$ is a family of nonintersecting lattice paths, $P_i$ running from $A_i$ to $E_i$ and lying in the ladder region $L$. We encode each lattice path $P_i$ by its corresponding two-rowed array, $LT_i$, say. So we have to figure out what the above properties of the family $P$ mean in terms of two-rowed arrays. First, observe that it is sufficient to require only $P_1$ to lie in the ladder region $L$ since nonintersectingness forces all other paths to lie below $P_1$ and thus also to lie in $L$. Secondly, note that for all $i$ the lattice path $P_i$ has to start with $i - 1$ horizontal steps and has to end with $i - 1$ vertical steps, because otherwise $(P_1, P_2, \ldots, P_n)$ would not be nonintersecting. Therefore the first NE-turn of $P_i$ must be in the region north-east of $(i - 1, n - i + 1)$, and the last NE-turn of $P_i$ must be in the region south-west of $(a-n+i-1, b-i+1)$, $i = 1, 2, \ldots, n$. Hence, by the encoding of lattice paths in terms of two-rowed arrays, the family
(\(P_1, P_2, \ldots, P_n\)) in \(\mathcal{P}^+_{\mathcal{E}}(\mathbf{A} \to \mathbf{E})\) corresponds to the family \((LT_1, LT_2, \ldots, LT_n)\) of two-rowed arrays,

\[
\begin{align*}
LT_1: & \quad 0 \leq n - 1 < \begin{array}{cccc}
\hline
\vdots & \vdots & \vdots & \vdots \\
\end{array} < a - n + 1 \\
& \quad 1 \leq n - 2 < \begin{array}{cccc}
\hline
\vdots & \vdots & \vdots & \vdots \\
\end{array} \leq b - 1 \\
LT_n: & \quad n - 1 \leq 0 < \begin{array}{cccc}
\hline
\vdots & \vdots & \vdots & \vdots \\
\end{array} < a \leq b - n + 1 ,
\end{align*}
\]

(each \[\] representing a two-rowed array, both rows of which containing the same number of entries and obeying the indicated bounds), such that

(1) \(LT_i\) and \(LT_{i+1}\) are nonintersecting for all \(i\), meaning that if \(LT_i\) is given by

\[
i - 1 < a_1 \quad a_2 \quad \ldots \quad a_k < a - n + i \\
n - i < b_1 \quad b_2 \quad \ldots \quad b_k \leq b - i + 1 ,
\]

and if \(LT_{i+1}\) is given by

\[
i \leq c_1 \quad c_2 \quad \ldots \quad c_l < a - n + i + 1 \\
n - i - 1 < d_1 \quad d_2 \quad \ldots \quad d_l \leq b - i ,
\]

then there is no \(I, J\) such that (4.4c,d,e) holds,

(2) the array \(LT_1\), say

\[
LT_1 = x_1 \quad x_2 \quad \ldots \quad x_k \\
y_1 \quad y_2 \quad \ldots \quad y_k ,
\]

"lies in the ladder", meaning

\[
y_s < f(x_s)
\]

for all \(s\).

Of course, (4.6b) comes from the definition (2.1) of \(L\).

Summarizing, taking into account (4.3), the left-hand side of (2.2) can be written as

\[
\sum z^{(\text{number of entries in } (LT_1, LT_2, \ldots, LT_n))/2} ,
\]

where the sum is over all families \((LT_1, LT_2, \ldots, LT_n)\) of two-rowed arrays (4.5) satisfying (1) and (2) above.

The two-rowed array interpretation of the right-hand side of (2.2). Now we turn our attention to the right-hand side of (2.2). It is obvious that the right-hand side is the generating function

\[
\sum \text{sgn } \sigma z^{\text{NE}(\mathbf{P})} ,
\]

where the sum is over all pairs \((\sigma, \mathbf{P})\), with \(\sigma\) being a permutation in \(\mathfrak{S}_n\) (\(\mathfrak{S}_n\) is the symmetric group of order \(n\)) and \(\mathbf{P} = (P_1, P_2, \ldots, P_n)\) being a family of lattice
paths, \( P_i \) running from \( A_{\sigma(i)} = (0, n - \sigma(i)) \) to \( E_i = (a - n + i, b) \) and lying in the ladder region \( L \). Thus, by encoding each lattice path \( P_i \) by its corresponding two-rowed array, \( RT_i \) say, each family \( \{P_1, P_2, \ldots, P_n\} \) of such lattice paths corresponds to the family \( \{RT_1, RT_2, \ldots, RT_n\} \) of two-rowed arrays,

\[
\begin{align*}
RT_1: & \quad \begin{array}{cccc}
0 \leq & n - \sigma(1) & < & a - n + 1 \\
& & \leq & b
\end{array} \\
RT_2: & \quad \begin{array}{cccc}
0 \leq & n - \sigma(2) & < & a - n + 2 \\
& & \leq & b
\end{array} \\
\vdots & \quad \begin{array}{cccc}
0 \leq & n - \sigma(n) & < & a \\
& & \leq & b
\end{array}
\end{align*}
\]

(4.9)

with the same permutation \( \sigma \), where \( RT_i \) “lies in the ladder region \( L \)”, meaning that if

\[
\begin{align*}
RT_i &= \begin{array}{cccc}
x_1 & x_2 & \cdots & x_k \\
y_1 & y_2 & \cdots & y_k
\end{array} \\
\end{align*}
\]

(4.10a)

then

\[
y_s < f(x_s)
\]

(4.10b)

for all \( s \). Of course, (4.10b) comes from the definition (2.1) of \( L \).

However, it is more convenient to represent these two-rowed arrays in a shifted manner. To be precise, the first row of \( RT_i \) is shifted \( i - 1 \) units to the right relative to the second row. Thus we obtain the family

\[
\begin{align*}
RT_1: & \quad \begin{array}{cccc}
0 \leq & n - \sigma(1) & < & a - n + 1 \\
& & \leq & b
\end{array} \\
RT_2: & \quad \begin{array}{cccc}
0 \leq & n - \sigma(2) & < & a - n + 2 \\
& & \leq & b
\end{array} \\
\vdots & \quad \begin{array}{cccc}
0 \leq & n - \sigma(n) & < & a \\
& & \leq & b
\end{array}
\end{align*}
\]

(4.11)

(For the moment the reader should ignore the vertical dotted lines and the labels \( M_1, F_2, M_2, B_2, \text{ etc.} \). Originally, (4.10a,b) was satisfied for all \( i \). By the shifting procedure, (4.10a,b) transformed into the property that if

\[
\begin{align*}
RT_i &= \begin{array}{cccc}
\cdots x_j & x_{j+1} \cdots \\
\cdots y_j & y_{j+1} \cdots
\end{array}
\end{align*}
\]

(4.12a)
then
\[(4.12b) \quad y_s < f(x_s+i-1)\]
for all \(s\).

There is one point to be made precise. The alert reader will ask, how to understand shifting the first row of \(RT_i\) in (4.9) by \(i - 1\) units if \(RT_i\) contains less than \(i - 1\) entries in each row. In this case we simply say that some entries in the “shifted” \(RT_i\) in (4.11) are empty. To be more specific, divide each \(RT_i\) in (4.11) into its front part \(F_i\) (consisting of the first \(i - 1\) entries of the second row of \(RT_i\)), its back part \(B_i\) (consisting of the last \(i - 1\) entries of the first row of \(RT_i\)), and its middle part \(M_i\) (consisting of the rest), as indicated in (4.11). Then, if the original \(RT_i\) in (4.9) contained less than \(i - 1\) entries in each row, then the middle part \(M_i\) will be empty, and the front and back parts \(F_i\) and \(B_i\) will contain some empty entries, but both \(F_i\) and \(B_i\) will contain the same number of nonempty entries, and hence also the same number of empty entries. By convention, we always assume that the nonempty entries of the front part \(F_i\) occupy the first of the \(i - 1\) entries, and that the nonempty entries of the back part \(B_i\) occupy the last of the \(i - 1\) entries. Note that both \(F_i\) and \(B_i\) cannot contain an empty entry if \(M_i\) is nonempty.

Summarizing, taking into account (4.3), the right-hand side of (2.2), which can be rewritten as (4.8), equals
\[(4.13) \quad \sum \text{sgn} \sigma z^\text{(number of entries in } (RT_1,RT_2,\ldots,RT_n))/2, \]
where the sum is over all pairs \((\sigma, (RT_1, RT_2, \ldots, RT_n))\), where \(\sigma\) is a permutation in \(S_n\), and where \((RT_1, RT_2, \ldots, RT_n)\) is a family of two-rowed arrays of the form (4.11), with \(RT_i\) satisfying (4.12a,b) and

(e1) \(RT_i\)'s front and back parts \(F_i\) and \(B_i\) contain the same number of empty entries, but contain no empty entries if \(RT_i\)'s middle part \(M_i\) is nonempty. The nonempty entries in \(F_i\) occupy the first of the \(i - 1\) entries, while the nonempty entries in \(B_i\) occupy the last of the \(i - 1\) entries.

The front, middle, and back part of \((RT_1, RT_2, \ldots, RT_n)\). We shall frequently refer to the collection \((F_1, F_2, \ldots, F_n)\) of all front parts as the front part of \((RT_1, RT_2, \ldots, RT_n)\), and think of it to be arranged as

\[(4.14) \quad \begin{array}{cccc}
\text{n=\sigma(1)} & F_1 \\
\text{n=\sigma(2)} & F_2 \\
\vdots & \vdots \\
\text{n=\sigma(n)} & F_n \\
\end{array} \]

Note that we include the (lower) bounds in the picture. This has several advantages. One immediate advantage is that the information about the permutation \(\sigma\) is thus encoded in the front part. So we do not have to mention it separately all the time. Note that still all the rows in (4.14) are strictly increasing.

Similar conventions are made when we speak of the middle part and the back part of \((RT_1, RT_2, \ldots, RT_n)\). In particular, we think of the middle part as
where \( ff_i \) is the right-most nonempty entry in \( F_i \), if there is no nonempty entry in \( F_i \) then \( ff_i = n - \sigma(i) \), and where \( bb_i \) is the left-most nonempty entry in \( B_i \), if there is no nonempty entry in \( B_i \) then \( bb_i = a - n + i \).

Finally, we think of the back part to be arranged as

\[
\begin{array}{c}
\hdots \\
B_2 & a-\sigma+2 \\
B_3 & a-\sigma+3 \\
\vdots \\
B_n & a \\
\end{array}
\]

(4.16)

Note that similar as with the front part we include the (upper) bounds in the picture. In (4.16) also, all the rows are still strictly increasing.

Recall that in both (4.14) and (4.16) some entries are allowed to be empty, according to the rules in (e1).

Outline of the proof. Our strategy of proving (2.2) consists of showing that the total contribution to the sum (4.13) of the families \( (RT_1, RT_2, \ldots, RT_n) \) of two-rowed arrays (4.11) satisfying (4.12a,b) and (e1), that

(b1) either contain an entry \( \geq n \) or an empty entry in the front part,

(b2) or contain two middle parts with consecutive indices, say \( M_i \) and \( M_{i+1} \), \( M_i \) represented by

\[
\begin{array}{c}
\hdots \\
0 \leq a_1 a_2 \ldots a_k < bb_i \\
ff_i < b_1 b_2 \ldots b_k \leq b \\
\end{array}
\]

(4.17a)

and \( M_{i+1} \) represented by

\[
\begin{array}{c}
\hdots \\
0 \leq c_1 c_2 \ldots c_l < bb_{i+1} \\
ff_{i+1} < d_1 d_2 \ldots d_l \leq b \\
\end{array}
\]

(4.17b)
that cross (see the remark below explaining this terminology; recall (4.15) for the meaning of $ff_i, bb_i$, etc.), meaning that there exist $I$ and $J$ such that

\begin{align}
(4.17c) & \quad c_J < a_I, \\
(4.17d) & \quad b_{I-1} < d_J,
\end{align}

and

\begin{equation}
(4.17e) \quad 1 \leq I \leq k + 1, \quad 0 \leq J \leq l.
\end{equation}

Here, by convention $a_{k+1}$ is defined to be $bb_i$ (the first nonempty entry of $B_i$, or $a - n + i$ if there is no nonempty entry in $B_i$; recall (4.15)), $b_0$ is defined to be $ff_i$ (the last nonempty entry in $F_i$, or $n - \sigma(i)$ if there is no nonempty entry in $F_i$), $c_0$ is defined to be $-1$, and $d_0$ is defined to be $ff_{i+1}$ (the last nonempty entry of $F_{i+1}$, or $n - \sigma(i+1)$ if there is no nonempty entry in $F_{i+1}$).

We call the point $(a_I, d_J)$ a crossing point of $M_i$ and $M_{i+1}$, (b3) or contain an entry $\leq a - n$ or an empty entry in the back part, equals 0.

The terminology “crossing of $M_i$ and $M_{i+1}$” comes from the fact that when $M_i$ and $M_{i+1}$ are interpreted as paths, the path corresponding to $M_i$ running from $(0, ff_i)$ to $(bb_i, b)$, the one corresponding to $M_{i+1}$ running from $(0, ff_{i+1})$ to $(bb_{i+1}, b)$, then they would cross if and only if (4.17c,d,e) is satisfied for some $I$ and $J$ (at least if $ff_i \geq ff_{i+1}$). This is easily derived from (4.4c,d,e) and the observation that, given two nonintersecting lattice paths, by shifting the lower path in direction $(-1, 1)$ one obtains two paths that might touch but do not cross. (The subtle discrepancy between (4.17e) and (4.4e) in the range of $J$ has its explanation in that we will also have to deal with cases in which $ff_i < ff_{i+1}$. In case that $ff_i \geq ff_{i+1}$ the “crossing condition” (4.17c,d,e) cannot hold with $J = 0$. Hence, in this case (4.17e) and (4.4e) are really the same.)

Suppose that we would have shown that the contribution to (4.13) of these families of two-rowed arrays equals zero. It implies that only those families $(RT_1, RT_2, \ldots, RT_n)$ of two-rowed arrays (4.11) satisfying (4.12a,b) and (e1) contribute to (4.13) that

\begin{enumerate}
  \item [(g1)] contain only nonempty entries, all of them $< n$, in the front part,
  \item [(g2)] have a “non-crossing” middle part,
  \item [(g3)] and contain only nonempty entries, all of them $> a - n$, in the back part.
\end{enumerate}

Now, item (g1) says in particular that the last entry in $F_i$ is $< n$. But because of the bounds in (4.11) and the strict increase along rows it is also $\geq n - \sigma(i) + i - 1$. Hence we have $i \leq \sigma(i)$ for all $i$. This implies that $\sigma$ must be the identity permutation, and that the front part looks like

\[
\begin{array}{cccc}
n-1 & n-2 & n-1 \\
n-3 & n-2 & n-1 \\
\vdots & \ddots & \vdots \\
0 & 1 & \ldots & n-2 & n-1
\end{array}
\]

Similarly, item (g3) implies that the back part looks like
So, front part and back part are uniquely determined. Besides, they impose the following bounds on the middle part,

\[
\begin{array}{c|cc}
0 & M_1 & \leq a-n+1 \\
\hline
n-1 & \leq b
\end{array}
\]

(4.20)

It is now easily seen that these families \((M_1,M_2,\ldots,M_n)\) of middle parts are in bijection with the “left-hand side” families (4.5) satisfying (l1) and (l2). Namely, for each \(M_i\) add \(i-1\) to the entries of the first row and subtract \(i-1\) from the entries of the second row. Then, obviously, the bounds are shifted into the bounds of (4.5). Moreover, the non-crossing condition (4.17c,d,e) translates into the non-intersectingness condition (4.4c,d,e) (recall the remark concerning the discrepancy between (4.17e) and (4.4e), which disappears in the case of (4.20), since in (4.20) we have \(ff_i = ff_{i+1} = n-1\) for all \(i\), therefore (l1) is satisfied after this addition/subtraction. Also, \(M_1\) is not changed by the addition/subtraction procedure. And, \(M_1 = RT_1\) satisfies (4.12a,b) with \(i = 1\). Therefore (l2) (with \(LT_1\) replaced by \(RT_1 = M_1\)) is satisfied after this addition/subtraction.

Finally note that the factor \(z^{-\left(\begin{array}{c}2 \\ 2\end{array}\right)}\) in equality (2.2) corrects the contribution to \(z^{(\text{number of entries in } (RT_1, RT_2, \ldots, RT_n))/2}\) (cf. (4.7) and (4.13)) of the \(2\left(\begin{array}{c}n \\ 2\end{array}\right)\) discarded entries in the superfluous front and back parts (4.18) and (4.19).

Thus Theorem 1 would be proved.

The involution \(\varphi\). To show that the contribution of the families \((RT_1, RT_2, \ldots, RT_n)\) of two-rowed arrays (4.11) satisfying (4.12a,b), (e1), and at least one of (b1), (b2), or (b3), equals 0, we construct an involution, \(\varphi\) say, on this set of families that maps a family \((RT_1, RT_2, \ldots, RT_n)\) with associated permutation \(\sigma\) to a family \((\overline{RT_1}, \overline{RT_2}, \ldots, \overline{RT_n})\) with associated permutation \(\overline{\sigma}\), such that

\[
(i1) \quad \text{sgn } \sigma = - \text{sgn } \overline{\sigma},
\]

and such that

\[
(i2) \quad (\text{number of entries in } (RT_1, RT_2, \ldots, RT_n)) = (\text{number of entries in } (\overline{RT_1}, \overline{RT_2}, \ldots, \overline{RT_n})).
\]
Clearly, this implies that the contribution to (4.13) of families that are mapped to
each other cancels.

So it remains to construct this involution \( \varphi \). To give an outline, we partition
the set of all families \((RT_1, RT_2, \ldots, RT_n)\) that satisfy (4.12a,b), (e1), and one
of (b1), (b2), or (b3) into two disjoint subsets, namely \( B^1 \), the set of all families
satisfying (b3), and \( B^2 \), the set of all families not satisfying (b3) (i.e., satisfying
(g3)) but satisfying (b1) or (b2). Then we define involutions \( \varphi^1 \) and \( \varphi^2 \) on each
of \( B^1 \) and \( B^2 \), respectively. The involution \( \varphi \) then is simply the union of the two
“subinvolutions” \( \varphi^1 \) and \( \varphi^2 \). Basically, \( \varphi^1 \) is defined by the operation (4.21)/(4.22)
(which is inspired by [12, operation \( \varphi_{1.0} \) on pp. 181/182]), and \( \varphi^2 \) is defined by the
operation (4.23)/(4.24) (which is inspired by [13, operation (5.3.6)/(5.3.10)], [15],
[14, operation (24),(25)/(27)]).

The subinvolution \( \varphi^1 \). Let \((RT_1, RT_2, \ldots, RT_n)\) be an element of \( B^1 \), i.e., it is a
family of two-rowed arrays of the form (4.11) satisfying (4.12a,b), (e1), and (b3).

Consider the north-west to south-east diagonals in the back part (i.e., the diagonal
s running parallel to the right border of the back part, consult (4.16)). Search for
the right-most such diagonal that contains an entry that is empty or smaller than
the corresponding entry in the “maximal back part” (4.19) (which must exist by
(b3)). Call these entries violation entries. Among the entries of the next diagonal
to the right, pick the bottom-most element that has such a violation entry below.
(There must be at least one by assumption). Let this entry, \( a_I \) say, be in \( B_i \). Let
the violation entry below be \( c_I \), say. We then have the following situation,

\[
\begin{align*}
RT_i : & \quad 0 \leq \begin{array}{c|c|c|c|c}
M^1_{i} & M^2_{i} & \cdots & a_{i-1} & a_i \\
\sigma(n) & F_i & M^1_{i+1} & \cdots & a_{n+i+1}
\end{array} \\
& \quad \leq b \\
(i=1, & \cdots, i-1, i+1, & \cdots, i-l) \\
RT_{i+1} : & \quad 0 \leq \begin{array}{c|c|c|c|c}
M^1_{i+1} & M^2_{i+1} & \cdots & c_i & c_{i+1} \\
\sigma(n) & F_{i+1} & M^1_{i+2} & \cdots & a_{n+i+2}
\end{array} \\
& \quad \leq b \\
(i=1, & \cdots, i-1, i+1, & \cdots, i-l) 
\end{align*}
\]

where \( c_I \) is empty or smaller than the corresponding entry in (4.19). \( M^1_i \) and
\( M^2_i \) denote the first and second row of \( M_i \), respectively, while \( M^1_{i+1} \) and \( M^2_{i+1} \)
denote the first and second row of \( M_{i+1} \), respectively. Then, by definition, \( \varphi^1 \) maps
\((RT_1, \ldots, RT_i, RT_{i+1}, \ldots, RT_n)\) to \((RT_1, \ldots, RT'_i, RT'_{i+1}, \ldots, RT_n)\), where

\[
\begin{align*}
RT_i' : & \quad 0 \leq \begin{array}{c|c|c|c|c}
M^1_{i+1} & M^2_{i+1} & \cdots & c_i & a_{i+1} \\
\sigma(n) & F_{i+1} & M^1_{i+2} & \cdots & a_{n+i+2}
\end{array} \\
& \quad \leq b \\
(i=1, & \cdots, i-1, i+1, & \cdots, i-l) \\
RT_{i+1}' : & \quad 0 \leq \begin{array}{c|c|c|c|c}
M^1_i & M^2_i & \cdots & a_{i-1} & c_{i+1} \\
\sigma(n) & F_i & M^1_{i+1} & \cdots & a_{n+i+1}
\end{array} \\
& \quad \leq b \\
(i=1, & \cdots, i-1, i+1, & \cdots, i-l)
\end{align*}
\]
It should be observed that the permutation $\sigma$ corresponding to $\varphi^1((RT_1,\ldots,RT_n)) = (RT_1,\ldots,RT_i,RT_{i+1}',\ldots,RT_n)$ is $\sigma = \sigma \circ (i,i+1)$. We remark that the displays (4.21)/(4.22) should be understood in the sense that $a_I$ and $c_{I+1}$ (which are shown to be located inside the back part) could also be the bounds $a - n + i$ and $a - n + i + 1$, respectively. A similar remark applies to $a_{i-1}$, which could also be the last entry of $M_{i1}$, the first row of the middle part $M_i$. However, the operation (4.21)/(4.22) is still well-defined in all these cases.

We must show that $\varphi^1$ is well-defined, i.e., that the new family $(RT'_1,\ldots,RT'_i,RT'_{i+1},\ldots,RT'_n)$ again belongs to $B^1$, and that $\varphi^1$ is an involution that satisfies (i1) and (i2).

For being well-defined, we have to check that the rows in $RT'_i$ and $RT'_{i+1}$ are still strictly increasing, and that (4.12a,b), (e1), and (b3) are satisfied.

For the first property, we have to show $c_I < a_I$ (if $c_I$ is nonempty, if otherwise nothing is to show) and $a_{I-1} < c_{I+1}$ (if $a_{I-1}$ is nonempty, if otherwise nothing is to show). Assuming that $c_I$ is nonempty, the first inequality holds because $a_I$ equals its corresponding entry in (4.19), but $c_I$ is less than the corresponding entry in (4.19), which is also $a_I$ (because entries in (4.19) are the same along columns). The second inequality holds since we must have $a_{I+1} = c_{I+1}$ (again, because entries in (4.19) are the same along columns), and hence $a_{I-1} < a_I < a_{I+1} \leq c_{I+1}$.

Next we check (e1) for $RT'_i$ and $RT'_{i+1}$. By assumption, $a_I$ and $c_{I+1}$ are nonempty. Therefore the correct number of entries will be empty in both $RT'_i$ and $RT'_{i+1}$ since it is in $RT_i$ and $RT_{i+1}$.

In order to see that (4.12a,b) holds for $RT'_i$ and $RT'_{i+1}$, observe that, because of the shift of the first rows relative to the second rows when passing from (4.21) to (4.22), entries in $RT'_i$ and $RT'_{i+1}$ related to each other by (4.12b) are still related to each other in $RT'_i$ and $RT'_{i+1}$, except for $a_I$ and $c_{I+1}$ and the entries to the right of these in the same rows, together with their related “partners” in the second rows, respectively. But $a_I$ and $c_{I+1}$ and all the entries to the right are $> a - n$ (compare with (4.19)), in which case (4.12b) is satisfied automatically because of our assumption $f(x) = b + 1$ for $x > a - n$.

Finally, also (b3) is satisfied by the new family. For, either the first entry in the back part of $RT'_{i+1}$ is empty. Or it is nonempty. In this case also $a_{I-1}$ is nonempty and so is smaller than $a_I$. Therefore $a_{I-1}$ (as entry in $RT'_{i+1}$) is already smaller than its corresponding entry in (4.19) (which is exactly $a_I$). By strict increase along rows this implies that the first entry in the back part of $RT'_{i+1}$ has to be $\leq a - n$.

Next we want to establish that $\varphi^1$ is an involution, i.e., that by applying $\varphi^1$ to $(RT_1,\ldots,RT'_i,RT'_{i+1},\ldots,RT_n)$ one obtains again $(RT_1,\ldots,RT_i,RT_{i+1},\ldots,RT_n)$. As already observed, $a_{I-1}$ (as entry in $RT'_{i+1}$) is either empty or smaller than its corresponding entry in (4.19). Furthermore, the diagonal containing $a_{I-1}$ is also the right-most diagonal in the back part of $\varphi^1((RT_1,\ldots,RT_n))$ that contains such a “violation entry”, because the diagonals to the right were not changed under application of $\varphi^1$. And, $a_I$ (as entry in $RT'_i$) is still the bottom-most entry in the next diagonal to the right that has such a violation entry below, because nothing was changed below or to the right of $c_{I+1}$ under application of $\varphi^1$. But this exactly means that we come back by applying $\varphi^1$ a second time, as required. Hence, $\varphi^1$ is an involution on $B^1$. That $\varphi^1$ satisfies (i1) is obvious since $\sigma$ is mapped to
\[ \sigma = \sigma \circ (i, i + 1), \] as mentioned before. Also, the numbers of entries in the families of two-rowed arrays are the same before and after application of \( \varphi^1 \), which settles (12).

The subinvolution \( \varphi^2 \). Let \( (RT_1, RT_2, \ldots, RT_n) \) be an element of \( B^2 \), i.e., it is a family of two-rowed arrays of the form (4.11) satisfying (4.12a,b), (e1), (g3), and at least one of (b1) or (b2). Recall that (g3) implies that the back part of \( (RT_1, RT_2, \ldots, RT_n) \) looks like (4.19). In particular, there are no empty entries in the back part, and hence there are no empty entries in the front part either.

We claim that if \( (RT_1, RT_2, \ldots, RT_n) \) satisfies (b1) then it already satisfies (b2). Suppose that there is an entry \( \geq n \) in the front part. Consider the top-most row of the front part, \( F_{i+1} \) say, containing such an entry. In particular, the last entry in \( F_{i+1} \), \( ff_{i+1} \), is \( \geq n \), and the last entry in \( F_i \), \( ff_i \), is \( < n \). Hence, \( ff_i < ff_{i+1} \). But then \( 0, ff_{i+1} \) is a crossing point of \( M_i \) and \( M_{i+1} \) (choose \( I = 1, J = 0 \) in (4.17a,b,c,d,e)). So (b2) is satisfied.

Now consider all crossing points of middle parts with consecutive indices. Among these points choose those with maximal \( x \)-coordinate, and among all these choose the crossing point with maximal \( y \)-coordinate. Denote this crossing point by \( S \). Let \( i \) be minimal such that \( S \) is a crossing point of \( M_i \) and \( M_{i+1} \). We have the following situation,

\[ \begin{align*}
RT_i : & \quad 0 \leq \begin{array}{c}
\ldots \ a_{i-1} : a_i \\
\vdots \\
F_i \\
\vdots \\
\ldots \ b_{i-1} : b_i \\
\leq b
\end{array} \\
\text{(4.23)} & \quad \begin{array}{c}
B_i \\
\ldots \ a-n+i
\end{array}
\end{align*} \]

\[ \begin{align*}
RT_{i+1} : & \quad 0 \leq \begin{array}{c}
\ldots \ c_{i+1} : c_i \\
\vdots \\
F_{i+1} \\
\vdots \\
\ldots \ d_{i-1} : d_i \\
\leq b
\end{array} \\
\text{(4.24)} & \quad \begin{array}{c}
B_{i+1} \\
\ldots \ a-n+i+1
\end{array}
\end{align*} \]

where \( S = (a_i, d_i) \) and where (4.17a,b,c,d,e) is satisfied for the middle parts \( M_i \) and \( M_{i+1} \). Then, by definition, \( \varphi^2 \) maps \( (RT_1, \ldots, RT_i, RT_{i+1}, \ldots, RT_n) \) to \( (RT_1, \ldots, RT''_i, RT''_{i+1}, \ldots, RT_n) \), where \( RT''_i \) and \( RT''_{i+1} \) are given by

\[ \begin{align*}
RT''_i : & \quad 0 \leq \begin{array}{c}
\ldots \ a_{i-1} : a_i \\
\vdots \\
F_{i+1} \\
\vdots \\
\ldots \ b_{i-1} : b_i \\
\leq b
\end{array} \\
\text{(4.25)} & \quad \begin{array}{c}
B_{i+1} \\
\ldots \ a-n+i+1
\end{array}
\end{align*} \]

\[ \begin{align*}
RT''_{i+1} : & \quad 0 \leq \begin{array}{c}
\ldots \ c_{i+1} : c_i \\
\vdots \\
F_i \\
\vdots \\
\ldots \ d_{i-1} : d_i \\
\leq b
\end{array} \\
\text{(4.26)} & \quad \begin{array}{c}
B_i \\
\ldots \ a-n+i
\end{array}
\end{align*} \]
Again, observe that the permutation \( \sigma \) corresponding to the new family of two-rowed arrays is \( \sigma = \sigma \circ (i, i + 1) \). We remark that for \( I = 1, k + 1 \) or \( J = 0, l \) (consult (4.17a,b) for the meaning of \( k \) and \( l \)) the displays (4.23)/(4.24) should be understood in the sense that some entries that in the displays are shown to be located in the middle part could be actually located in the front or back part (e.g. \( b_{l-1} \) for \( I = 1 \) would be located in the front part, while \( a_{l-1} \) for \( I = 1 \) would even be non-existent). However, the operation (4.23)/(4.24) is still well-defined in all these cases.

Also here, we have to show that \( \varphi^2 \) is well-defined, i.e., that \((RT_1', \ldots, RT_i'', RT_{i+1}'', \ldots, RT_n'')\) belongs to \( B^2 \), and that \( \varphi^2 \) is an involution that satisfies (i1) and (i2).

For being well-defined, we have to check that the rows in \( RT_i'' \) and \( RT_{i+1}'' \) are still strictly increasing, and that (4.12a,b), (e1), (g3), and (b2) are satisfied. For proving that the rows of \( RT_i'' \) and \( RT_{i+1}'' \) are strictly increasing we have to show \( c_J < a_I, d_{J-1} < b_I, a_{I-1} < c_{J+1} \) (if \( I > 1 \), there is nothing to show if \( I = 1 \), and \( b_{I-1} < d_J \). The first and the last inequality hold by (4.17c) and (4.17d), respectively. Supposing that the second inequality does not hold, we obtain \( c_J < a_I < a_{I+1} \) and \( b_I \leq d_{J-1} < d_J \). But by (4.17c,d,e) this means that \((a_I, d_J)\) is a crossing point of \( M_i \) and \( M_{i+1} \), contradicting the “maximality” of \( S = (a_I, d_J) \).

Similarly, if we suppose that the third inequality does not hold, we obtain \( c_{I+1} < a_I \) and \( b_{I-1} < d_J < d_{J+1} \). By (4.17c,d,e) this means that \((a_I, d_{J+1})\) is a crossing point of \( M_i \) and \( M_{i+1} \), again contradicting the “maximality” of \( S = (a_I, d_J) \).

The property (e1) is trivially satisfied since, as we already observed, there are no empty entries in \( RT_i \) and \( RT_{i+1} \), and so there are no empty entries in \( RT_i'' \) and \( RT_{i+1}'' \).

The next property to be checked is (4.12a,b). Of course, there is no problem with entries in the second rows to the right of the cuts in \( RT_i'' \) and \( RT_{i+1}'' \) (indicated by vertical dotted lines in (4.24)) and their “partners” related by (4.12b) in the first rows, since this part was not affected by \( \varphi^2 \). Also there is no problem with entries in the first rows to the left of the cuts in \( RT_i'' \) and \( RT_{i+1}'' \) and their partners related by (4.12b) in the second rows because of the shift of the first rows relative to the second rows that happens in these parts when passing from (4.23) to (4.24). However, there could be a problem if partners were “divorced” under application of \( \varphi^2 \). This is the situation of one entry (being in a first row to the right of the cut staying at its place, the related other entry (being in a second row to the left of the cut) moving to the other two-rowed array.

The following is the typical situation,

\[
RT_i : \quad \ldots \quad a_{I-1} \quad | \quad a_I \quad \ldots \quad a_{I+r-1} \\
\quad \quad \quad \quad b_{I+r-i} \quad \ldots \quad b_{I-1} \quad | \quad b_I \quad \ldots
\]

\[
RT_{i+1} : \quad \ldots \quad c_J \quad | \quad c_{J+1} \quad \ldots \quad c_{J+r} \\
\quad \quad \quad \quad d_{J+r-i} \quad \ldots \quad d_{J-1} \quad \quad \quad | \quad d_J \quad \ldots
\]  

and after application of \( \varphi^2 \),
where \( r \) satisfies \( 1 \leq r \leq i - 1 \). Of course, any of \( a_{I+r-1}, b_{I+r-i}, c_{J+r}, d_{J+r-i} \) could be non-existent. We shall take care of these possibilities, too.

Since \( RT_i \) satisfies (4.12a,b) we have \( b_s < f(a_{s+i-1}) \) for all \( s \), and since \( RT_{i+1} \) also satisfies (4.12a,b) we have \( d_s < f(c_{s+i}) \) for all \( s \). In particular, there hold
\[
b_{I+r-i} < f(a_{I+r-1}) \quad \text{(provided both \( a_{I+r-1} \) and \( b_{I+r-i} \) exist)} \quad \text{and} \quad d_{J+r-i} < f(c_{J+r}) \quad \text{(provided both \( c_{J+r} \) and \( d_{J+r-i} \) exist), for} \quad 1 \leq r \leq i - 1.\]

We have to show that \( RT_i'' \) and \( RT_{i+1}'' \) also satisfy (4.12a,b), i.e., that
\[
d_{J+r-i} < f(a_{I+r-1}), \tag{4.27}
\]
provided both \( a_{I+r-1} \) and \( d_{J+r-i} \) exist (if either \( a_{I+r-1} \) or \( d_{J+r-i} \) does not exist there is nothing to show), and that
\[
b_{I+r-i} < f(c_{J+r}), \tag{4.28}
\]
provided both \( b_{I+r-i} \) and \( c_{J+r} \) exist (if either \( b_{I+r-i} \) or \( c_{J+r} \) does not exist there is nothing to show).

For convenience, if \( a_{I+r-1} \) does not exist we put \( f(a_{I+r-1}) := b + 1 \), and if \( c_{J+r} \) does not exist we put \( f(c_{J+r}) := b + 1 \). Also, we put \( b_{k+1} := b + 1 \) (consult (4.17a) for the meaning of \( k \)).

We distinguish between two cases.

Case 1: \( f(a_{I+r-1}) \leq f(c_{J+r}) \). This immediately proves (4.28), since if both \( b_{I+r-i} \) and \( c_{J+r} \) exist then we have \( b_{I+r-i} < f(a_{I+r-1}) \leq f(c_{J+r}) \). For proving (4.27), first consider the case \( I = k + 1 \) (consult (4.17a,e) for the meaning of \( k \)). Then \( a_{I+r-1} = a_{k+r} \) is in the back part \( B_i \) of \( RT_i \), hence \( a_{k+r-1} > a - n \) (recall that the back part has to equal (4.19)). By one of the assumptions about \( f \), this implies \( f(a_{k+r-1}) = b + 1 \) in which case (4.27) is automatically satisfied. Therefore, in the sequel we may assume \( I \leq k \). Now suppose that (4.27) does not hold, i.e., that both \( a_{I+r-1} \) and \( d_{J+r-i} \) exist and that \( d_{J+r-i} \geq f(a_{I+r-1}) \). We claim that \( d_{J+r-i} \leq b_I \). For suppose \( d_{J+r-i} > b_I \). By (4.17c) we have \( c_j < a_I < a_{I+1} \), and by our assumption \( b_I < d_{J+r-i} < d_J \). But by (4.17c,d,e) this means that \( a_{I+1}, d_J \) is a crossing point of \( M_i \) and \( M_{i+1} \), contradicting the “maximality” of \( S = (a_I, d_J) \).

Therefore, we have indeed \( d_{J+r-i} \leq b_I \). This implies \( b_I \geq d_{J+r-i} \geq f(a_{I+r-1}) \geq f(a_I) \), the last inequality being true because of \( a_{I+r-1} \geq a_I \) and the monotonicity of \( f \). This says that \( (a_I, b_I) \) is a point outside the ladder region \( L \). Now, to give an intuitive argument first, view the middle parts \( M_1, M_2, \ldots, M_i \) as paths, \( P_1, P_2, \ldots, P_i \) say, \( P_j \) running from \( (0, ff_{j}) \) to \( (a - n + 1, b) \), \( j = 1, 2, \ldots, i \) (recall (4.15) for the meaning of \( ff_j \) and \( bb_j \), also recall that for all \( j \) we have \( bb_j = a - n + 1 \) because the back part looks like (4.19)). The point \( (a_I, b_I) \) is a NE-turn of the path \( P_i \) corresponding to \( M_i \). Furthermore, \( M_1, M_2, \ldots, M_i \) cannot have a crossing point strictly to the right of \( (a_I, b_I) \) because \( (a_I, d_J) \) is the “maximal” crossing point. Therefore the subpath of \( P_i \) from \( (a_I, b_I) \) to its end point \( (a - n + 1, b) \) forms
a barrier above which all the other paths $P_1, \ldots, P_{l-1}$ have to stay, being allowed to touch $P_l$ but not being allowed to cross. The point $(a_l, b_l)$ lies outside (to be precise: above) the ladder region $L$. Therefore all the other paths, staying above or on $(a_l, b_l)$, have to leave $L$. In particular, the first path $P_1$, which corresponds to $M_1$, would leave the ladder region $L$. This contradicts the fact that $M_1 = RT_1$ satisfies (4.12a,b) with $i = 1$.

To give a rigorous argument, consider $M_{i-1}$ and $M_i$. Let $M_{i-1}$ be given by

$$0 \leq x_1 x_2 \ldots x_m < a - n + 1,$$

$$ff_{i-1} < y_1 y_2 \ldots y_m \leq b$$

and, as before, let $M_i$ be given by

$$0 \leq a_1 a_2 \ldots a_k < a - n + 1,$$

$$ff_i < b_1 b_2 \ldots b_k \leq b$$

For convenience, set $x_0 := -1$, $y_0 := ff_{i-1}$, and $x_{m+1} := a - n + 1$. Let $s$ be maximal such that $x_s \leq a_l$. Since $l \leq k$, and so $a_l \leq a - n$, we have $s \leq m$. Suppose that $y_s < b_l$. By (4.17a,b,c,d,e) (with the appropriate replacements, namely applied to $M_{i-1}$ and $M_i$) it follows that $(x_{s+1}, b_l)$ is a crossing point of $M_{i-1}$ and $M_i$, where $x_{s+1} > a_l$. Once more this contradicts the “maximality” of $S = (a_l, d_l)$. Therefore we actually have $y_s \geq b_l$. Continuing in this manner with $M_{i-2}$ and $M_{i-1}, \ldots, M_1$ and $M_2$, we would find a sequence $(a_1, b_1), (x_s, y_s), \ldots, (u_t, v_t)$, with $a_l, b_l$ being entries in $M_{i-1}$, $a_l$ exactly above $b_l$, with $x_s, y_s$ being entries in $M_{i-1}$ (also being allowed to be $-1, ff_{i-1}$), $x_s$ exactly above $y_s$, \ldots, with $u_t, v_t$ being entries in $M_1$ (also being allowed to be $-1, ff_1$), $u_t$ exactly above $v_t$, and with $a_l \geq x_s \geq \cdots \geq u_t$ and $b_l \leq y_s \leq \cdots \leq v_t$. Since $(a_l, b_l)$ lies outside $L$, the latter inequality chains imply that also $(u_t, v_t)$ lies outside $L$. But this is impossible since $M_1 = RT_1$ satisfies (4.12a,b) with $i = 1$. (Also $v_t = ff_1$ is impossible since then we would have $ff_1 = v_t \geq b_l \geq f(a_l)$, but also $ff_1 = n - \sigma(1) \leq n - 1 < f(a_l)$, the last inequality being due to one of the assumptions about $f$.) Therefore it follows that $d_{J+r-1} < f(a_{I+r-1})$, which settles (4.27).

Case 2: $f(a_{I+r-1}) \geq f(c_{J+r})$. This immediately proves (4.27), since if both $a_{I+r-1}$ and $d_{J+r-1}$ exist then we have $a_{I+r-1} < f(c_{J+r}) \leq f(a_{I+r-1})$. For proving (4.28), first consider the case $J = l$ (consult (4.17b,e) for the meaning of $l$). Then $c_{J+r} = c_{l+r}$ is in the back part $B_{l+1}$ of $RT_{l+1}$, hence $c_{J+r} > a - n$ (recall that the back part has to equal (4.19)). By one of the assumptions about $f$, this implies $f(c_{J+r}) = b + 1$ in which case (4.28) is automatically satisfied. Therefore, in the sequel we may assume $J < l$. Now suppose that (4.28) does not hold, i.e., that both $b_{I+r-1}$ and $c_{J+r}$ exist and that $b_{I+r-1} \geq f(c_{J+r})$. First observe that there holds $a_1 \leq c_{J+1}$, because otherwise we would have $c_{J+1} < a_l$ and by (4.17d) also $b_{I-1} < d_J < d_{J+1}$, which by (4.17c,d,e) means that $(a_l, d_{J+1})$ is a crossing point of $M_1$ and $M_{I+1}$, contradicting the “maximality” of $S = (a_l, d_l)$. Therefore, by monotonicity of $f$, we have $f(a_l) \leq f(c_{J+1}) \leq f(c_{J+r}) < b_{I+r-1} < b_l$. Again, this says that $(a_l, b_l)$ is a point outside the ladder region $L$. Also here we have $I < k$ since $a_l \leq c_{J+1} \leq a - n$, the last inequality being true because of $J < l$. As before, this contradicts the fact that $M_1 = RT_1$ satisfies (4.12a,b) with $i = 1$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Finally, condition (g3) is certainly satisfied for $RT''_i$ and $RT''_{i+1}$ since they have the same back parts as $RT_i$ and $RT_{i+1}$, respectively. And, because of $a_I - 1 < a_I$ and $d_{I-1} < d_I$, condition (h2) is satisfied for $RT''_i$ and $RT''_{i+1}$, again with crossing point $(a_I, d_I)$.

The last observation, together with the fact that under application of $\varphi^2$ nothing is changed to the right of $a_I, b_I, c_{I+1}, d_I$, shows that application of $\varphi^2$ to $(RT_1, \ldots, RT'_i, RT''_i, \ldots, RT_n)$ gives back $(RT_1, \ldots, RT_i, RT_{i+1}, \ldots, RT_n)$. Hence, $\varphi^2$ is an involution on $B^2$. It satisfies (ii) since $\sigma$ is mapped to $\sigma = \sigma \circ (i, i + 1)$, as observed before. Also, the numbers of entries in the families of two-rowed arrays are the same before and after application of $\varphi^2$, which settles (12).

This finishes the proof of Theorem 1.

5. A weighted generalization of Theorem 1

Examining the proof of Theorem 1, it is easily seen that we actually proved a stronger result.

Let $x = (x_0, x_1, x_2, \ldots)$ and $y = (y_0, y_1, y_2, \ldots)$ be sequences of indeterminates. Given a path $P_i$ with array representation (4.4a) we define a weight for $P_i$ by

$$w_{x,y}(P_i) = \prod_{i=1}^{k} x_i y_i.$$  

(5.1)

For example, the weight of the path displayed in Figure 1 is $w_{x,y}(P_0) = x_1 x_2 x_5 y_1 y_3 y_4$. This weight is extended in a shifted manner to families $P = (P_1, P_2, \ldots, P_n)$ of lattice paths by

$$w_{x,y}(P) = \prod_{j=1}^{r} w_{(x_{i+j+1}, x_{j-1}, \ldots), (y_{i+j}, y_j, \ldots)}(P_j).$$  

(5.2)

Then we have the following generalization of Theorem 1.

Theorem 3. With the assumptions of Theorem 1 there holds

$$GF(P_L^+(A \rightarrow E); w_{x,y}) = (x_{a-1} x_{a-2} \cdots x_{a-n+1} y_{n-1} \cdots y_2 y_1)^{-1} \cdot \det_{1 \leq i, j \leq n} (GF(P_L(A_i \rightarrow E_s); w_{x,y})).$$  

(5.3)

Clearly, Theorem 3 specializes to Theorem 1 by setting $x_i = z^{1/2}$ and $y_i = z^{1/2}$ for all $i$.

6. Enumeration of single lattice paths in a ladder with respect to turns

The entries in the determinant in (2.2) are all generating functions $\sum z^{NE(P)}$ for single lattice paths $P$ with given starting and end point that stay in the ladder region $L$, where $L$ is of the form (2.1). Hence, we have to say how these can be computed. Of course, a “nice” formula cannot be expected in general. There are only two cases in which “nice” formulas exist, the case of the trivial ladder (i.e., $f(x) \equiv b+1$; see (6.4)), and the case of a ladder determined by a diagonal boundary.
(i.e., \( f(x) = x + D + 1 \), for some positive integer \( D \); see (6.6)). In all other cases one has to be satisfied with answers of recursive nature.

Actually, the problem was already treated by Kulkarni [18]. We restate his formula here, and include a proof since the proof in [18] is obscured by notation. Then we derive an alternative formula, which has the same complexity as Kulkarni’s formula, but is superior for ladder regions of a particular kind. This is discussed in more detail in the paragraph after (6.10).

Kulkarni’s formula [18, Theorem 4] reads as follows. (Note that Kulkarni considers one-sided lower ladders instead of one-sided upper ladders as we do. It is however no difficulty to translate everything by a rotation by 180°.) It is formulated only with the starting point of the lattice paths being \((0,0)\) and the end point being \((a,b)\), the top-right corner of the ladder \(L\). Of course, this is sufficient to determine the generating function for lattice paths inside the ladder region \(L\) with an arbitrary starting point and an arbitrary end point, since a change of coordinates will always bring us into this situation.

**Proposition 4.** Let the ladder region \(L\) be given by the weakly increasing function \( f : [0, a] \rightarrow [1, b + 1] \) by means of (2.1), as before. Furthermore, let \(-1 = a_k < a_{k-1} < \cdots < a_1 < a_0 = a - 1\) be a partition of the (integer) interval \([0, a]\) such that \( f \) is constant on each subinterval \([a_i + 1, a_{i-1}], i = k, k - 1, \ldots, 1\). Then the generating function \( \sum z^{\text{NE}(P)} \) for all lattice paths \( P \) from \((0,0)\) to \((a,b)\) in the ladder region \(L\) equals

\[
\text{GF}_L(P((0,0)\rightarrow(a,b)); z^{\text{NE}()}) = \sum_{\mathbf{e} \geq \mathbf{f} \geq 0} z^{\mathbf{e}} \prod_{i=1}^{k} \left( \frac{a_i - a_i - a_{i-1}}{e_i - e_{i-1}} \right) \left( \frac{f(a_{i-1} - f(a_i))}{f_i - f_{i-1}} \right),
\]

where \( \mathbf{e} = (e_1, e_2, \ldots, e_k) \) and \( \mathbf{f} = (f_1, f_2, \ldots, f_k) \), where, by definition, \( e_0 = f_0 = f(-1) = 0 \), and where \( \mathbf{e} \geq \mathbf{f} \geq 0 \) means \( e_i \geq f_i \geq 0 \), \( i = 1, 2, \ldots, k \).

**Proof.** Let \( P \) be a lattice path from \((0,0)\) to \((a,b)\) that stays in the ladder region \(L\). Let the NE-turns of \( P \) be \((x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\), ordered from top-right to bottom-left, i.e., \( x_1 > x_2 > \cdots > x_l \) and \( y_1 > y_2 > \cdots > y_l \). Suppose that there are \( e_i \) NE-turns with \( x\)-coordinates in the interval \([a_i + 1, a_0] = [a_i + 1, a - 1]\), and that there are \( f_i \) NE-turns with \( y\)-coordinates in the interval \([f(a_i) + 1, f(a_0)] = [f(a_i) + 1, f(a - 1)]\), \( i = 1, 2, \ldots, k \). In particular, there holds \( e_k = f_k = \ell \). Equivalently, we have

\[
a - 1 = a_0 \geq x_1 > \cdots > x_{e_1} > a_1 \geq x_{e_1 + 1} > \cdots > x_{e_2} > a_2 \geq \cdots \geq a_{k-1} \geq x_{e_{k-1} + 1} > \cdots > x_{e_k} > a_k = -1
\]

and

\[
b = f(a_0) \geq y_1 > \cdots > y_{e_1} > f(a_1) \geq y_{e_1 + 1} > \cdots > y_{e_2} > f(a_2) \geq \cdots \geq f(a_{k-1}) \geq y_{e_{k-1} + 1} > \cdots > y_{e_k} > f(a_k) = 0.
\]

It is immediate that we must have \( e_i \geq f_i \geq 0 \). (This is most easily seen by considering a picture of some ladder.) Conversely, given integer vectors \( \mathbf{e} \) and \( \mathbf{f} \)
with \( e_i \geq f_i \geq 0 \) and \( e_k = f_k \), by (6.2) and (6.3) there are

\[
\prod_{i=1}^{k} \left( a_{i-1} - a_i \right) \left( f(a_{i-1}) - f(a_i) \right)
\]

possible choices for the \( x- \) and \( y- \) coordinates of the NE-turns of a lattice path that has \( e_i \) NE-turns with \( x- \) coordinates in the interval \([a_i + 1, a_0]\), and \( f_i \) NE-turns with \( y- \) coordinates in the interval \([f(a_i) + 1, f(a_0)]\). Thus, (6.1) is established.

Now we describe the announced alternative method to compute the generating function \( \sum z^{\text{NE}(P)} \) for lattice paths \( P \) with given starting point \( A \) and given end point \( E \) that stay in the ladder region \( L \), where \( L \) is of the form (2.1). Let \( \mathcal{P}^h(A \rightarrow E) \) (respectively \( \mathcal{P}^v_h(A \rightarrow E) \)) denote the subset of \( \mathcal{P}(A \rightarrow E) \) (respectively \( \mathcal{P}_L(A \rightarrow E) \)) consisting of all lattice paths that start with a horizontal step. Similarly, let \( \mathcal{P}^v(A \rightarrow E) \) (respectively \( \mathcal{P}^v_L(A \rightarrow E) \)) denote the subset of \( \mathcal{P}(A \rightarrow E) \) (respectively \( \mathcal{P}_L(A \rightarrow E) \)) consisting of all lattice paths that start with a vertical step.

This second method is based on the following simple facts.

**Proposition 5.** Let \( A = (\alpha_1, \alpha_2) \) and \( E = (\eta_1, \eta_2) \) be lattice points. Then there hold

\[
\text{GF}(\mathcal{P}(A \rightarrow E); z^{\text{NE}(.)}) = \sum_{k \geq 0} \binom{\eta_1 - \alpha_1}{k} \binom{\eta_2 - \alpha_2}{k} z^k, \quad \text{if } \eta_1 > \alpha_1, \quad \eta_1 = \alpha_1.
\]

**Proposition 6.** Let \( L_D \) be a “diagonal” ladder determined by the function \( f(x) = x + D + 1 \) by means of (2.1). Let \( A = (\alpha_1, \alpha_2) \) and \( E = (\eta_1, \eta_2) \) be lattice points in the ladder \( L_D \). Then there hold

\[
\text{GF}(\mathcal{P}_{L_D}(A \rightarrow E); z^{\text{NE}(.)}) = \sum_{k \geq 0} \binom{\eta_1 - \alpha_1 - 1}{k} \binom{\eta_2 - \alpha_2}{k} z^k, \quad \eta_1 > \alpha_1,
\]

\[
\text{GF}(\mathcal{P}^v_{L_D}(A \rightarrow E); z^{\text{NE}(.)}) = \sum_{k \geq 0} \binom{\eta_1 - \alpha_2 + D - 1}{k} \binom{\eta_2 - \alpha_1 - D}{k} z^k, \quad \eta_1 = \alpha_1.
\]
Proposition 7. Let \( L \) be an arbitrary ladder of the form (2.1), and let \( A, E \) be lattice points in \( L \). Then, for all \( x \in [0, a] \) there holds
\[
\text{GF}(P_L(A \rightarrow E); z^{\text{NE}(j)}) = \sum_{j \geq x} \text{GF}(P_L(A \rightarrow (j, f(x) - 1)); z^{\text{NE}(j)}) \cdot \text{GF}(P_L((j, f(x) - 1) \rightarrow E); z^{\text{NE}(j)}).
\] (6.8)

Proof of the Propositions. Identity (6.4) is immediate from a glance at (4.4a).

For (6.5) we observe
\[
(6.9)\quad \text{GF}(P^v(A \rightarrow E); z^{\text{NE}(j)}) = \text{GF}(P(A \rightarrow E); z^{\text{NE}(j)}) - \text{GF}(P^h(A \rightarrow E); z^{\text{NE}(j)}) = \text{GF}(P(A \rightarrow E); z^{\text{NE}(j)}) - \text{GF}(P(A + (1, 0) \rightarrow E); z^{\text{NE}(j)}).
\]

Application of (6.4) in the right-hand side of (6.9) then leads to (6.5).

Identity (6.6) is the special case \( q = 1, \mu_i = \alpha_i, \lambda_i = \eta_i, c \to \infty, d = D, \) of [16, Theorem 1]. Equation (6.9) is also true with \( P_{L,D} \) instead of \( P \). Application of (6.6) in the right-hand side of (6.9), with \( P \) replaced by \( P_{L,D} \), then leads to (6.7).

The proof of (6.8) more or less consists only of Figure 6 (the dots indicating some ladder \( L \)) and the observation that any path from \( A \) to \( E \) has to run through exactly one of the marked vertical pieces.

Now, here is the second method for determining \( \text{GF}(P_L(A \rightarrow E); z^{\text{NE}(j)}) \) for any given ladder \( L \) of the form (2.1), with points \( A = (\alpha_1, \alpha_2) \) and \( E = (\eta_1, \eta_2) \) located inside \( L \); Partition the border of \( L \), i.e., the set of points \( \{(x, f(x)) : x \in [0, a]\} \) into horizontal and diagonal pieces, say \( L^1, L^2, \ldots, L^m \), where \( L^j = \{(x, f(x)) : x_{i-1} < x \leq x_i\} \), for some \(-1 = x_0 < x_1 < x_2 < \cdots < x_m = a\), each \( L^j \) being either horizontal or diagonal. Then apply the recurrence (6.8) in succession with \( x = x_{m-1}, x_{m-2}, \ldots, x_1 \) and use (6.4)–(6.7) to compute all the occurring generating functions. To give an example, in the case of the ladder of Figure 2 we would choose \( m = 3, x_1 = 3, x_2 = 7, x_3 = 13 \), and the resulting formula reads
\[
(6.10a)\quad \text{GF}(P_L(A \rightarrow E); z^{\text{NE}(j)}) = \sum_{j=4}^{13} \sum_{l=8}^{13} \sum_{k_1, k_2, k_3 \geq 0} z^{k_1+k_2+k_3} \binom{j-\alpha_1}{k_1} \binom{6-\alpha_2}{k_1}
\]
\[
(6.10b)\quad \left( \binom{l-1}{k_2-1} \binom{6}{k_2} - \binom{l-2}{k_2-1} \binom{7-j}{k_2} \right)
\]
\[
(6.10c)\quad \left( \binom{\eta_1-l-1}{k_3-1} \binom{\eta_2-12}{k_3} \right)^{j_1}
\]
where, in abuse of notation, the expression in (6.10b) is understood to be 1 if \( j = l \) and the expression in (6.10c) is understood to be 1 if \( l = \eta_1 \) (compare with (6.5) and (6.7)).

If \( L \) consists of not too many pieces, both methods are feasible methods, see our Example in section 3. Both methods yield \((2m - 1)\)-fold sums if the partition of the border consists of horizontal pieces throughout. However, the second method is preferable in case of long diagonal portions in the border of \( L \), since then Kul-Karni’s formula involves a lot more summations. Of course, in the worst case,
when \( L \) consists of 1-point pieces throughout, both methods are nothing else than plain counting, and therefore useless. For computation in case of such “fractal” boundaries it is more promising to avoid Theorem 1 and instead try to extend the dummy path method in [17] such that it also applies to the enumeration of nonintersecting lattice paths with respect to turns.

We remark that the above methods that we proposed for the computation of

\[ \text{GF}(\mathcal{P}_L(A \to E); z^{\text{NE}(\cdot)}) \]

also apply for computing the more general \( \text{GF}(\mathcal{P}_L(A \to E); w_{x,y}) \) from section 5. Then the binomials in (6.1), (6.4)–(6.7) would be replaced by elementary symmetric functions (cf. [20, p. 19]) in certain subsets of \( x \) and \( y \), respectively. For sake of brevity, we omit the details.

### 7. Conclusion

When examining our proof of Theorem 1, it can be realized that the validity of formula (2.2) is an almost incredible coincidence. There is no freedom for generalizations whatsoever. For example, in order to compute the Hilbert series of more general one-sided ladder determinantal rings than mentioned in this paper, it is necessary to solve the same enumeration problem as in Theorem 1, but with more general starting and end points, namely \( A_i = (0, \alpha_i) \) and \( E_i = (\eta_i, b) \) for some \( \alpha_i, \eta_i \) with \( \alpha_1 > \alpha_2 > \cdots > \alpha_n \) and \( \eta_1 < \eta_2 < \cdots < \eta_n \). One could write down the determinant in (2.2) in this case, too, and one could also give it a combinatorial meaning in terms of two-rowed arrays by using basically the same involution as in our proof of Theorem 1. But now the front and back part would not be uniquely determined and can therefore not be discarded. Hence, there is no lattice path meaning for the determinant in this more general case. Still, the enumeration problem with general starting and end points for one-sided ladders has been solved by the first author [15]. The solution cannot be formulated in terms of lattice paths any more, but in terms of two-rowed arrays only.

As we mentioned in section 3, more generally, also two-sided ladder determinantal rings are considered [1], [2], [5], [10]. In the associated enumeration problem one counts nonintersecting lattice paths in a two-sided ladder region, meaning that there is also a ladder-shaped lower boundary, but it is not a boundary for the paths but only for their NE-turns. We do not know how to generalize Theorem 1 to this situation.
References


Institut für Mathematik der Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria

E-mail address: kratt@pap.univie.ac.at
URL: http://www.radon.mat.univie.ac.at/people/kratt

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use