

# Dinv and Area

Adriano Garsia\*

Department of Mathematics  
UC San Diego  
California, U.S.A.  
garsiaadriano@gmail.com

Guoce Xin†

School of Mathematical Sciences  
Capital Normal University  
Beijing 100048, PR China  
guoce.xin@gmail.com

Submitted: Sep 29, 2016; Accepted: Mar 22, 2017; Published: Mar 31, 2017  
Mathematics Subject Classifications: 05A19, 05E99

## Abstract

We give a new combinatorial proof of the well known result that the  $\text{dinv}$  of an  $(m, n)$ -Dyck path is equal to the area of its sweep map image. The first proof of this remarkable identity for co-prime  $(m, n)$  is due to Loehr and Warrington. There is also a second proof (in the co-prime case) due to Gorsky and Mazin and a third proof due to Mazin.

**Keywords:** Rational Dyck paths, sweep map,  $\text{dinv}$

## 1 Introduction

Our main goal in this paper is to obtain a simpler proof that, under the sweep map, the  $\text{dinv}$  statistic of a rational Dyck path  $\overline{D}$  becomes the  $\text{area}$  statistic of its image path  $D$ . The first proof of this remarkable identity is due to Loehr and Warrington in [4]. There is also a second proof due to Gorsky and Mazin in [3], and a third proof due to Mazin [5]. See Section 3 for further explanation.

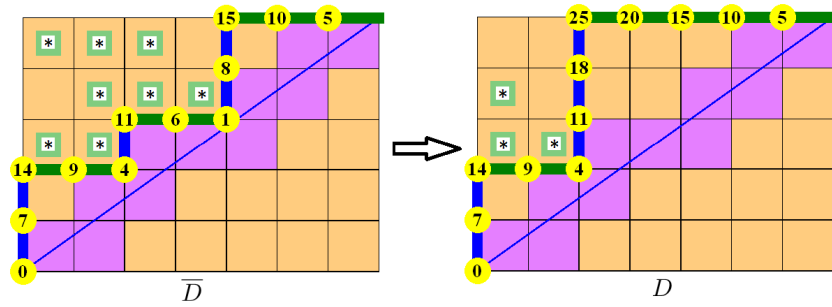
Inspired by a recent work of [6], we have come to depict rational Dyck paths in a manner which makes the ranks of the vertices of a path consistent with its visual representation. This very simple change turns out to be conducive to considerable simplifications in proving many of the properties of rational Dyck paths. For instance, we give a geometric proof of the invertibility of the rational Sweep Map in [2].

We shall always use  $(m, n)$  for a co-prime pair of positive integers, South end (by letter  $S$ ) for the starting point of a North step and West end (by letter  $W$ ) for the starting point of an East step. This is convenient and causes no confusion because we usually talk about the starting points of these steps.

---

\*Supported by NFS grant DMS13-62160.

†Partially supported by National Natural Science Foundation of China (11171231).



**Figure 1:** A  $(7,5)$ -Dyck path and its sweep map image.

In Figure 7 we have depicted a path  $\overline{D}$  in the  $7 \times 5$  lattice rectangle and its sweep map image  $D$  as they are traditionally depicted. The ranks of the vertices of an  $(m,n)$ -path are constructed by assigning 0 to the origin  $(0,0)$  and adding an  $m$  after a North step and subtracting an  $n$  after an East step.

To obtain the Sweep image  $D$  of  $\overline{D}$ , we let the main diagonal (with slope  $n/m$ ) sweep from right to left and successively draw the steps of  $D$  as follows: i) draw a South end (and hence a North step) when we sweep a South end of  $\overline{D}$ ; ii) draw a West end (hence an East step) when we sweep a West end of  $\overline{D}$ . The steps of  $D$  can also be obtained by rearranging the steps of  $\overline{D}$  by increasing ranks of their starting vertices.

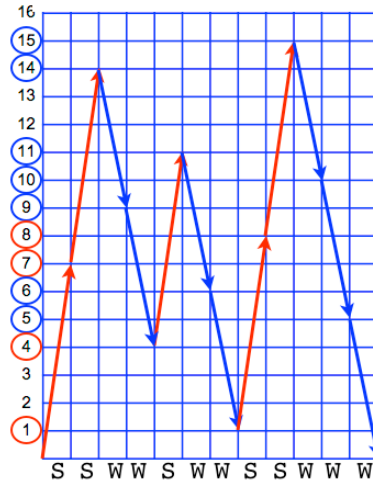
For  $(dm, dn)$ -rational Dyck paths, we compute the ranks of the starting points in the same way, but we may have ties for the starting ranks. When this happens, we sweep the right starting point first. Geometrically, we may simply sweep the starting points of the steps of  $\overline{D}$  from right to left using lines of slope  $n/m + \epsilon$  for sufficiently small  $\epsilon > 0$ , which will be written as  $0 < \epsilon \ll 1$ .

The **area** of a rational Dyck path is equal to the number of lattice cells between the path and the main diagonal. The **dinv** statistic we are using is the same as the  $h_{m/n}^-$  statistic in [1, Lemma 11]: A cell  $c$  above a  $(dm, dn)$ -Dyck path  $\overline{D}$  contributes a unit to its dinv statistic if and only if the starting rank  $a$  of the East step of  $\overline{D}$  below  $c$  and the starting rank  $b$  of the North step of  $\overline{D}$  to the right of  $c$  satisfy the inequality  $1 \leq (b + m) - (a - n) \leq m + n$ , which is equivalent to  $0 \leq a - b < m + n$ . In Figure 7, the cells contributing to  $\text{dinv}(\overline{D})$  are distinguished by a green square<sup>1</sup>. A quick count reveals that  $\text{dinv}(\overline{D}) = 8 = \text{area}(D)$ .

To proceed we need some notation. A  $(dm, dn)$  path diagram  $T$  consists of a list of  $dn$  red arrows and  $dm$  blue arrows, placed on a  $(dm + dn) \times dm n$  lattice rectangle. A red arrow is the up vector  $(1, m)$  and a blue arrow is the down vector  $(1, -n)$ <sup>2</sup>. The rows of lattice cells will be referred to as **rows** and the horizontal lattice lines will be simply referred to as **lines**. On the left of each line we have placed its  $y$  coordinate which we will refer to as its **level**. The level of the starting point of an arrow is called its **starting rank**, and similarly its **end rank** is the level of its end point. It will be convenient to call

<sup>1</sup>We also add a \* inside for black-white print.

<sup>2</sup>For black-white print, red arrows are up arrows and blue arrows are down arrows.



**Figure 2:** Illustration of the transformed  $\overline{D}$ .

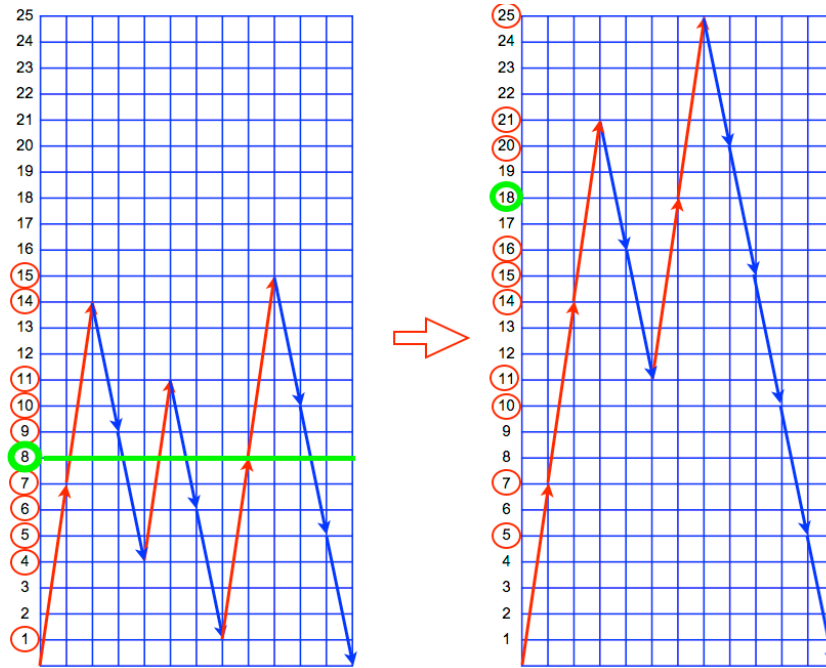
row  $i$  the row of lattice cells delimited by the lines at levels  $i$  and  $i + 1$ . Let  $\Sigma$  be a list consisting of  $dn$  letters  $S$  and  $dm$  letters  $W$ , and let  $R = (r_1, \dots, r_{dm+dn})$  be a sequence of  $dn + dm$  non-negative ranks. The path diagram  $T(\Sigma, R)$  (see Figure 2) is obtained by placing the letters of  $\Sigma$  at the bottom of the lattice columns and if the  $i^{\text{th}}$  letter of  $\Sigma$  is an  $S$  (resp.,  $W$ ) then we draw a red (resp., blue) arrow with starting rank  $r_i$  in the  $i^{\text{th}}$  column. Figure 2 depicts our manner of drawing the path  $\overline{D}$ . The ranks of  $\overline{D}$  are now the circled levels of the starting vertices.

Now the sweep order is from bottom to top and from right to left within each level. Geometrically, the Sweep lines are of slope  $\epsilon$  for  $0 < \epsilon \ll 1$ . Note that the co-prime case (i.e., when  $d = 1$ ) simplifies since no two starting points have the same rank, and thus the sweep lines are just the level lines.

Notice that each lattice cell may contain a segment of a red arrow or a segment of a blue arrow or no segment at all. The red segment count of row  $j$  will be denoted  $c^r(j)$  and the blue segment count is denoted  $c^b(j)$ . We will denote by  $c(j) = c^r(j) - c^b(j)$  and refer to it the  $j$ -th row count. Observe that in every row of a path diagram, the red segments and blue segments have to alternate. In particular, for Dyck paths as in Figure 2, every row must start with a red segment and end with a blue segment, and hence  $c(j) = 0$  holds for all  $j$ . This is called the zero-row-count property. It has the following immediate consequence.

**Proposition 1.** *The starting rank of any arrow  $A$  of  $D$  may be simply obtained by drawing in green (thick) the line of slope  $0 < \epsilon \ll 1$  at the starting point of its preimage  $\overline{A}$ , then counting the segments above the green line of any red arrow of  $\overline{D}$  that starts below the green line and adding to that count the number of segments below the green line of any blue arrow of  $\overline{D}$  that starts above the green line.*

*Proof.* The desired rank is  $bm - an$ , when  $\overline{D}$  has  $b$  red arrows and  $a$  blue arrows that start below the green line. We interpret this number as the segment count of these arrows



**Figure 3:** Here we see that the rank of the last red arrow of  $D$  is 18. We have altogether 4 red arrows and 2 blue arrows of  $\overline{D}$  that start below the green line, giving  $18 = 4 \times 7 - 2 \times 5$ . We count a total of 9 red segments above the green line and 9 blue segments below the green line. The zero-row-count property gives that these 9 blue segments are precisely the needed  $4 \times 7 - 2 \times 5$  minus 9.

and apply the zero-row-count property for rows below the green line. Indeed, the count of red segments above the green line is certainly needed. All the remaining portion of  $bm - an$  are segments below the green line. On the other hand, we need to add some blue segments to have all segments below the green line to apply the zero-row-count property. These blue segments are exactly as described in the proposition. See Figure 3 for an example.  $\square$

This proposition leads to the basic formula we will use to compute the area of the image  $D$  by working on the preimage  $\overline{D}$ .

**Theorem 2.** For any  $(dm, dn)$ -Dyck path  $D$ , with  $(m, n)$  co-prime, we have

$$\text{area}(D) = \frac{1}{n} \left( \sum_{j=1}^{dn} r(S_j(D)) \right) - d \frac{n-1}{2}, \quad (1)$$

where “ $r(S_j(D))$ ” denotes the rank of the  $j^{\text{th}}$  South end of  $D$ .

*Proof.* Notice that if we define the rank of the lattice cell whose South-West coordinates are  $(i, j)$  by setting  $r(i, j) = mj - ni$ , then the least positive rank in the  $j^{\text{th}}$  row is none other than the remainder of  $mj \bmod n$ . Since the residues modulo  $n$  are  $0, 1, \dots, n-1$

and the least positive ranks are distinct for  $0 \leq j \leq n-1$ , it follows that their sum is  $n(n-1)/2$ . Summing over all  $j$  gives  $dn(n-1)/2$ . Calling  $lpr_j$  the least nonnegative rank in row  $j$ , it is evident that  $(r(S_j(D)) - lpr_j)/n$  is the contribution of row  $j$  to the area of  $D$ . This given, we see that (1) is simply obtained by summing all these contributions.  $\square$

## 2 Proof that dinv sweeps to area

Given a co-prime pair  $(m, n)$  and  $d$ , our argument is to show that as we decrease the area of the preimage  $\overline{D}$  by one unit, both dinv and area satisfy the same recursion.

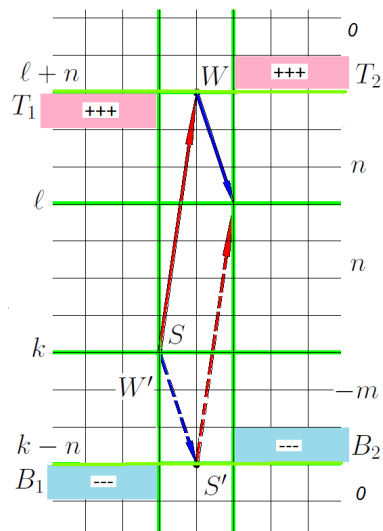
We have depicted in Figure 4 the cell that we have subtracted from the preimage  $\overline{D}$  to obtain a preimage  $\overline{D}'$  with one less unit of area. This operation may be viewed as replacing a red arrow  $S$  of  $\overline{D}'$  by a dashed red arrow  $S'$  and a blue arrow  $W$  by a dashed blue arrow  $W'$ . Calling  $D$  and  $D'$  the sweep map images of  $\overline{D}$  and  $\overline{D}'$ , our task is to determine the difference  $\text{area}(D) - \text{area}(D')$ . Our tool will be formula (1) and the fact that the starting rank of any arrow  $A$  of  $D$  is  $bm - an$  where  $\overline{D}$  has  $b$  red arrows and  $a$  blue arrows that start below the preimage  $\overline{A}$  of  $A$  in  $\overline{D}$ . This difference will be calculated in  $\overline{D}$  and  $\overline{D}'$  by means of our tool. It should be mentioned that the following argument will be significantly simplified if we choose the starting rank of the displayed  $W$  to be the largest in the sweep order. Then some of the cases can not happen. In particular, the regions  $T_1$  and  $T_2$  can not have any segments.

Below we will talk about four lines of levels  $l+n, l, k, k-n$  respectively. In the  $d \neq 1$  case, we actually mean the lines of slope  $\epsilon$  that passing through the corresponding four vertices of middle parallelogram. This is due to the modification of the Sweep lines in the non-coprime case.

Now there are 4 distinct cases. Firstly, red arrows that start above level  $l+n$  (see Figure 4) or below level  $k-n$  are not affected by the replacements. Thus their contribution to the difference is 0. Secondly in the case of any red arrow starting strictly below level  $l+n$  and strictly above level  $k$ , its contribution to the difference is  $n$ . The reason for this is that both red arrows in the display will increase the ranks of the arrows of  $D$  and  $D'$ , whose starting levels are in this range by an equal amount, therefore they cancel each other. By contrast the dashed blue arrow will affect the ranks of the red arrows of  $D'$  so that the contribution of each to the difference is  $-(-n)$ .

Thirdly, notice that each red arrow of  $D$  or  $D'$  that starts strictly below rank  $k$  and strictly above rank  $k-n$ , is not affected by either of the blue arrows. But each of the red arrows of  $D'$  is affected by the dashed red arrow and thus contributes a  $-m$  to the difference.

Finally we must include the contribution to the areas of  $D$  and  $D'$  by the ranks of red



**Figure 4:** Contribution for the area difference.

arrows in the display itself. We claim that

$$\begin{aligned} \text{rank}(S) - \text{rank}(S') \\ = m \times \#\{\text{red arrows that start (strictly) below level } k \text{ and above level } n - k\} \\ - n \times \#\{\text{blue arrows that start (strictly) below level } k \text{ and above level } n - k\}. \end{aligned} \quad (2)$$

The reason is that the arrows that start below level  $k - n$  contribute equally to the areas of  $S$  and  $S'$ . Thus they cancel in computing the desired difference. On the other hand all the arrows accounted for in (2) do contribute to the area of  $S$  but not to the area of  $S'$ .

Notice first that the contribution to the area difference in the third case is the negative of the first part of the contribution obtained in (2). After cancellation, all the remaining contributions are multiples of  $n$ . (A convenient fact since, according to (1),  $n$  has to be divided out.)

Furthermore, in the second case this multiple counts the number of red arrows that have a red segment in  $T_1$  or  $T_2$ . Finally, we see in (2) that the factor of  $n$  that survives the cancellation, counts exactly the arrows that have a blue segment in  $B_1$  or  $B_2$ .

These observations imply the following result.

**Proposition 3.** *Let  $\overline{D'}$  be obtained from  $\overline{D}$  by removing an area cell and let  $D$  and  $D'$  be their sweep map images. Let  $B_1$  and  $B_2$  be the blue regions and  $T_1$  and  $T_2$  be the red regions<sup>3</sup> in Figure 4. Then*

$$\text{area}(D) - \text{area}(D') = c^r(T_1) + c^r(T_2) - c^b(B_1) - c^b(B_2), \quad (3)$$

where  $c^r(T_1), c^r(T_2)$  denote red segment counts and  $c^b(B_1), c^b(B_2)$  denote blue segment counts in the corresponding regions.

To obtain the recursion satisfied by  $\text{dinv}$  we will make use of the following remarkable fact. By abuse of notation, we will use  $W_i$  (resp.  $S_j$ ) for the  $i$ -th blue ( $j$ -th red) arrow.

**Proposition 4.** *The  $\text{dinv}$  statistic of a rational Dyck path  $D$  (given in our stretched form) may simply be obtained by counting the pairs  $(W_i \rightarrow S_j)$  consisting of a blue arrow to the left of a red arrow, such that  $W_i$  sweeps  $S_j$ , i.e.,  $W_i$  intersects  $S_j$  when we move it along a line of slope  $\epsilon$  (with  $0 < \epsilon \ll 1$ ) to the right past  $S_j$ .*

*Proof.* Suppose that  $W_i$  has starting rank  $a$  and  $S_j$  has starting rank  $b$ . By the quoted result of Loehr-Warrington, the pair  $(W_i, S_j)$  contribute a unit to  $\text{dinv}(D)$  if and only if  $0 \leq a - b < m + n$ . This is equivalent to requiring that  $a \geq b$  and  $a - n < b + m$ . But these two inequalities are precisely what is needed to guarantee that  $W_i$  sweeps  $S_j$ .  $\square$

This given, our  $\text{dinv}$  recursion can be stated as follows.

---

<sup>3</sup>For black-white print, we add “+ + +” to red regions and “- - -” to blue regions.

**Proposition 5.** Let  $\overline{D}'$  be obtained from  $\overline{D}$  by removing an area cell. Let  $B_1$  and  $T_1$  be the blue regions and  $B_2$  and  $T_2$  be the red regions in Figure 5. Then

$$\mathbf{dinv}(\overline{D}) - \mathbf{dinv}(\overline{D}') = c^b(T_1) + c^r(T_2) - c^b(B_1) - c^r(B_2) - 1, \quad (4)$$

where  $c^b(T_1), c^b(B_1)$  denote blue segment counts and  $c^r(T_2), c^r(B_2)$  denote red segment counts in the corresponding regions.

*Proof.* We will use the visual fact  $\mathbf{dinv}(\overline{D}) = \#\{(W_i \rightarrow S_j) : W_i \text{ sweeps } S_j\}$ .

Since  $\overline{D}'$  is obtained from  $\overline{D}$  by replacing the solid arrows  $S, W$  by dashed arrows  $S', W'$  (see Figure 5), we can divide the contribution of a pair  $(W_i \rightarrow S_j)$  to the difference  $\mathbf{dinv}(\overline{D}) - \mathbf{dinv}(\overline{D}')$  into four cases.

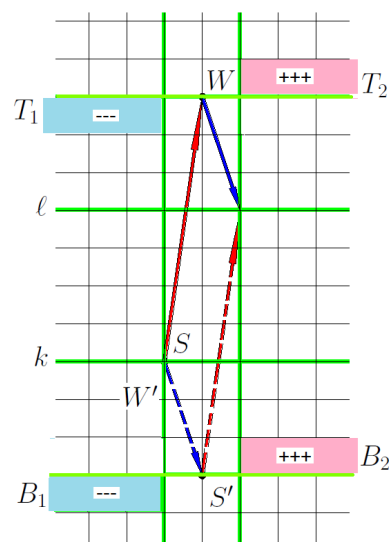
(1) Both  $W_i$  and  $S_j$  are not the displayed arrows. The contribution in this case is always 0.

(2) Both  $W_i$  and  $S_j$  are in the displayed arrows. This can only happen in two ways:  $(W, S)$  in  $\overline{D}$  (no  $\mathbf{dinv}$ ) or  $(W', S')$  in  $\overline{D}'$  (1  $\mathbf{dinv}$ ). Therefore the contribution to the difference in this case is  $-1$ .

(3) Only  $W_i$  is one of the displayed arrows. Then we need to consider the pairs  $(W, S_j)$  in  $\overline{D}$  and  $(W', S_j)$  in  $\overline{D}'$ . Their contribution to the difference is 1 if  $S_j$  has a red segment in  $T_2$ ,  $-1$  if  $S_j$  has a red segment in  $B_2$  and 0 if  $S_j$  does not have a segment in neither  $T_2$  or  $B_2$ .

(4) Only  $S_j$  is in the displayed arrows. Then we need to consider the pairs  $(W_i, S)$  in  $\overline{D}$  and  $(W_i, S')$  in  $\overline{D}'$ . Their contribution to the difference is 1 if  $W_i$  has a blue segment in  $T_1$ ,  $-1$  if  $W_i$  has a blue segment in  $B_1$  and 0 if  $W_i$  does not have a segment in neither  $T_1$  or  $B_1$ .

This proves the identity in (4).  $\square$



**Figure 5:** Contribution for the  $\mathbf{dinv}$  difference.

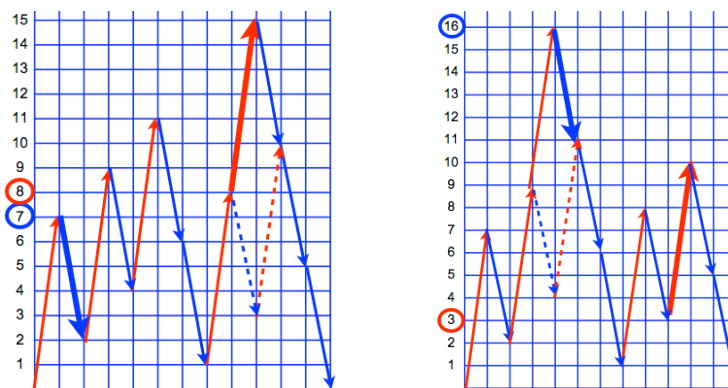
*Proof that  $\mathbf{dinv}$  sweeps to area.* We first show that the recursions in (3) and (4) are identical. To do this it is sufficient to observe that  $c^r(T_1) = c^b(T_1)$  and  $c^r(B_2) + 1 = c^b(B_2)$ . The reason for this is the alternating colors of segments in each row that always begin with a red segment and end with a blue segment.

Thus it is sufficient to verify the base case where  $\mathbf{area}(\overline{D}) = 0$ . The only  $(dm, dn)$ -Dyck path with area 0 is the path  $\overline{D}$  that remains as close as possible to the main diagonal. Thus the ranks of the South ends of such a path are a rearrangement of  $0, 1, 2, \dots, n-1$ , with each appearing  $d$  times. This forces the image  $D$  of  $\overline{D}$  to start with  $dn$  North steps and end with  $dm$  East steps. This is the path of maximum area. It remains to prove that  $\overline{D}$  has maximum  $\mathbf{dinv}$ , or equivalently every cell above  $\overline{D}$  contributes to its  $\mathbf{dinv}$ . By contradiction, suppose that for a pair  $(W_i \rightarrow S_j)$ , the  $i$ -th blue arrow  $W_i$  does not sweep the  $j$ -th red arrow  $S_j$ . We have the following two cases. See Figure 6.

(1)  $S_j$  starts at a level above  $W_i$ . Assume we have  $S_j S_{j+1} \cdots S_{j+r} W$  for some  $r \geq 0$ . Then consider the path obtained from  $\overline{D}$  by changing this to  $S_j S_{j+1} \cdots W S_{j+r}$ . This is a new Dyck path with area one less than that of  $\overline{D}$ . A contradiction!

(2)  $S_j$  ends below the level of  $W_i$ , then assume we have  $SW_{i-r} \cdots W_{i-1} W_i$  for some  $r \geq 0$ . Then consider the path obtained from  $\overline{D}$  by changing this to  $W_{i-r} S \cdots W_{i-1} W_i$ . This is a new Dyck path with area one less than that of  $\overline{D}$ . A contradiction!

This completes our proof that  $\text{dinv}$  sweeps to area.  $\square$



**Figure 6:** Two examples of our proof that the  $\text{dinv}$  of the area zero  $(m, n)$ -Dyck path is  $\frac{(dm-1)(dn-1)+d-1}{2}$ , where the blue arrow  $W_i$  and red arrow  $S_j$  are thickened.

### 3 Remarks

We terminate our presentation by a few comments. To begin we should note that our argument does not use Proposition 1. We have nevertheless included it in this writing for two reasons. Firstly because it is too surprising a result to leave out, but more importantly, because it gives a simple proof of the nontrivial result that the sweep image of a Dyck path is also a Dyck path. The reason for this is that it is implicit in the conclusion of the Proposition that the starting ranks of all the arrows of the image are non-negative. The latter is the only property needed to guarantee that the image of a  $(dm, dn)$ -Dyck path is a  $(dm, dn)$ -Dyck path.

There are three proof of the “ $\text{dinv}$  sweeps to area” result as we said in Section 1. The first two proof only deal with the co-prime case. This case simplifies a lot since all starting ranks of the steps are distinct. Gregory Warrington told us their proof in [1] can be extended for the non-coprime case. Our first draft of this paper also deal with the co-prime case, but explained how to (naturally) extend our approach to the non-coprime case. After we put this draft on the arXiv, Mazin told us immediately that the non-coprime case is [5, Corollary 1] after some translation of terminology. This write up is modified (suggested by the referee) to fit the non-coprime case.

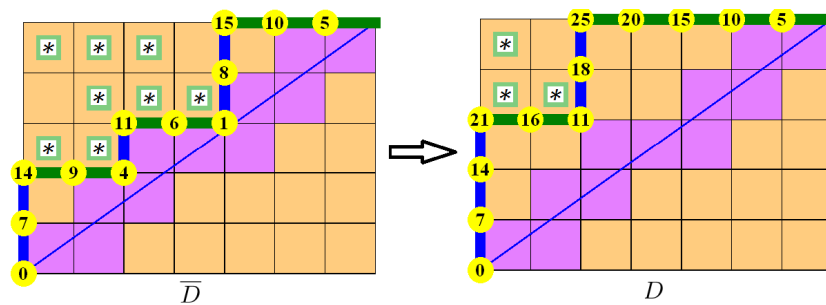


## References

- [1] Drew Armstrong, Nicholas A. Loehr, and Gregory S. Warrington, Rational parking functions and Catalan numbers, *Annals Combin.*, 20(1):21–58, 2016.
- [2] Adriano M. Garsia and Guoce Xin, Inverting the rational sweep map, *J. Comb.*, to appear, [arXiv:1602.02346](#).
- [3] E. Gorsky and M. Mazin, Compactified Jacobians and  $q, t$ -Catalan Numbers I, *J. Combin. Theory Ser. A*, 120:49–63, 2013.
- [4] Nicholas A. Loehr and Gregory S. Warrington, A continuous family of partition statistics equidistributed with length, *J. Combin. Theory Ser. A*, 116:379–403, 2009.
- [5] Mikhail Mazin, A bijective proof of Loehr-Warrington’s formulas for the statistics  $\text{ctot}_p^q$  and  $\text{mid}_p^q$ , *Annals Combin.*, 18:709–722, 2014.
- [6] H. Thomas and N. Williams, Sweeping up zeta, [arXiv:1512.01483](#).

# Corrigendum – added 9th June 2017

1. Page 2 Figure 1: it should be replaced by the following figure, where the right picture was not correctly drawn in the published version.



**Figure 7:** A  $(7, 5)$ -Dyck path and its sweep map image.

2. Page 6 equation (2): the  $n - k$  (appear twice) should be replaced by  $k - n$ .