

A bijection between bargraphs and Dyck paths

Emeric Deutsch*

Sergi Elizalde†

Abstract

Bargraphs are a special class of convex polyominoes. They can be identified with lattice paths with unit steps north, east, and south that start at the origin, end on the x -axis, and stay strictly above the x -axis everywhere except at the endpoints. Bargraphs, which are used to represent histograms and to model polymers in statistical physics, have been enumerated in the literature by semiperimeter and by several other statistics, using different methods such as the wasp-waist decomposition of Bousquet-Mélou and Rechnitzer, and a bijection with certain Motzkin paths.

In this paper we describe an unusual bijection between bargraphs and Dyck paths, and study how some statistics are mapped by the bijection. As a consequence, we obtain a new interpretation of Catalan numbers, as counting bargraphs where the semiperimeter minus the number of peaks is fixed.

Keywords: bargraph, bijection, Dyck path, Catalan number

1 Introduction

A bargraph is a lattice path with steps $U = (0, 1)$, $H = (1, 0)$ and $D = (0, -1)$ that starts at the origin and ends on the x -axis, stays strictly above the x -axis everywhere except at the endpoints, and has no pair of consecutive steps of the form UD or DU . Sometimes it is convenient to identify a bargraph with the corresponding word on the alphabet $\{U, D, H\}$, and sometimes with the sequence of heights (i.e., y -coordinates) of the H steps of the path, thus interpreting the bargraph as a composition.

Bargraphs appear in the literature with different names. As a special class of convex polyominoes, they have been studied and enumerated by Bousquet-Mélou and Rechnitzer [6], Prellberg and Brak [10], and Feretić [8]. The distribution of several bargraph statistics has been given in a series of papers by Blecher, Brennan, Knopfmacher and Prodinger [1, 2, 3, 4, 5]. All of the above papers rely on a decomposition of bargraphs that is used to obtain equations satisfied by the corresponding generating functions. Recently, the enumeration of bargraphs with respect to some additional statistics has been obtained in [7] by using a simple bijection between bargraphs and Motzkin paths without peaks or valleys, together with the recursive structure of Motzkin paths.

The *semiperimeter* of a bargraph B is defined as the number of U steps plus the number of H steps, and denoted by $\text{sp}(B)$. It coincides with the semiperimeter of the closed polyomino determined by the bargraph and the x -axis. Let \mathcal{B} denote the set of bargraphs, and let \mathcal{B}_n denote

*NYU Tandon School of Engineering, Brooklyn, NY 11201, USA.

†Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA. sergi.elizalde@dartmouth.edu. Partially supported by grant #280575 from the Simons Foundation.

those of semiperimeter n . An important bargraph statistic in this paper is the *number of peaks*, which are occurrences of $UH^\ell D$ for some $\ell \geq 1$.

A Dyck path is a lattice path with steps $u = (1, 1)$ and $d = (1, -1)$ that starts at the origin, ends on the x -axis, and stays weakly above the x -axis. Let \mathcal{D} denote the set of Dyck paths, and let \mathcal{D}_n denote those of semilength n , where the semilength of a path P is defined as half of the number of steps, and denoted by $\text{sl}(P)$. It is well known that $|\mathcal{D}_m| = C_m = \frac{1}{m+1} \binom{2m}{m}$, the m th Catalan number.

In Section 2 we introduce a bijection between bargraphs and Dyck paths. An unusual feature of our bijection is that it does not preserve “size” under the natural definitions, such as semiperimeter or semilength. Instead, the statistics preserved by the bijection are more subtle, as discussed in Section 3. As a consequence of the bijection, we obtain a new interpretation of the Catalan numbers in terms of bargraphs. An alternative proof of this result, using generating functions, is given in Section 4.

2 The bijection

In this section we define a bijection ϕ from Dyck paths to bargraphs, with the property that the semilength of a Dyck path becomes the semiperimeter minus the number of peaks of the corresponding bargraph.

Given a Dyck path P , define the height of each step of P as the y -coordinate of its highest point. Consider the sequence of heights of the steps of P . To construct the bargraph $\phi(P)$ from this sequence, turn each maximal block of c consecutive letters h into $\lfloor \frac{c}{2} \rfloor$ columns of height h .

For example, the Dyck path in Figure 1 has height sequence **1111123333222233332221111221**, and thus its corresponding bargraph has column heights **113322332112**.

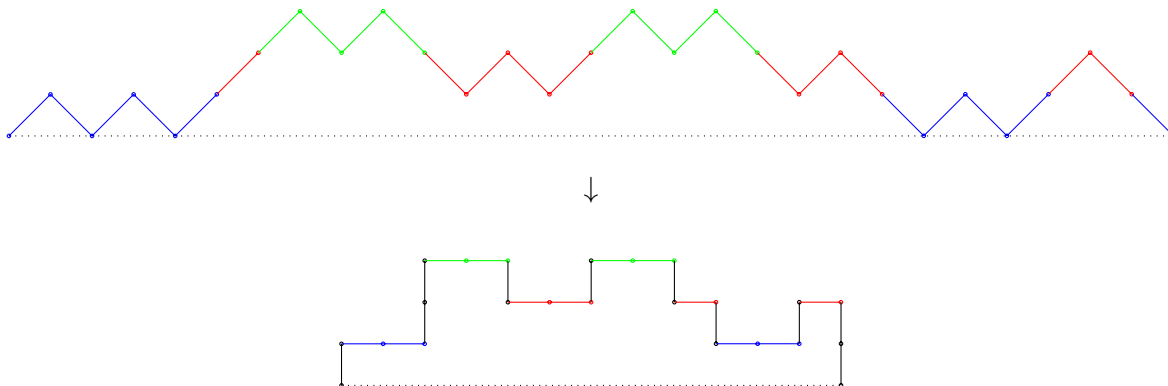


Figure 1: A Dyck path and its corresponding bargraph.

Next we describe the inverse. Consider a bargraph described by a sequence of column heights $h_1^{a_1} h_2^{a_2} \dots h_r^{a_r}$. By inserting terms with $a_j = 0$ if necessary, we can assume that $h_1 = h_r = 1$ and $|h_{i+1} - h_i| = 1$ for all i . Define $h_0 = h_{r+1} = 0$ by convention. To construct a Dyck path, for

$i = 1, 2, \dots, r$, we replace $h_i^{a_i}$ with

$$\begin{cases} (ud)^{a_i} & \text{if } h_{i-1} < h_i > h_{i+1}, \\ (ud)^{a_i}u & \text{if } h_{i-1} < h_i < h_{i+1}, \\ (du)^{a_i}d & \text{if } h_{i-1} > h_i > h_{i+1}, \\ (du)^{a_i} & \text{if } h_{i-1} > h_i < h_{i+1}. \end{cases}$$

For example, the bargraph with column heights 113322332112 is first written as $1^22^03^22^23^22^11^22^11^0$, which then becomes $(ud)^2uu(ud)^2(du)^2(ud)^2(du)d(du)^2(ud)d$.

3 Properties of ϕ

Next we show how some statistics on Dyck paths are mapped by the bijection ϕ into statistics on bargraphs. We start by defining some statistics on bargraphs and Dyck paths.

For a bargraph B , denote by $\#H(B)$, $\#U(B)$ and $\#D(B)$ the number of H , U and D steps in B , respectively. Note that, by definition, $\text{sp}(B) = \#H(B) + \#U(B)$. The *area* of the region under a bargraph B and above the x -axis is denoted by $\text{area}(B)$.

A *peak* (resp. *valley*) in a bargraph is an occurrence of $UH^\ell D$ (resp. $DH^\ell U$) for some $\ell \geq 1$. Denote by $\text{pk}(B)$ the number of peaks of B , and by $\text{val}(B)$ its number of valleys. Note that $\text{pk}(B) = \text{val}(B) + 1$, since peaks and valleys in a bargraph alternate, always starting and ending with a peak. Let $\text{shv}(B)$ be the sum of the heights of the valleys of B , where the height of a valley is the y -coordinate of its H steps. Denote by $\#H_1(B)$ the number of H steps of B at height 1 (equivalently, number of columns of B of height 1).

A *peak* in a Dyck path is an occurrence of ud , and a *return* is a d step that ends on the x -axis. For $P \in \mathcal{D}$, let $\Lambda(P)$ denote its number of peaks, and $\text{ret}(P)$ its number of returns. Defining the height of a peak as the y -coordinate of its highest point, let $\text{shp}(P)$ denote the sum of the heights of the peaks of P .

Define the height of a Dyck path or a bargraph to be the maximum y -coordinate that it reaches. Denote by $\text{ini}_u(P)$ (resp. $\text{fin}_d(P)$) the number of initial u steps (resp. final d steps) in a Dyck path P , and define $\text{ini}_U(B)$ and $\text{fin}_D(B)$ on a bargraph B similarly.

Table 1 summarizes the notation that we use for statistics on bargraphs and Dyck paths.

Statistics on bargraphs		Statistics on Dyck paths	
sp	semiperimeter	sl	semilength
pk	number of peaks	Λ	number of peaks
val	number of valleys	ret	number of returns
shv	sum of heights of valleys	shp	sum of heights of peaks
ini_U	number of initial U steps	ini_u	number of initial u steps
fin_D	number of final D steps	fin_d	number of final d steps
$\#H, \#U, \#D$	number of H, U, D steps	eb	number of even blocks in height sequence
$\#H_1$	number of H steps at height 1	ob	number of odd blocks in height sequence

Table 1: Notation for statistics on bargraphs and Dyck paths

Example 3.1. The bargraph B in Figure 1 has $\text{sp}(B) = 17$, $\#H(B) = 12$, $\#U(B) = \#D(B) = 5$, $\text{area}(B) = 24$, $\text{pk}(B) = 3$, $\text{shv}(B) = 3$, $\#H_1(B) = 4$, $\text{height}(B) = 3$, $\text{ini}_U(B) = 1$, $\text{fin}_D(B) = 2$.

Theorem 3.2. *Let $P \in \mathcal{D}$ and let $B = \phi(P) \in \mathcal{B}$. Then*

- (a) $\text{sl}(P) = \text{sp}(B) - \text{pk}(B)$,
- (b) $\Lambda(P) = \#H(B) - \text{val}(B)$,
- (c) $\text{shp}(P) = \text{area}(B) - \text{shv}(B)$,
- (d) $\text{height}(P) = \text{height}(B)$,
- (e) $\text{ini}_u(P) = \text{ini}_U(B)$,
- (f) $\text{fin}_d(P) = \text{fin}_D(B)$.
- (g) $\text{ret}(P) = \#H_1(B) + 1$ (unless P and B have height 1, in which case $\text{ret}(P) = \#H_1(B)$).

Proof. Let us first prove part (a). The height sequence of P can be decomposed into maximal blocks of steps of the same height. Let $\text{eb}(P)$ be the number of even blocks, and let $\text{ob}(P)$ be the number of odd blocks. In the construction of $B = \phi(P)$, each block of size c produces $\lfloor \frac{c}{2} \rfloor$ horizontal steps in the bargraph. Thus, the number of H steps in B equals $\#H(B) = \text{sl}(P) - \frac{\text{ob}(P)}{2}$.

Additionally, the height changes by one from each block to the next, which translates into a U or a D in the bargraph. Thus $\#U(B) + \#D(B) = \text{eb}(P) + \text{ob}(P) + 1$. More specifically, the two consecutive height changes occurring at the beginning and at the end of a given block (with the convention that the height is 0 outside the path) have the same sign if and only if the block has odd length.

For example, in the path of Figure 1, denoting by \uparrow and \downarrow height changes of $+1$ and -1 respectively, we get $\uparrow 11111 \uparrow 2 \uparrow 3333 \downarrow 2222 \uparrow 3333 \downarrow 222 \downarrow 1111 \uparrow 22 \downarrow 1 \downarrow$. In particular, the number of even blocks equals the number of peaks plus the number of valleys (which in turn equals the number of peaks minus one), and so $\text{eb}(P) = 2 \text{pk}(B) - 1$.

Putting the above observations together,

$$\begin{aligned} \text{sp}(B) &= \#H(B) + \frac{\#U(B) + \#D(B)}{2} = \text{sl}(P) - \frac{\text{ob}(P)}{2} + \frac{\text{eb}(P) + \text{ob}(P) + 1}{2} \\ &= \text{sl}(P) + \frac{\text{eb}(P) + 1}{2} = \text{sl}(P) + \text{pk}(B), \end{aligned}$$

as desired.

To prove parts (b) and (c), note that in the above decomposition of P into maximal blocks of steps of the same height, all peaks of P occur within the blocks, since the two steps in a peak UD have the same height. Additionally, the height of such a peak coincides with the height of the block containing it, and thus with the height of the corresponding columns in the bargraph. Each odd block in P of size $2i + 1$ contains i peaks. Similarly, each even block of size $2i$ contains i peaks, unless the two blocks surrounding it are higher than the block itself, in which case it contains $i - 1$ peaks. Blocks surrounded by higher blocks correspond to valleys of B . It follows that $\Lambda(P) = \text{sl}(P) - \frac{\text{ob}(P)}{2} - \text{val}(B) = \#H(B) - \text{val}(B)$. Additionally, since the area of a bargraph is the sum of the heights of its columns, we get $\text{shp}(P) = \text{area}(B) - \text{shv}(B)$.

Parts (d), (e) and (f) follow immediately from the construction of ϕ .

Finally, to prove part (g), note that all returns of P occur in blocks of height 1. If P (equivalently, B , by part (c)) has height 1, then it is clear that $\text{ret}(P) = \text{sl}(P) = \#H_1(B)$. If P has height at least 2, then the blocks of height 1 are the first and last block, which are necessarily odd, plus possibly some even blocks. If the first block has length $2i + 1$ (for some i), it contributes i returns to P and i H steps at height 1 to B . If the last block has length $2i + 1$, it contributes $i + 1$ returns to P and i H steps at height 1 to B . Any even block at height 1 of length $2i$ contributes i returns to P and i H steps at height 1 to B . The equality $\text{ret}(P) = \#H_1(B) + 1$ follows. \square

The following is an immediate consequence of Theorem 3.2(a), giving a new interpretation of the Catalan numbers.

Corollary 3.3.

$$|\{B \in \mathcal{B} : \text{sp}(B) - \text{pk}(B) = m\}| = C_m.$$

Example 3.4. For $m = 1, 2, 3$, the set $\{B \in \mathcal{B} : \text{sp}(B) - \text{pk}(B) = m\}$ consists of all bargraphs of semiperimeter 2, 3, 4, respectively, since each of these bargraphs has 1 peak. For $m = 4$, this set, drawn in Figure 2, consists of all 13 bargraphs of semiperimeter 5 (each of them having 1 peak) and the only bargraph of semiperimeter 6 that has 2 peaks.

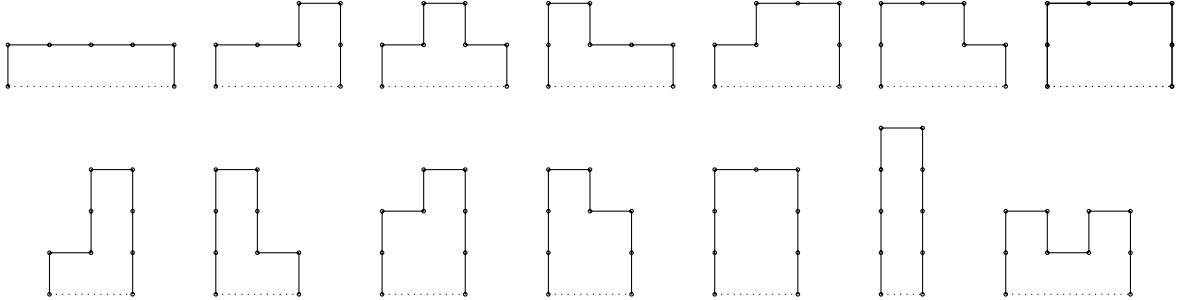


Figure 2: The 14 bargraphs with $\text{sp}(B) - \text{pk}(B) = 4$.

Table 2 displays the number of bargraphs with a given semiperimeter (up to 12) and a given number of peaks. Corollary 3.3 states that the diagonal sums of this table (which is the triangular sequence [9, A271940]) are the Catalan numbers. This observation for small values of n led to the results in this paper.

We end this section by describing other properties of ϕ , which can be used to give a recursive description of this map. Interpreting a bargraph B as a lattice path, the path UBD is the bargraph whose height sequence is obtained by adding one to each entry of the height sequence of B . On the other hand, given bargraphs B and B' , the expression $B \circ B'$ denotes the concatenation of B and B' viewed as height sequences. Let $\mathbf{1}$ denote the bargraph consisting of one column of height 1.

Proposition 3.5. *Let $P, P' \in \mathcal{D}$, and let $B = \phi(P)$, $B' = \phi(P')$. Then*

- (i) $\phi(uPd) = UBD$,
- (ii) $\phi(PP') = \begin{cases} B \circ \mathbf{1} \circ B' & \text{if } \text{height}(P) \geq 2 \text{ and } \text{height}(P') \geq 2, \\ B \circ B' & \text{otherwise.} \end{cases}$

$n \setminus k$	1	2	3	4
2	1			
3	2			
4	5			
5	13			
6	34	1		
7	89	8		
8	233	42		
9	610	183	1	
10	1597	717	13	
11	4181	2622	102	
12	10946	9134	624	1

Table 2: Number of bargraphs with semiperimeter n having k peaks. The diagonal sums of this table, which count bargraphs where $n - k$ is fixed, are the Catalan numbers $1, 2, 5, 14, 42, 132, 429, \dots$

Proof. To prove part (i), note that elevating the path P increases the entries of its height sequence by one, and adds a 1 at the beginning and a 1 at the end of it, neither of which contributes new columns to the bargraph.

Part (ii) follows from the observation that, if $P \in \mathcal{D}$ has $\text{height}(P) \geq 2$, then the initial and final maximal blocks of 1's in the height sequence of P have odd sizes. \square

4 Generating functions

In this subsection we give an alternative proof of Corollary 3.3 using generating functions. Let

$$G(t, z) = \sum_{B \in \mathcal{B}} t^{\text{pk}(B)} z^{\text{sp}(B)}$$

be the generating function for bargraphs according to the number of peaks and the semiperimeter. It was shown in [2] (and can be easily derived from [7]) that G satisfies

$$z(1-z)G(t, z)^2 - (1-3z+z^2+tz^3)G(t, z) + tz^2(1-z) = 0. \quad (1)$$

Let

$$H(t, z) = G(1/t, tz) = \sum_{B \in \mathcal{B}} t^{\text{sp}(B) - \text{pk}(B)} z^{\text{sp}(B)}$$

be the generating function where t keeps track of the semiperimeter minus the number of peaks. Then, substituting into (1), we see that H satisfies

$$tz(1-tz)H(t, z)^2 - (1-3tz+t^2z^2+t^2z^3)H(t, z) + tz^2(1-tz) = 0. \quad (2)$$

Setting $z = 1$ and simplifying, we get

$$tH(t, 1)^2 - (1-2t)H(t, 1) + t = 0,$$

yielding

$$H(t, 1) = \sum_{B \in \mathcal{B}} t^{\text{sp}(B) - \text{pk}(B)} = \frac{1 - \sqrt{1-4t}}{2t} - 1 = \sum_{m \geq 1} C_m t^m.$$

References

- [1] A. Blecher, C. Brennan and A. Knopfmacher, Levels in bargraphs, *Ars Math. Contemp.* 9 (2015), 287–300.
- [2] A. Blecher, C. Brennan and A. Knopfmacher, Peaks in bargraphs, *Transactions of the Royal Society of South Africa* 71 (2016), 97–103.
- [3] A. Blecher, C. Brennan and A. Knopfmacher, Combinatorial parameters in bargraphs, *Quaest. Math.* 39 (2016), 619–635.
- [4] A. Blecher, C. Brennan and A. Knopfmacher, Walls in bargraphs, preprint.
- [5] A. Blecher, C. Brennan, A. Knopfmacher and H. Prodinger, The height and width of bargraphs, *Discrete Appl. Math.* 180 (2015), 36–44.
- [6] M. Bousquet-Mélou and A. Rechnitzer, The site-perimeter of bargraphs, *Adv. in Appl. Math.* 31 (2003), 86–112.
- [7] E. Deutsch and S. Elizalde, Statistics on bargraphs viewed as cornerless Motzkin paths, *Discrete Appl. Math.* 221 (2017), 54–66.
- [8] S. Feretić, A perimeter enumeration of column-convex polyominoes, *Discrete Math. Theor. Comput. Sci.* 9 (2007), 57–84.
- [9] The On-Line Encyclopedia of Integer Sequences, published electronically at <https://oeis.org>, 2016.
- [10] T. Prellberg and R. Brak, Critical exponents from nonlinear functional equations for partially directed cluster models, *J. Stat. Phys.* 78 (1995), 701–730.