Combinatorial enumeration of lattice paths by flaws with respect to a linear boundary of rational slope

Federico Firoozi, Jonathan Jedwab, Amarpreet Rattan
Department of Mathematics, Simon Fraser University, Burnaby BC, Canada

29 May 2024

Abstract

Let a, b be fixed positive coprime integers. For a positive integer g, write $N_k(g)$ for the set of lattice paths from the startpoint $(0,0)$ to the endpoint (ga, gb) with steps restricted to $\{(1,0),(0,1)\}$, having exactly k flaws (lattice points lying above the linear boundary). We wish to determine $|N_k(g)|$. The enumeration of lattice paths with respect to a linear boundary while accounting for flaws has a long and rich history, dating back to the 1949 results of Chung and Feller. The only previously known values of $|N_k(g)|$ are the extremal cases $k = 0$ and $k = g(a+b) - 1$, determined by Bizley in 1954. Our main result is that a certain subset of $N_k(g)$ is in bijection with $N_{k+1}(g)$. One consequence is that the value $|N_k(g)|$ is constant over each successive set of $a + b$ values of k. This in turn allows us to derive a recursion for $|N_k(g)|$ whose base case is given by Bizley's result for $k = 0$. We solve this recursion to obtain a closed form expression for $|N_k(g)|$ for all k and g. Our methods are purely combinatorial.

1 Introduction

The lattice path shown in Figure 1.1 contains exactly five lattice points that lie above the linear boundary joining the startpoint $(0,0)$ to the endpoint $(8,6)$.

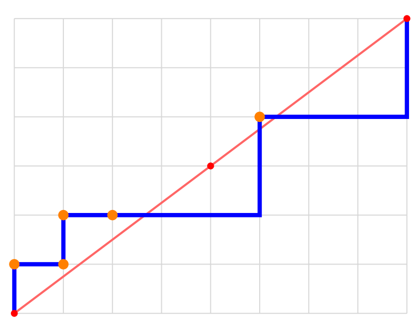


Figure 1.1: Lattice path from $(0,0)$ to $(8,6)$ with five lattice points above the line connecting $(0,0)$ to $(8,6)$.

Throughout, a, b are fixed positive coprime integers and g is a positive integer. Our objective is to count the number of lattice paths from the startpoint $(0,0)$ to the endpoint (ga, gb) with steps restricted to $\{(1,0),(0,1)\}$, having exactly k lattice points lying above the linear boundary joining the startpoint to the endpoint.

Let p be a path. The *boundary* of p is the line joining its startpoint to its endpoint. The path p *contains* the lattice point $(x+i, y+j)$ (equivalently, $(x+i, y+j)$ is *a point of* p) if p starts at (x, y), and the first $i+j \geq 0$ steps of p consist of i of the $(1,0)$ steps and j of the $(0,1)$ steps (in any order). We consider the points of p to be ordered according to increasing values of $i+j$. A point of p is a *flaw* if it lies strictly above the boundary of p. For example, the path in Figure 1.1 has the five flaws $(0,1), (1,1), (1,2), (2,2), (5,4)$ denoted in orange.

Definition 1.1 (Sets $N(g)$ and $N_k(g)$)**.** Let $N(g)$ be the set of all paths from $(0,0)$ to (ga, gb), and let $N_k(g)$ be the subset of such paths having exactly k flaws. ⌟

The possible values for the number k of flaws of a path are those satisfying $0 \leq k < g(a+b)$. The central objective of this paper is to find an explicit formula for $|N_k(g)|$ for all g, k satisfying $0 \leq k < g(a+b)$.

The extremal values $|N_0(g)|$ and $|N_{g(a+b)-1}(g)|$ were found by Bizley [2] in 1954 (see Theorem 1.10 below). Until now, the value of $|N_k(g)|$ was unknown for all other k.

In general, the values $|N_0(g)|$ and $|N_{g(a+b)-1}(g)|$ are not equal because a path can contain lattice points on the boundary other than its startpoint and endpoint. Points on the boundary are not counted as flaws, and so rotation of the path through 180° does not map the set $N_k(g)$ to the set $N_{g(a+b)-1-k}(g)$. We are therefore unable to use standard rotational symmetry arguments that are commonly applied to other path enumeration problems.

Table 1.1 displays the numerical value of $|N_k(4)|$ for $(a,b) = (3,2)$, obtained by computer. We note two apparent properties suggested by these values:

k	$\lvert N_k(4)\rvert$	$\lvert N_k(4)\rvert - \lvert N_{k+1}(4)\rvert$
0	7229	0
1	7229	0
2	7229	0
3	7229	0
4	7229	754
5	6475	0
6	6475	0
7	6475	0
8	6475	0
9	6475	437
10	6038	0
11	6038	0
12	6038	0
13	6038	0
14	6038	586
15	5452	0
16	5452	0
17	5452	0
18	5452	0
19	5452	

Table 1.1: Computer enumeration of $|N_k(4)|$ for $(a,b) = (3,2)$.

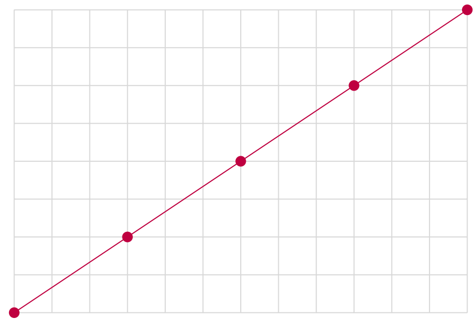


Figure 1.2: The boundary of a path whose endpoint is $(ga, gb) = (4 \cdot 3, 4 \cdot 2)$. The boundary contains $g + 1 = 5$ lattice points (red vertices).

P1 (**Constant on blocks**). The value $|N_k(g)|$ is constant on each of g distinct "blocks" of $a+b$ consecutive values of k.

P2 (**Strictly decreasing**). The value $|N_k(g)|$ is strictly decreasing between successive blocks.

We shall show that properties P1 and P2 both hold for all values of g, a, b.

Table 1.1, in addition to displaying the value of $|N_k(4)|$ for $(a, b) = (3, 2)$, also displays the value of the difference $|N_k(4)| - |N_{k+1}(4)|$. These differences suggest a strategy for achieving our central objective: identify a subset $S_k(g)$ of $N_k(g)$ having cardinality $|N_k(g)| - |N_{k+1}(g)|$, and show that the sets $N_k(g) \setminus S_k(g)$ and $N_{k+1}(g)$ are in bijection. We achieve this in our main result (Theorem 1.5). Properties P1 and P2 follow as consequences of the main result.

We introduce some additional vocabulary before defining the subset $S_k(g)$.

Definition 1.2 (Path concatenation)**.** Let p_1 and p_2 be paths having arbitrary startpoints. The *path concatenation* p_1p_2 is the path that starts at the startpoint of p_1, takes all the (ordered) steps of p_1, and then takes all the (ordered) steps of p_2. ⌟

Definition 1.3 (Boundary points)**.** The *boundary points* of a path $p \in N(g)$ are the points of p that lie on its boundary. Boundary points of p other than the startpoint $(0, 0)$ and endpoint (ga, gb) are *interior boundary points* of p. The startpoint and interior boundary points are the *non-terminal* boundary points. ⌟

The lattice points lying on the boundary joining $(0, 0)$ to (ga, gb) are the $g + 1$ points of the form (ja, jb) for $0 \leq j \leq g$ (see Figure 1.2). The number of interior boundary points of a path $p \in N(g)$ therefore lies between 0 and $g - 1$.

Recall that the number k of flaws of a path in $N(g)$ satisfies $0 \leq k < g(a + b)$. A path in $N(g)$ containing $g(a + b) - 1$ flaws has *max flaws*. Equivalently, the set of paths with max flaws is $N_{g(a+b)-1}(g)$.

Definition 1.4 (Subset $S_k(g)$)**.** For $0 \leq k < g(a+b)-1$, let $S_k(g)$ be the subset of $N_k(g)$ containing all paths of the form p_1p_2, where $p_1 \in N_0(g - j)$ and $p_2 \in N_k(j)$ for some j satisfying $0 < j < g$, and p_2 has max flaws. We write $S(g) \coloneqq \bigcup_k S_k(g)$. ⌟

See Figure 1.3 for two example paths in $S(4)$. A path in $S(g)$ is, for some j, the concatenation of a path p_1 from $(0, 0)$ to $\big((g-j)a, (g-j)b\big)$ having no flaws with a path p_2 from $(0, 0)$ to (ja, jb) having max flaws. The condition that p_2 has max flaws implies that $S_k(g)$ is empty unless $k = j(a + b) - 1$

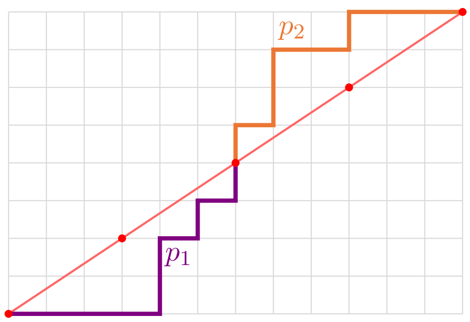
for some j satisfying $0 < j < g$ (and in particular $S_0(g)$ is empty). So we have

$$S_k(g) = \varnothing \quad \text{for } k \not\equiv -1 \pmod{a+b}, \tag{1.1}$$

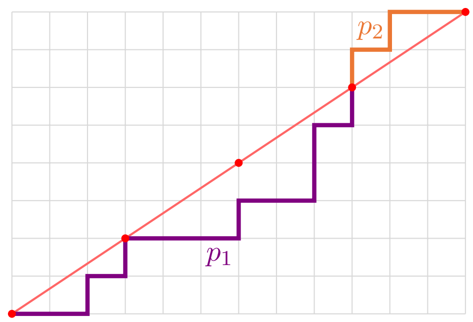
and, for each j satisfying $0 < j < g$,

$$S_{j(a+b)-1}(g) = \Big\{ p_1 p_2 : p_1 \in N_0(g-j) \text{ and } p_2 \in N_{j(a+b)-1}(j) \Big\}. \tag{1.2}$$

Note that the path p_2 in Definition 1.4 does not contain an interior boundary point (because it has max flaws), but the path p_1 might (see Figure 1.3).



(a) A path p_1p_2 in $S_9(4)$, where $p_1 \in N_0(2)$ and $p_2 \in N_9(2)$.



(b) A path p_1p_2 in $S_4(4)$, where $p_1 \in N_0(3)$ and $p_2 \in N_4(1)$.

Figure 1.3: Two example paths in $S(4)$ for $(a, b) = (3, 2)$.

1.1 Main result and consequences

Theorem 1.5 (Main result)**.** *Let g, k satisfy $0 \le k < g(a+b) - 1$. Then*

$$|N_k(g) \setminus S_k(g)| = |N_{k+1}(g)|.$$

We shall prove Theorem 1.5 in Section 2. Define

$$\mu_j(g) \coloneqq |N_{j(a+b)}(g)| \quad \text{for each } j \text{ satisfying } 0 \le j < g. \tag{1.3}$$

The following result is a first consequence of Theorem 1.5 and establishes property P1.

Corollary 1.6 (Constant on blocks)**.** *Let g be a positive integer. Then*

$$|N_k(g)| = \mu_j(g) \quad \textit{for all } j, k \textit{ satisfying } 0 \le j < g \textit{ and } j(a+b) \le k < (j+1)(a+b).$$

Proof. The result follows directly from Theorem 1.5 and (1.1). □

We now observe two further consequences of Theorem 1.5.

Corollary 1.7 (Recurrence relation)**.** *We have*

$$\mu_{j-1}(g) - \mu_0(g-j)\mu_{j-1}(j) = \mu_j(g) \quad \textit{for each } j \textit{ satisfying } 0 < j < g.$$

Proof. Let j satisfy $0 < j < g$ and let $k = j(a+b) - 1$. Since $S_k(g)$ is a subset of $N_k(g)$, we have by Theorem 1.5 that

$$|N_k(g)| - |S_k(g)| = |N_{k+1}(g)|. \tag{1.4}$$

We know from Corollary 1.6 that $|N_k(g)| = \mu_{j-1}(g)$ and $|N_{k+1}(g)| = \mu_j(g)$, and from (1.2) and Corollary 1.6 that $|S_k(g)| = |S_{j(a+b)-1}(g)| = \mu_0(g-j)\mu_{j-1}(j)$. Substitute these values into (1.4) to obtain the result. $\square$

The next corollary establishes property P2.

Corollary 1.8 (Strictly decreasing)**.** *We have* $\mu_0(g) > \mu_1(g) > \cdots > \mu_{g-1}(g)$.

Proof. This follows from Corollary 1.7, noting that $\mu_0(g-j)\mu_{j-1}(j) > 0$ for $0 < j < g$ by the definition of $\mu_j(g)$ given in (1.3). $\square$

1.2 The value of $\mu_j(g)$

Recall that our central objective is to find an explicit formula for $|N_k(g)|$. By Corollary 1.6, it is sufficient to determine the values $\mu_j(g)$.

The recurrence relation of Corollary 1.7 for $\mu_j(g)$ has a unique solution for each j, g satisfying $0 \le j < g$, provided the initial values $\mu_0(g)$ (contained in the top row of Table 1.2) are known for all g. The required initial values $\mu_0(g) = |N_0(g)|$ were given by Bizley [2] in 1954. We shall express these values in Corollary 1.11 in terms of a quantity H_g, introduced below. A derivation of $\mu_j(g)$ using only the values of $\mu_0(g)$ is given in [6, Chapter 4], making use of an identity from the theory of symmetric functions (see also Remark 1.9).

$\mu_0(1)$	$\mu_0(2)$	$\mu_0(3)$	$\mu_0(4)$	$\mu_0(5)$	$\cdots$
	$\mu_1(2)$	$\mu_1(3)$	$\mu_1(4)$	$\mu_1(5)$	$\cdots$
		$\mu_2(3)$	$\mu_2(4)$	$\mu_2(5)$	$\cdots$
			$\mu_3(4)$	$\mu_3(5)$	$\cdots$
				$\mu_4(5)$	$\cdots$
					$\cdots$

Table 1.2: All values $\mu_j(g)$ can be derived using only the values in the top row of the table, but can be more easily derived by also using the values in the coloured diagonal.

However, the values $\mu_j(g)$ can be derived more easily without relying on this identity, by additionally making use of the values $\mu_{g-1}(g)$ (contained in the coloured diagonal of Table 1.2). These additional values can be obtained using Corollary 1.6 from the values $|N_{g(a+b)-1}(g)|$ given by Bizley [2], and can be expressed in terms of another quantity E_g. This is the approach we shall use to establish the value of $\mu_j(g)$ in Theorem 1.12.

We now define the quantities H_g and E_g as sums over all integer partitions of g. Recall that a weakly increasing sequence of positive integers λ whose entries sum to g is a *partition of* g; we write $\lambda \vdash g$ to indicate this. Each entry of λ is called a *part*. We use the notation $\lambda = \langle 1^{m_1} 2^{m_2} \cdots \rangle$ to mean that λ has m_i parts equal to i, and then $g = \sum_{i \ge 1} i m_i$. For example, the partition $(1, 1, 2, 3)$ of 7 is also written $\langle 1^2 2^1 3^1 \rangle \vdash 7$.

For $i > 0$, let

$$c_i \coloneqq \frac{1}{i(a+b)}\binom{i(a+b)}{ia}. \tag{1.5}$$

In the case $i = 1$, this quantity is often referred to as a *rational Catalan number.*

For a partition $\lambda = \langle 1^{m_1} 2^{m_2} \cdots \rangle \vdash g$, let its *length* be $l(\lambda) \coloneqq \sum_{i\geq 1} m_i$, and let

$$c_\lambda \coloneqq \prod_{i\geq 1} \frac{c_i^{m_i}}{m_i!}. \tag{1.6}$$

Now let

$$H_g \coloneqq \sum_{\lambda\vdash g} c_\lambda, \tag{1.7}$$

$$E_g \coloneqq \sum_{\lambda\vdash g} (-1)^{g-l(\lambda)} c_\lambda, \tag{1.8}$$

and for convenience let

$$E_0 \coloneqq 1. \tag{1.9}$$

Remark 1.9. For each positive integer g, the quantities H_g and E_g are in fact specializations of (one part) *complete* and *elementary* symmetric functions h_g and e_g, respectively. The standard relationship between the *power sums* p_i and the complete and elementary symmetric functions are

$$h_g = \sum_{\lambda\vdash g} \frac{p_\lambda}{z_\lambda} \quad \text{and} \quad e_g = \sum_{\lambda\vdash g} (-1)^{g-l(\lambda)} \frac{p_\lambda}{z_\lambda},$$

where for a partition $\lambda = \langle 1^{m_1} 2^{m_2} \cdots \rangle$ of g, we write $z_\lambda \coloneqq \prod_{i\geq 1} i^{m_i} m_i!$ and $p_\lambda \coloneqq \prod_{i\geq 1} p_i^{m_i}$. Then the quantities H_g and E_g in (1.7) and (1.8) are obtained from the specialization given by $p_i = ic_i$. The identity

$$\sum_{i=0}^{g} (-1)^i E_i H_{g-i} = 0 \tag{1.10}$$

results from another well-known relationship between the complete and elementary symmetric functions. See [1] or the classic reference [11] for details on the above and other relevant background on symmetric functions. ⌟

Theorem 1.10 (Bizley [2])**.** *We have that*

$$|N_0(g)| = H_g,$$
$$|N_{g(a+b)-1}(g)| = (-1)^{g+1} E_g.$$

Using Corollary 1.6, we obtain the value of $\mu_0(g)$ and of $\mu_{g-1}(g)$.

Corollary 1.11 (Value of $\mu_0(g)$ and $\mu_{g-1}(g)$)**.** *We have that*

$$\mu_0(g) = H_g, \tag{1.11}$$

$$\mu_{g-1}(g) = (-1)^{g+1} E_g. \tag{1.12}$$

We now give a closed form expression for the value of $\mu_j(g)$, which we remark is a truncated version of the sum from (1.10).

Theorem 1.12 (Path enumeration formula)**.** *We have*

$$\mu_j(g) = \sum_{k=0}^{j} (-1)^k E_k H_{g-k} \quad \textit{for } 0 \le j < g.$$

Proof. It follows from Corollary 1.7 and (1.11) and (1.12) that when $0 < j < g$ we have

$$\mu_j(g) = \mu_{j-1}(g) + (-1)^j E_j H_{g-j}.$$

Therefore (by an implicit induction) for $0 \le j < g$ we have

$$\begin{aligned} \mu_j(g) &= \mu_0(g) + \sum_{k=1}^{j} (-1)^k E_k H_{g-k} \\ &= \sum_{k=0}^{j} (-1)^k E_k H_{g-k} \end{aligned} \tag{1.13}$$

using (1.9) and (1.11). □

Remark 1.13. The expression (1.13) can be written more compactly as a specialization of a Schur function via the identity $\sum_{k=0}^{j} (-1)^k e_k h_{g-k} = (-1)^j s_{\langle 1^j (g-j)^1 \rangle}$, where $s_{\langle 1^j (g-j)^1 \rangle}$ is the Schur function indexed by the hook partition $\langle 1^j (g-j)^1 \rangle$. See [11, Section I.3, Example 9]. ⌟

We next state separately the important special case $g = 1$ of the path enumeration formula Theorem 1.12.

Theorem 1.14 (Special case $g = 1$)**.** *We have*

$$|N_k(1)| = \frac{1}{a+b} \binom{a+b}{a} \quad \textit{for all } k \textit{ satisfying } 0 \le k < a+b.$$

Proof. Let $g = 1$ and let k satisfy $0 \le k < a+b$. Then $j = 0$ in Corollary 1.6, and this gives $|N_k(1)| = \mu_0(1)$. Similarly, we have $j = 0$ in Theorem 1.12, which after applying (1.9) gives $\mu_0(1) = H_1$. Equate these two expressions for $\mu_0(1)$ to give

$$|N_k(1)| = H_1. \tag{1.14}$$

Now by (1.7), (1.6), and (1.5) we have $H_1 = c_{\langle 1^1 \rangle} = c_1 = \frac{1}{a+b}\binom{a+b}{a}$. Substitute in (1.14) to give the result. □

Theorem 1.14 gives that $|N_k(1)|$ is independent of k. We highlight this special case because our general construction used to prove Theorem 1.5 simplifies greatly in the case $g = 1$. We re-examine this special case in Section 2.5.

We now consider the special case $a = b = 1$, when the slope of the boundary is 1. Although Theorem 1.12 already provides an expression for $\mu_j(g)$ in this case, we now derive an alternative formula involving Catalan numbers that appears to be considerably simpler. In Section 1.3.2 we shall compare this formula to a theorem due to Chung and Feller.

Theorem 1.15 (Alternative formula for $a = b = 1$)**.** *Let $a = b = 1$. Then*

$$\mu_j(g) = |N_{2j}(g)| = |N_{2j+1}(g)| = \sum_{k=j+1}^{g} C_{k-j-1}C_{g-k+j} \quad \textit{for all } j \textit{ satisfying } 0 \le j < g,$$

where

$$C_i := \frac{1}{i+1}\binom{2i}{i} \tag{1.15}$$

is the i^{th} Catalan number.

Proof. Fix the positive integer g. The first two equalities of the theorem hold by Corollary 1.6. We establish the third equality by induction for $0 \le j < g$.

Since $a = b = 1$, the set $N(g)$ comprises all paths from $(0,0)$ to (g,g). A classical result states that the number of such paths with zero flaws (known as *Dyck* or *Catalan* paths) is C_g, and so $\mu_0(g) = C_g$. The well-known identity $C_g = \sum_{k=1}^{g} C_{k-1}C_{g-k}$ then gives the base case $j = 0$ of the induction. Assume all cases up to $j-1$ hold for some j satisfying $0 < j < g$. Then the recurrence relation of Corollary 1.7 gives

$$\mu_j(g) = \mu_{j-1}(g) - C_{g-j}\,\mu_{j-1}(j). \tag{1.16}$$

By Corollary 1.6, the value $\mu_{j-1}(j) = |N_{2j-1}(j)|$ is the number of paths from $(0,0)$ to (j,j) having max flaws. By rotation through $180°$, this equals the number of paths from $(0,0)$ to (j,j) with zero flaws having no interior boundary points, which (by removing the initial $(1,0)$ step and the final $(0,1)$ step) equals the number of Dyck paths from $(0,0)$ to $(j-1,j-1)$, namely C_{j-1}. Substitute in (1.16) and use the inductive hypothesis to give

$$\begin{aligned}
\mu_j(g) &= \sum_{k=j}^{g} C_{k-j}C_{g-k+j-1} - C_{g-j}C_{j-1} \\
&= \sum_{k=j}^{g-1} C_{k-j}C_{g-k+j-1} \\
&= \sum_{K=j+1}^{g} C_{K-j-1}C_{g-K+j}
\end{aligned}$$

by reindexing with $K = k+1$. Therefore case j holds and the induction is complete. □

Example 1.16 (Computation using the path enumeration formula)**.** Let $(a,b) = (3,2)$ and $g = 4$. We illustrate the use of the path enumeration formula Theorem 1.12 to calculate the number $|N_k(4)|$ of paths from $(0,0)$ to $(12,8)$ having k flaws, for each k satisfying $0 \le k < 20$. By Corollary 1.6, it is sufficient to determine $\mu_j(4)$ for each $j = 0,1,2,3$.

We begin by listing the partitions of the integers $1,2,3,4$.

$$\begin{aligned}
&\text{Partitions of } 4 : \langle 4^1\rangle,\ \langle 1^1 3^1\rangle,\ \langle 2^2\rangle,\ \langle 1^2 2^1\rangle,\ \langle 1^4\rangle, \\
&\text{Partitions of } 3 : \langle 3^1\rangle,\ \langle 1^1 2^1\rangle,\ \langle 1^3\rangle, \\
&\text{Partitions of } 2 : \langle 2^1\rangle,\ \langle 1^2\rangle,
\end{aligned}$$

$$\text{Partitions of } 1 : \langle 1^1 \rangle.$$

Using (1.5), we compute

$$c_1 = 2,\ c_2 = 21,\ c_3 = \frac{1001}{3},\ c_4 = \frac{12597}{2}.$$

Using (1.6), we then compute (for example)

$$c_{\langle 1^2 2^1 \rangle} = \left(\frac{c_1^2}{2!}\right)\left(\frac{c_2^1}{1!}\right) = 42, \quad c_{\langle 1^3 \rangle} = \left(\frac{c_1^3}{3!}\right) = \frac{4}{3}.$$

The full set of c_λ values for $\lambda \vdash j$ where $1 \leq j \leq 4$ is

$$\begin{aligned}
&c_{\langle 4^1 \rangle} = \frac{12597}{2}, \quad && c_{\langle 1^1 3^1 \rangle} = \frac{2002}{3}, \quad && c_{\langle 2^2 \rangle} = \frac{441}{2}, \quad && c_{\langle 1^2 2^1 \rangle} = 42, \quad && c_{\langle 1^4 \rangle} = \frac{2}{3},\\
&c_{\langle 3^1 \rangle} = \frac{1001}{3}, && c_{\langle 1^1 2^1 \rangle} = 42, && c_{\langle 1^3 \rangle} = \frac{4}{3}, && && \\
&c_{\langle 2^1 \rangle} = 21, && c_{\langle 1^2 \rangle} = 2, && && && \\
&c_{\langle 1^1 \rangle} = 2. && && && &&
\end{aligned}$$

Using (1.7) and (1.8), we next calculate (for example)

$$\begin{aligned}
H_3 &= c_{\langle 3^1 \rangle} + c_{\langle 1^1 2^1 \rangle} + c_{\langle 1^3 \rangle} = \frac{1001}{3} + 42 + \frac{4}{3} = 377,\\
E_3 &= (-1)^{3-1} c_{\langle 3^1 \rangle} + (-1)^{3-2} c_{\langle 1^1 2^1 \rangle} + (-1)^{3-3} c_{\langle 1^3 \rangle} = \frac{1001}{3} - 42 + \frac{4}{3} = 293.
\end{aligned}$$

The full set of H_k and E_k values is

$$\begin{aligned}
H_4 &= 7229, \qquad & E_4 &= -5452,\\
H_3 &= 377, & E_3 &= 293,\\
H_2 &= 23, & E_2 &= -19,\\
H_1 &= 2, & E_1 &= 2,\\
& & E_0 &= 1.
\end{aligned}$$

Using Theorem 1.12, we then determine that

$$\begin{aligned}
\mu_0(4) &= E_0 H_4 = 1 \cdot 7229 = 7229\\
\mu_1(4) &= E_0 H_4 - E_1 H_3 = 1 \cdot 7229 - 2 \cdot 377 = 6475\\
\mu_2(4) &= E_0 H_4 - E_1 H_3 + E_2 H_2 = 1 \cdot 7229 - 2 \cdot 377 - 19 \cdot 23 = 6038\\
\mu_3(4) &= E_0 H_4 - E_1 H_3 + E_2 H_2 - E_3 H_1 = 1 \cdot 7229 - 2 \cdot 377 - 19 \cdot 23 - 293 \cdot 2 = 5452.
\end{aligned}$$

(Alternatively, we may use (1.12) for a more direct calculation of the last value $\mu_3(4) = (-1)^{4+1} E_4 = 5452$.)

Using Corollary 1.6, we may now determine the value of $|N_k(4)|$ for each k satisfying $0 \leq k < 20$. The resulting values agree with the computer enumeration shown in Table 1.1. ⌟

We remark, as noted by Bizley [2], that both H_g and E_g are necessarily integers because of the counting result Theorem 1.10, even though this is not readily apparent from the forms (1.6), (1.7), and (1.8). We further remark that although the quantity c_i defined in (1.5) is not necessarily an integer, it is not difficult to show that ic_i is an integer.

1.3 Previous results

In order to place our results in a wider context, we briefly review the literature on the enumeration of lattice paths in the presence of a linear boundary.

1.3.1 How much of a path lies above the boundary?

The path enumeration setting we consider involves paths with step set $\{(1,0),(0,1)\}$ in the two-dimensional lattice $\mathbb{Z}^2$; a boundary line joining the startpoint $(0,0)$ of a path to its endpoint (ga, gb), where g, a, b are positive integers such that a and b are coprime; and k flaws. Previous authors have defined a flaw differently from us, namely as a certain type of step (usually a $(0,1)$ step) of the path that lies above the boundary. Each of the references [4, 8, 9, 12, 13] adopts this step-based definition of flaw and a notion of the "wrong" side of the boundary, although the precise definition is not identical in all five references.

In the more general case that we consider here, where the value of b/a need not necessarily be an integer, the definition of a flaw as a step is no longer appropriate since some steps can lie only partially above the boundary (see Figure 1.4). Our definition of a flaw, as a lattice point of the path that lies above the boundary, does not have this ambiguity.

Note that these two definitions of flaws are genuinely different: Figure 1.5 shows that even in the case $b/a = 1$ there is no simple relationship between the number of $(0,1)$ steps lying above the boundary and the number of lattice points lying above the boundary.

We review previous results relating to these two definitions of flaws in Sections 1.3.2 and 1.3.3.

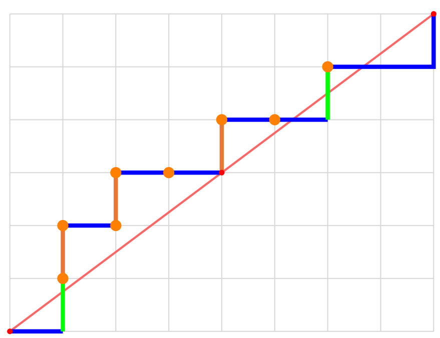


Figure 1.4: A path with a boundary whose slope is not an integer. This shows that both $(0,1)$ steps and $(1,0)$ steps can lie partially above and below the boundary simultaneously, whereas lattice points cannot.

1.3.2 Boundaries of integer slope with (0,1) steps as flaws

In this part of the review, we take k to be the number of $(0,1)$ steps of a path from $(0,0)$ to (g, gb) that lie above the boundary of integer slope b.

Firstly consider paths from $(0,0)$ to (g,g). The number of such paths having $k = 0$ (no $(0,1)$ steps lying above the boundary, which is equivalent to having no lattice points above the boundary) is the Catalan number C_g given by (1.15), as mentioned in the proof of Theorem 1.15. Chung and Feller's influential 1949 work [5] showed that, remarkably, the same count applies for all k.

(a) A path in which two $(0,1)$ steps and three lattice points lie above the boundary.

(b) A path in which two $(0,1)$ steps and two lattice points lie above the boundary.

Figure 1.5: Even in the case $b/a = 1$, there is no simple relationship between the number of $(0,1)$ steps lying above the boundary and the number of lattice points lying above the boundary.

Theorem 1.17 (Chung-Feller [5, Theorem 2A])**.** *Let k satisfy $0 \leq k \leq g$. Then the number of paths from $(0,0)$ to (g,g) having k of the $(0,1)$ steps lying above the boundary is C_g.*

Theorem 1.17 can be proven using bijective methods [3]. The reader is invited to compare Theorems 1.15 and 1.17: both apply to a boundary of slope 1, but Theorem 1.15 takes flaws to be points above the boundary whereas Theorem 1.17 takes flaws to be $(0,1)$ steps above the boundary.

Huq generalized Theorem 1.17 to paths from $(0,0)$ to (g, gb).

Theorem 1.18 (Huq [9, Corollary 5.1.2])**.** *Let k satisfy $0 \leq k \leq gb$. Then the number of paths from $(0,0)$ to (g, gb) having k of the $(0,1)$ steps lying above the boundary is*

$$\frac{1}{gb+1}\binom{(b+1)g}{g}.$$

Further variations on Theorem 1.17 have been found [8, 10, 12].

1.3.3 Boundaries of rational slope with lattice points as flaws

In this part of the review, we take k to be the number of lattice points of a path from $(0,0)$ to (ga, gb) that lie strictly above the boundary. (This is the measure k used for $N_k(g)$ in Definition 1.1.)

In 1950, Grossman [7] conjectured an explicit formula for the number $|N_0(g)|$ of paths from $(0,0)$ to (ga, gb) having no flaws. In 1954, Bizley [2, Eq. (10)] proved Grossman's formula using generating functions. Bizley [2, Eq. (8)] also obtained an explicit formula for the number of paths having neither flaws nor interior boundary points. In other words, such paths remain strictly below the boundary except at the endpoints. Since the set of such paths is in bijection with the set of paths having max flaws (via rotation), this second result of Bizley's gives the value $|N_{g(a+b)-1}(g)|$. The values $|N_0(g)|$ and $|N_{g(a+b)-1}(g)|$ are stated in Theorem 1.10.

2 Proof of the main result

For convenience, we restate our main result here.

Theorem 1.5. *Let g, k satisfy $0 \leq k < g(a+b) - 1$. Then*

$$|N_k(g) \setminus S_k(g)| = |N_{k+1}(g)|.$$

2.1 Proof outline

We shall prove our main result by considering fixed g, k satisfying $0 \leq k < g(a+b) - 1$ and constructing injective maps

$$\begin{aligned} \phi &\colon N_k(g) \setminus S_k(g) \to N_{k+1}(g) \\ \psi &\colon N_{k+1}(g) \to N_k(g) \setminus S_k(g). \end{aligned}$$

In fact, the map ψ we shall construct is the inverse of ϕ, although we shall not require this fact in our proof. We partition the set $N_k(g) \setminus S_k(g)$ into subsets X and Y, and partition (using a different rule) the set $N_{k+1}(g)$ into subsets $\mathcal{X}$ and $\mathcal{Y}$. We allow each of the partitioning subsets to be empty. Using these partitions, we then specify the action of ϕ using injective submaps ϕ^X and ϕ^Y, and the action of ψ using injective submaps $\psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$ (see Figure 2.1).

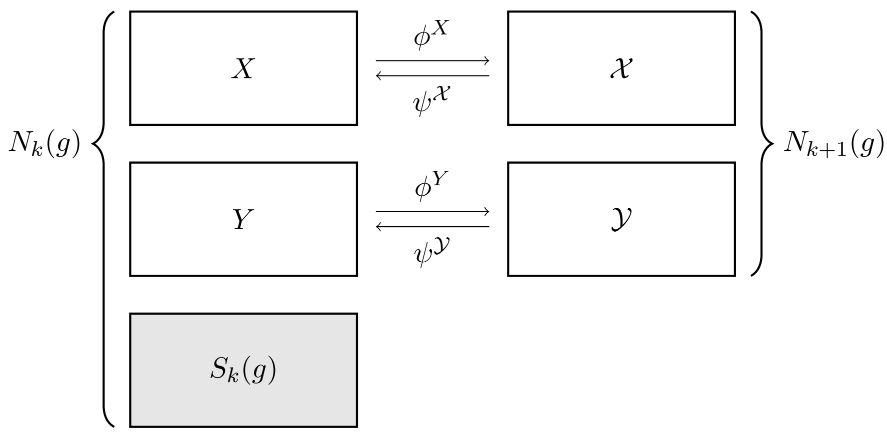


Figure 2.1: The map $\phi\colon N_k(g) \setminus S_k(g) \to N_{k+1}(g)$ is defined piecewise using the maps $\phi^X : X \to \mathcal{X}$ and $\phi^Y : Y \to \mathcal{Y}$. The map $\psi\colon N_{k+1}(g) \to N_k(g) \setminus S_k(g)$ is likewise defined piecewise using the maps $\psi^{\mathcal{X}} : \mathcal{X} \to X$ and $\psi^{\mathcal{Y}} : \mathcal{Y} \to Y$.

To prove Theorem 1.5, it suffices to

1. specify the partition of $N_k(g) \setminus S_k(g)$ and of $N_{k+1}(g)$ as illustrated in Figure 2.1, and
2. define the maps ϕ and ψ, and show that they are both injective.

2.2 Partition of sets $N_k(g) \setminus S_k(g)$ and $N_{k+1}(g)$

We first introduce some additional terminology. Recall that the boundary of a path in $N(g)$ is the line from $(0, 0)$ to (ga, gb).

Definition 2.1 (Elevation)**.** Let (i, j) be a point of a path in $N(g)$. The *elevation* of (i, j) is $ja - ib$.

The elevation of a point of a path in $N(g)$ is a measure of the directed distance from the point to the boundary. Points on the boundary have zero elevation; points above the boundary have positive elevation; points below the boundary have negative elevation.

Definition 2.2 (Highest points below, lowest points above)**.** Let p be a path. The *highest points below the boundary (HPBs)* of p are those points of p (if any) lying strictly below the boundary and attaining the closest elevation to zero. The *lowest points above the boundary (LPAs)* of p are defined analogously. ⌟

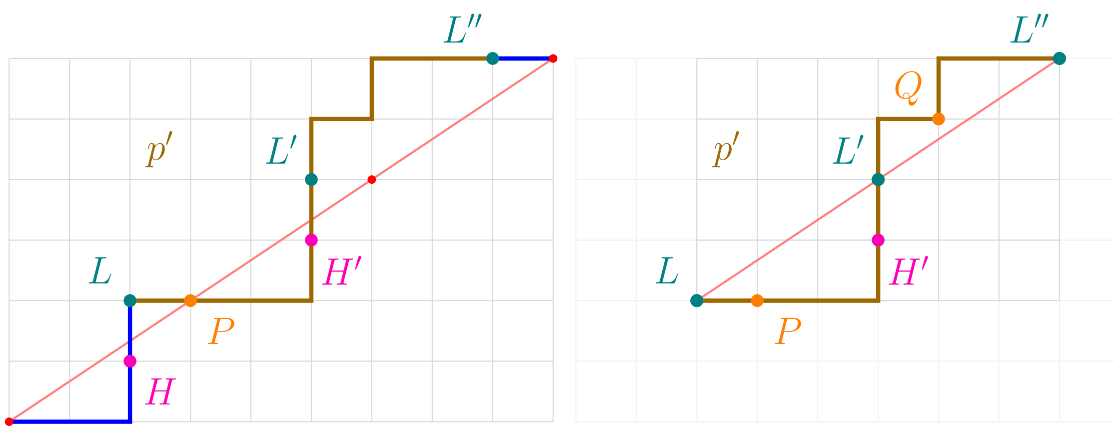
See Figure 2.2a for an illustration of a path p with HPBs H, H' and LPAs L, L', L''. We note that the possible elevation values for an LPA are $1, 2, \ldots, \min(a, b)$, and that the possible elevation values for an HPB are $-1, -2, \ldots, -\min(a, b)$. Both the number of HPBs and the number of LPAs of a path in $N(g)$ lie in $\{0, 1, \ldots, g\}$.

We shall often consider a *subpath* p' of a path p, namely a consecutive sequence of steps of p. When viewed as a separate path in its own right, the boundary of p' need not coincide with the boundary of p (nor even have the same slope) and so its HPBs and LPAs need not necessarily be the same as those of p (see Figure 2.2). When we wish to view p' as a path in its own right, we shall refer to "the path p'"; when we wish to view p' as a part of p we shall refer to "the subpath p'".

Definition 1.2 describes the combination of paths p_1 and p_2 to form the concatenated path p_1p_2. To reverse this process, we *split* the path p_1p_2 at the endpoint of p_1 into component paths p_1, p_2. We can similarly split a path at two distinct points to form component paths p_1, p_2, p_3. If the elevation of the startpoint and endpoint of p_i (viewed as a subpath) are equal, then p_i (viewed as a path) has a boundary with the same slope (the same values of a and b) as the full path.

We make the following observation about the change in the number of flaws when a path is split at an HPB or LPA and the resulting subpaths are interchanged.

Observation 2.3. Let p be a path containing exactly β interior boundary points and exactly λ LPAs. Suppose that p is split at an HPB H into p_1p_2, so that H is the endpoint of p_1 and the startpoint of p_2. Then the rearranged path p_2p_1 has exactly $\beta + 1$ more flaws than p, namely all β interior boundary points of p together with the endpoint of p_2. If instead p is split at an LPA into p_1p_2, then the rearranged path p_2p_1 has exactly λ fewer flaws than p, namely all λ LPAs of p. ⌟



(a) A path p and (in brown) its subpath p'.

(b) The subpath p' as a path in its own right.

Figure 2.2: Let $(a, b) = (3, 2)$. The path p is a member of $N_7(3) \setminus S_7(3)$ containing the interior boundary point P and the LPAs L, L', L'' and HPBs H, H'. The subpath p' contains the same interior boundary point and LPAs as p, as well as the HPB H' of p. The path p' (on its own) has the same slope as p, but is a member of $S_4(2) \subseteq N_4(2)$ containing the boundary points L, L', L'', the unique HPB P, and the unique LPA Q. We note that H' is not an HPB of the path p'.

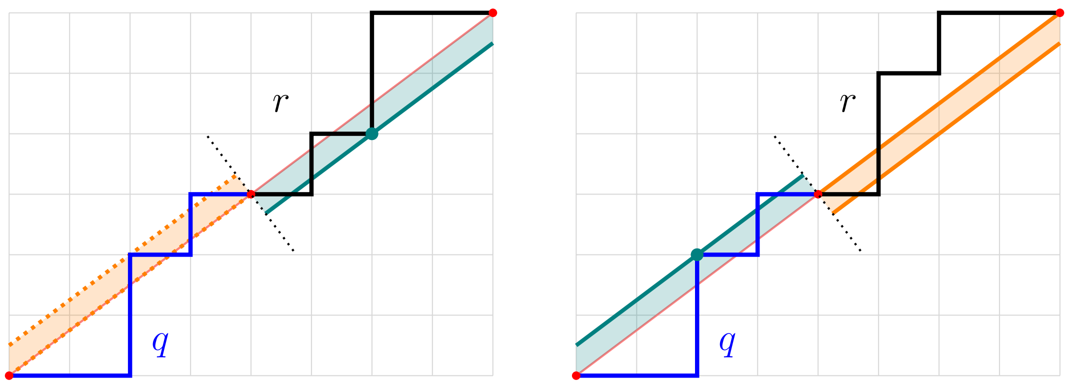
We now define the subsets X and Y of $N_k(g) \setminus S_k(g)$ by reference to an arbitrary path $p \in N_k(g) \setminus S_k(g)$. Split p at its last non-terminal boundary point into qr, and regard q and r as paths in their own right. If p has no interior boundary points, then p is split at its startpoint and q is empty. Since $p \notin S_k(g)$, either q has at least one flaw or r has non-max flaws; in the latter case, r has at least one HPB because p splits at its last non-terminal boundary point into qr and so r itself has no interior boundary points. Therefore exactly one of three cases holds:

Case 1: q has no flaws and r has non-max flaws. Then $p \in X$.

Case 2: q has at least one flaw and r has max flaws. Then $p \in Y$.

Case 3: q has at least one flaw and r has non-max flaws. If the LPAs of q are closer to the boundary of p than are the HPBs of r, then $p \in Y$. Otherwise $p \in X$.

See Figure 2.3 for an illustration of Case 3.



(a) A path p in X. (b) A path p in Y.

Figure 2.3: Let $(a, b) = (4, 3)$ and split the path p into qr at its last non-terminal boundary point. In diagram (a), we have $p \in N_6(2)$ and the green region is determined by the elevation of the HPBs of the subpath r; this in turn determines an open orange "forbidden region" that the subpath q must avoid so that $p \in X$. In diagram (b), we have $p \in N_7(2)$ and the green region is determined by the elevation of the LPAs of the subpath q; this in turn determines a closed orange "forbidden region" that the subpath r must avoid so that $p \in Y$.

We now give a more concise definition of the subsets X and Y. Recall that k is fixed and satisfies $0 \leq k < g(a + b) - 1$ throughout this section.

Definition 2.4 (The subsets X and Y)**.** Let $p \in N_k(g) \setminus S_k(g)$. Split p at its last non-terminal boundary point into qr, and regard q and r as paths. The path p lies in Y provided:

(i) q has at least one flaw, and

(ii) the elevation of the LPAs of q is smaller than the magnitude of the elevation of the HPBs of r (if any).

Otherwise, p lies in X. ⌟

Note that Y is empty if $k = 0$. We now use Definition 2.4 to specify a canonical representation for a path in each of X and Y as a concatenation of paths.

Definition 2.5 (Canonical representation of paths in X and Y)**.** Let $p \in N_k(g) \setminus S_k(g)$. Split p at its last non-terminal boundary point into $p = qr$.

Case $p \in X$**:** the path r has at least one HPB. Split r at its last HPB into $r = r_1r_2$. The canonical representation of p is qr_1r_2.

Case $p \in Y$**:** the path q has at least one LPA. Split q at its last LPA into $q = q_1q_2$. The canonical representation of p is q_1q_2r.

We now define the subsets $\mathcal{X}$ and $\mathcal{Y}$ of $N_{k+1}(g)$ by reference to an arbitrary path $\mathbb{p} \in N_{k+1}(g)$. Since $\mathbb{p}$ has at least one flaw, it has at least one LPA.

Definition 2.6 (The subsets $\mathcal{X}$ and $\mathcal{Y}$)**.** Let $\mathbb{p} \in N_{k+1}(g)$. The path $\mathbb{p}$ lies in $\mathcal{Y}$ provided:

(i) $\mathbb{p}$ has at least two LPAs, and

(ii) the subpath of $\mathbb{p}$ lying between the last two LPAs of $\mathbb{p}$ contains no boundary points of $\mathbb{p}$.

Otherwise, $\mathbb{p}$ lies in $\mathcal{X}$.

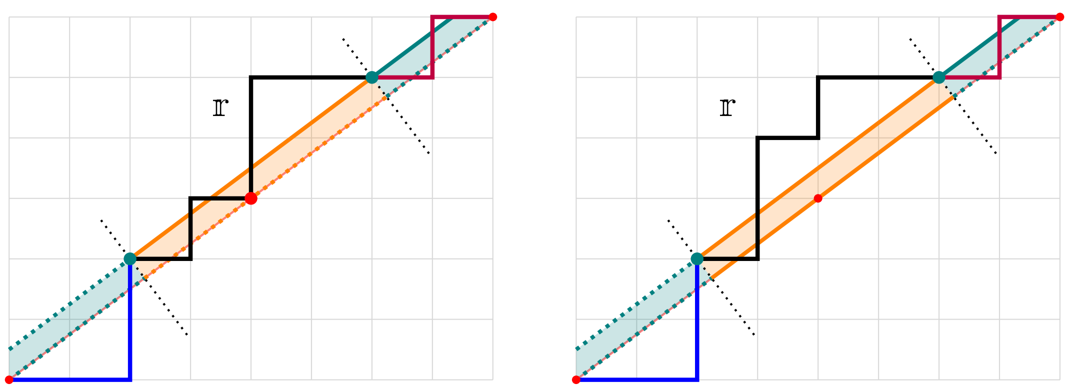
Note that $\mathcal{Y}$ is empty if $k = 0$. See Figure 2.4 for an illustration of Definition 2.6.

We now use Definition 2.6 to specify a canonical path split for a path in each of $\mathcal{X}$ and $\mathcal{Y}$.

Definition 2.7 (Canonical representation of paths in $\mathcal{X}$ and $\mathcal{Y}$)**.** Let $\mathbb{p} \in N_{k+1}(g)$.

Case $\mathbb{p} \in \mathcal{X}$**:** let L be the last LPA of $\mathbb{p}$, and let B be the boundary point of $\mathbb{p}$ (possibly the startpoint of $\mathbb{p}$) which immediately precedes L. Split $\mathbb{p}$ at B and L into $\mathbb{p} = \mathbb{q}\mathbb{r}_2\mathbb{r}_1$. The canonical representation of $\mathbb{p}$ is $\mathbb{q}\mathbb{r}_2\mathbb{r}_1$.

Case $\mathbb{p} \in \mathcal{Y}$**:** the path $\mathbb{p}$ has at least two LPAs. Split $\mathbb{p}$ at its last two LPAs into $\mathbb{p} = \mathbb{q}_1\mathbb{r}\mathbb{q}_2$. The canonical representation of $\mathbb{p}$ is $\mathbb{q}_1\mathbb{r}\mathbb{q}_2$.

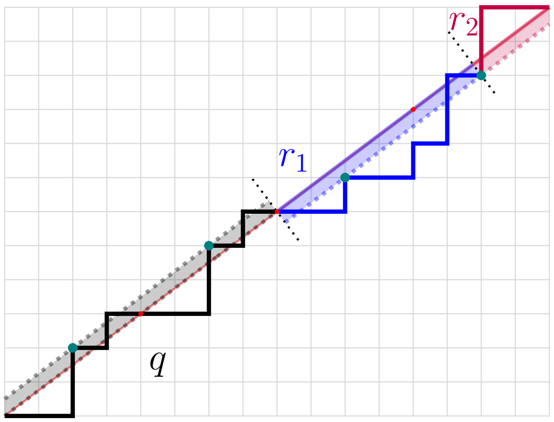


(a) A path $\mathbb{p}$ in $\mathcal{X}$. (b) A path $\mathbb{p}$ in $\mathcal{Y}$.

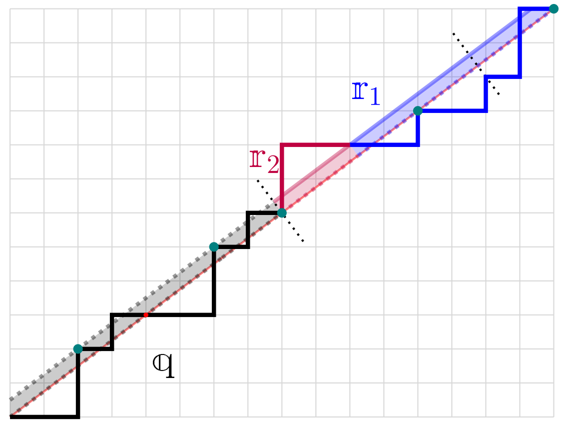
Figure 2.4: Let $(a, b) = (4, 3)$. In diagram (a), we have $\mathbb{p} \in N_7(2)$ and the subpath $\mathbb{r}$ between the (last) two LPAs of $\mathbb{p}$ contains a boundary point of $\mathbb{p}$. In diagram (b), we have $\mathbb{p} \in N_8(2)$ and the subpath $\mathbb{r}$ between the (last) two LPAs of $\mathbb{p}$ contains no boundary points of $\mathbb{p}$.

2.3 The actions of ϕ^X, ϕ^Y, $\psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$

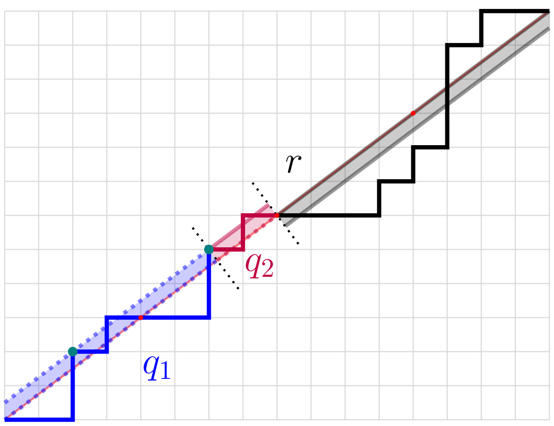
We now define the maps $\phi^X, \phi^Y, \psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$, whose domains and codomains are given in Figure 2.1. Illustrations of these maps are given in Figure 2.5.



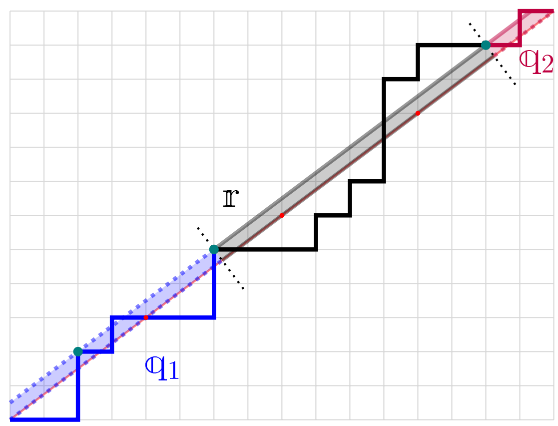
(a) Canonical representation of a path p in X.



(b) Canonical representation of a path $\mathbb{p}$ in $\mathcal{X}$.



(c) Canonical representation of a path p in Y.



(d) Canonical representation of a path $\mathbb{p}$ in $\mathcal{Y}$.

Figure 2.5: The maps ϕ^X and $\psi^{\mathcal{X}}$ act on the paths in diagrams (a) and (b), respectively, and their images are (b) and (a), respectively. Similarly the maps ϕ^Y and $\psi^{\mathcal{Y}}$ act on paths in diagram (c) and (d), respectively, and their images are (d) and (c), respectively. The LPAs and HPBs of the paths determine open or closed forbidden regions within which no points of the path can lie.

2.3.1 The actions of ϕ^X and ϕ^Y

We now define the maps ϕ^X and ϕ^Y.

Definition 2.8 (Actions of ϕ^X and ϕ^Y)**.** Let $p \in N_k(g) \setminus S_k(g)$.

Case $p \in X$: Write $p = qr_1r_2$ according to Definition 2.5. Then $\phi^X : X \to \mathcal{X}$ is given by

$$\phi^X(qr_1r_2) = qr_2r_1.$$

Case $p \in Y$: Write $p = q_1q_2r$ according to Definition 2.5. Then $\phi^Y : Y \to \mathcal{Y}$ is given by

$$\phi^Y(q_1q_2r) = q_1rq_2.$$

⌟

Proposition 2.9. *The map ϕ^X is well-defined.*

Proof. Let $p \in X$. We must check that $\phi^X(p) = qr_2r_1$ belongs to $\mathcal{X}$. Let H be the startpoint of r_2. By Definition 2.5, H is the last HPB of the path r_1r_2. Since p is split at its last non-terminal boundary point into paths q and r_1r_2, we have:

1. the path r_1r_2 has no interior boundary points.

Since $p \in X$, by Definition 2.4 we have:

2. the elevation of the LPAs of the path q (if any) is greater than or equal to the magnitude of the elevation of H in the path r_1r_2.

It follows from statement 1 and Observation 2.3 that:

3. the path r_2r_1 has exactly one more flaw than does r_1r_2.

It follows from statements 1 and 2 that:

4. the subpath r_2r_1 contains exactly one of the LPAs of $\phi^X(p)$ (namely the endpoint of r_2).

It follows from statement 3 that $\phi^X(p)$ contains exactly one more flaw than p. Furthermore, since the startpoint H of r_2 is a boundary point of the path $\phi^X(p) = qr_2r_1$, statement 4 implies that $\phi^X(p)$ cannot simultaneously satisfy both conditions *(i)* and *(ii)* of Definition 2.6. Therefore $\phi^X(p) \in \mathcal{X}$, as required. □

Remark 2.10. Continue with the notation from the proof of Proposition 2.9. We note for use in Section 2.4 that, since the startpoint H of r_2 is the last HPB of the path r_1r_2 and is a boundary point of the path $\phi^X(p)$, we have that H is the only boundary point of $\phi^X(p) = qr_2r_1$ contained in the subpath r_2. ⌟

Proposition 2.11. *The map ϕ^Y is well-defined.*

Proof. Let $p \in Y$. We must check that $\phi^Y(p) = q_1rq_2$ belongs to $\mathcal{Y}$. Let L be the endpoint of q_1. By Definition 2.5, L is the last LPA of the path q_1q_2. Since $p \in Y$, by Definition 2.4 the elevation of the LPAs of the path q_1q_2 (including L) is smaller than the magnitude of the elevation of the HPBs of the path r (if any). Therefore:

1. the path $\phi^Y(p)$ contains exactly one more flaw than p, namely the startpoint L' of q_2.
2. the points L and L' are the (distinct) last two LPAs of $\phi^Y(p)$ (since the path r has no interior boundary points by Definition 2.5).
3. the subpath r contains no boundary points of $\phi^Y(p)$.

This shows by Definition 2.6 that $q_1rq_2 \in \mathcal{Y}$, as required. □

2.3.2 The actions of $\psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$

We now define the maps $\psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$.

Definition 2.12 (Actions of $\psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$)**.** Let $\mathbb{p} \in N_{k+1}(g)$.

Case $\mathbb{p} \in \mathcal{X}$: Write $\mathbb{p} = \mathbb{q}\mathbb{r}_2\mathbb{r}_1$ according to Definition 2.7. Then $\psi^{\mathcal{X}} : \mathcal{X} \to X$ is given by

$$\psi^{\mathcal{X}}(\mathbb{q}\mathbb{r}_2\mathbb{r}_1) = \mathbb{q}\mathbb{r}_1\mathbb{r}_2.$$

Case $\mathbb{p} \in \mathcal{Y}$: Write $\mathbb{p} = \mathbb{q}_1\mathbb{r}\mathbb{q}_2$ according to Definition 2.7. Then $\psi^{\mathcal{Y}} : \mathcal{Y} \to Y$ is given by

$$\psi^{\mathcal{Y}}(\mathbb{q}_1\mathbb{r}\mathbb{q}_2) = \mathbb{q}_1\mathbb{q}_2\mathbb{r}.$$ $\lrcorner$

Proposition 2.13. *The map $\psi^{\mathcal{X}}$ is well-defined.*

Proof. Let $\mathbb{p} \in \mathcal{X}$. We must check that $\psi^{\mathcal{X}}(\mathbb{p}) = \mathbb{q}\mathbb{r}_1\mathbb{r}_2$ belongs to X.

Let L be the endpoint of the path $\mathbb{r}_2$. By Definition 2.7, we have:

1. L is the last LPA of $\mathbb{p} = \mathbb{q}\mathbb{r}_2\mathbb{r}_1$, and the startpoint of $\mathbb{r}_2$ is the boundary point of $\mathbb{p}$ which immediately precedes L.

By Definition 2.6, we have:

2. either $\mathbb{p}$ has exactly one LPA, or the subpath of $\mathbb{p}$ lying between the last two LPAs of $\mathbb{p}$ contains a boundary point of $\mathbb{p}$.

It follows from statements 1 and 2 that:

3. the subpath $\mathbb{r}_2\mathbb{r}_1$ of $\mathbb{p}$ contains exactly one LPA of $\mathbb{p}$, namely the point L.

Statement 3 and Observation 2.3 imply that:

4. the path $\psi^{\mathcal{X}}(\mathbb{p}) = \mathbb{q}\mathbb{r}_1\mathbb{r}_2$ splits at its last non-terminal boundary point into the paths $\mathbb{q}$ and $\mathbb{r}_1\mathbb{r}_2$.
5. the path $\psi^{\mathcal{X}}(\mathbb{p}) = \mathbb{q}\mathbb{r}_1\mathbb{r}_2$ has exactly one fewer flaw than $\mathbb{p}$.

The elevation of L in the path $\mathbb{r}_2\mathbb{r}_1$ equals the magnitude of the elevation of the HPBs of the path $\mathbb{r}_1\mathbb{r}_2$. Since L is an LPA of $\mathbb{p}$ by statement 1, this gives:

6. the elevation of the LPAs of the path $\mathbb{q}$ (if any) is greater than or equal to the magnitude of the elevation of the HPBs of the path $\mathbb{r}_1\mathbb{r}_2$.

Statements 4, 5 and 6 show by Definition 2.4 that $\psi^{\mathcal{X}}(\mathbb{p}) \in X$. $\square$

Remark 2.14. Continue with the notation from the proof of Proposition 2.13. We note for use in Section 2.4 that statement 1 implies the endpoint of $\mathbb{r}_1$ is the last HPB of the path $\mathbb{r}_1\mathbb{r}_2$. $\lrcorner$

Proposition 2.15. *The map $\psi^{\mathcal{Y}}$ is well-defined.*

Proof. Let $\mathbb{p} \in \mathcal{Y}$. We must check $\psi^{\mathcal{Y}}(\mathbb{p}) = \mathbb{q}_1\mathbb{q}_2\mathbb{r}$ belongs to Y. Let L be the endpoint of the path $\mathbb{q}_1$, and let L' be the startpoint of the path $\mathbb{q}_2$. By Definition 2.7, we have:

1. L and L' are the last two LPAs of $\mathbb{p} = \mathbb{q}_1\mathbb{r}\mathbb{q}_2$.

It follows from statement 1 that

2. the path $\psi^{\mathcal{Y}}(\mathbb{p}) = \mathbb{q}_1\mathbb{q}_2\mathbb{r}$ splits at its last non-terminal boundary point into the paths $\mathbb{q}_1\mathbb{q}_2$ and $\mathbb{r}$.

The LPAs L and L' of $\mathbb{p}$ combine to form a single point in $\psi^{\mathcal{Y}}(\mathbb{p}) = \mathbb{q}_1\mathbb{q}_2\mathbb{r}$, and so:

3. the subpaths $\mathbb{q}_1$ and $\mathbb{q}_2$ of $\psi^{\mathcal{Y}}(\mathbb{p})$ collectively contain exactly one fewer flaw than the subpaths $\mathbb{q}_1$ and $\mathbb{q}_2$ of $\mathbb{p}$.
4. the path $\mathbb{q}_1\mathbb{q}_2$ has at least one LPA, namely the point $L = L'$.

By statement 1 and Definition 2.6*(ii)*, the subpath $\mathbb{r}$ contains no boundary points of $\mathbb{p}$. This, together with statement 4, implies:

5. the elevation of the LPAs of the path $\mathbb{q}_1\mathbb{q}_2$ is smaller than the magnitude of the elevation of the HPBs of the path $\mathbb{r}$ (if any).

It follows from statement 5 that:

6. disregarding its startpoint and endpoint, the subpath $\mathbb{r}$ of $\psi^{\mathcal{Y}}(\mathbb{p}) = \mathbb{q}_1\mathbb{q}_2\mathbb{r}$ contains the same number of flaws as the subpath $\mathbb{r}$ of $\mathbb{p}$.

By statements 3 and 6, the path $\psi^{\mathcal{Y}}(\mathbb{p})$ contains exactly one fewer flaw than $\mathbb{p}$. By statement 4, the path $\mathbb{q}_1\mathbb{q}_2$ has at least one flaw. Together with statements 2 and 5, this shows by Definition 2.4 that $\psi^{\mathcal{Y}}(\mathbb{p}) \in Y$. $\square$

Remark 2.16. Continue with the notation from the proof of Proposition 2.15. We note for use in Section 2.4 that statement 1 implies the endpoint L of $\mathbb{q}_1$ is the last LPA of the path $\mathbb{q}_1\mathbb{q}_2$. $\lrcorner$

2.4 The maps ϕ and ψ are injective

We complete the proof of Theorem 1.5 by showing in turn that each of the maps ϕ^X, ϕ^Y, $\psi^{\mathcal{X}}$, $\psi^{\mathcal{Y}}$ is injective. We give the proof for ϕ^X and ϕ^Y in detail, and for $\psi^{\mathcal{X}}$ and $\psi^{\mathcal{Y}}$ in abbreviated form.

The map ϕ^X is injective:

Let $p, p' \in X$, and write $p = qr_1r_2$ and $p' = q'r_1'r_2'$ according to Definition 2.5. We suppose that $\phi^X(p) = \phi^X(p')$, and wish to show that $p = p'$.

By statement 4 in the proof of Proposition 2.9, the endpoint L of r_2 is the last LPA of $\phi^X(p) = qr_2r_1$. By Remark 2.10, the startpoint H of r_2 is the boundary point of $\phi^X(p) = qr_2r_1$ immediately preceding L.

Therefore $\phi^X(p) = qr_2r_1$ splits into qr_2 and r_1 at the last LPA L of $\phi^X(p)$, and the subpath qr_2 splits into q and r_2 at the boundary point of $\phi^X(p)$ immediately preceding L. The corresponding statement holds for $\phi^X(p')$. Since $\phi^X(p)$ and $\phi^X(p')$ are equal by assumption, their LPAs and boundary points are identical. Therefore $q = q'$ and $r_2 = r_2'$ and $r_1 = r_1'$ and so $p = qr_1r_2 = q'r_1'r_2' = p'$, as required.

The map ϕ^Y is injective:

Let $p, p' \in Y$, and write $p = q_1q_2r$ and $p' = q_1'q_2'r'$ according to Definition 2.5. We suppose that $\phi^Y(p) = \phi^Y(p')$, and wish to show that $p = p'$.

By statement 2 in the proof of Proposition 2.11, the endpoint L of q_1 and the startpoint L' of q_2 are the last two LPAs of $\phi^Y(p) = q_1rq_2$, and so $\phi^Y(p)$ splits at its last two LPAs into q_1 and r and q_2. Likewise, $\phi^Y(p')$ splits at its last two LPAs into q_1' and r' and q_2'. But $\phi^Y(p)$

and $\phi^Y(p')$ are equal by assumption, so their last two LPAs are identical. Therefore $q_1 = q_1'$ and $r = r'$ and $q_2 = q_2'$ and so $p = q_1q_2r = q_1'q_2'r' = p'$, as required.

The map $\psi^{\mathcal{X}}$ is injective:

Let $\mathbb{p}, \mathbb{p}' \in \mathcal{X}$, and write $\mathbb{p} = \mathbb{q}\mathbb{r}_2\mathbb{r}_1$ and $\mathbb{p}' = \mathbb{q}'\mathbb{r}_2'\mathbb{r}_1'$ according to Definition 2.7. We suppose that $\psi^{\mathcal{X}}(\mathbb{p}) = \psi^{\mathcal{X}}(\mathbb{p}')$, and wish to show that $\mathbb{p} = \mathbb{p}'$.

By statement 4 in the proof of Proposition 2.13, the path $\psi^{\mathcal{X}}(\mathbb{p}) = \mathbb{q}\mathbb{r}_1\mathbb{r}_2$ splits at its last non-terminal boundary point into $\mathbb{q}$ and $\mathbb{r}_1\mathbb{r}_2$. By Remark 2.14 the path $\mathbb{r}_1\mathbb{r}_2$ splits at its last HPB into $\mathbb{r}_1$ and $\mathbb{r}_2$.

It follows that $\mathbb{p} = \mathbb{q}\mathbb{r}_2\mathbb{r}_1 = \mathbb{q}'\mathbb{r}_2'\mathbb{r}_1' = \mathbb{p}'$, as required.

The map $\psi^{\mathcal{Y}}$ is injective:

Let $\mathbb{p}, \mathbb{p}' \in \mathcal{Y}$, and write $\mathbb{p} = \mathbb{q}_1\mathbb{r}\mathbb{q}_2$ and $\mathbb{p}' = \mathbb{q}_1'\mathbb{r}'\mathbb{q}_2'$ according to Definition 2.7. We suppose that $\psi^{\mathcal{Y}}(\mathbb{p}) = \psi^{\mathcal{Y}}(\mathbb{p}')$, and wish to show that $\mathbb{p} = \mathbb{p}'$.

By statement 2 in the proof of Proposition 2.15, the path $\psi^{\mathcal{Y}}(\mathbb{p}) = \mathbb{q}_1\mathbb{q}_2\mathbb{r}$ splits at its last non-terminal boundary point into $\mathbb{q}_1\mathbb{q}_2$ and $\mathbb{r}$. By Remark 2.16, the path $\mathbb{q}_1\mathbb{q}_2$ splits at its last LPA into $\mathbb{q}_1$ and $\mathbb{q}_2$.

It follows that $\mathbb{p} = \mathbb{q}_1\mathbb{r}\mathbb{q}_2 = \mathbb{q}_1'\mathbb{r}'\mathbb{q}_2' = \mathbb{p}'$, as required.

2.5 The special case $g = 1$ (Theorem 1.14)

We finally re-examine the special case $g = 1$ (Theorem 1.14), involving paths from $(0, 0)$ to (a, b), to show how the proof of Theorem 1.5 described in Section 2 simplifies significantly. In doing so, we shall obtain a simple self-contained proof of Theorem 1.14.

Let k satisfy $0 \leq k < a + b - 1$ and let p be a path in $N_k(1)$. Since

1. the path p cannot have any interior boundary points,

it follows that:

2. the subset $S_k(1)$ of $N_k(1)$ is empty by Definition 1.4,
3. the subset Y of $N_k(1)$ is empty by Definition 2.4.

Since $g = 1$, the path p has at most one LPA and so

4. the subset $\mathcal{Y}$ of $N_{k+1}(1)$ is empty by Definition 2.6.

By reference to Figure 2.1, statements 2, 3 and 4 show that $N_k(1) = X$ and $N_{k+1}(1) = \mathcal{X}$. We shall show that ϕ^X and ψ^X are inverse maps, so that $|N_k(1)| = |N_{k+1}(1)|$. We may then conclude that $|N_0(1)| = |N_1(1)| = \cdots = |N_{a+b-1}(1)|$, which gives Theorem 1.14 because the total number of paths from $(0, 0)$ to (a, b) is $\binom{a+b}{a}$.

It remains to show that ϕ^X and ψ^X are inverse maps. Write $p \in X$ as its canonical representation $p = qr_1r_2$ according to Definition 2.5, where p has at least one HPB. By statement 1, we have that q is empty and so $p = r_1r_2$. Since $g = 1$, the path p has at most one HPB. Therefore by Definition 2.8 the map ϕ^X splits p at its unique HPB into r_1r_2 and replaces it by r_2r_1. That is, ϕ^X cyclically permutes the steps of p by bringing the unique HPB to the origin.

Similarly, write $\mathbb{p} \in \mathcal{X}$ as its canonical representation $\mathbb{p} = \mathbb{q}\mathbb{r}_2\mathbb{r}_1$ according to Definition 2.7, where $\mathbb{q}$ is empty by statement 1. By Definition 2.12, the map ψ^X splits $\mathbb{p}$ at its unique LPA into $\mathbb{r}_2\mathbb{r}_1$ and replaces it by $\mathbb{r}_1\mathbb{r}_2$. That is, ψ^X cyclically permutes the steps of p by bringing the unique LPA to the origin.

Comparison of the descriptions of ϕ^X and ψ^X shows that they are inverse maps, as required.

3 Conclusion

Our central objective was to find an explicit formula for $|N_k(g)|$, the number of simple lattice paths from $(0,0)$ to (ga, gb) having exactly k lattice points lying strictly above the linear boundary joining the startpoint to the endpoint. This is given by the closed form expression in Theorem 1.12, using the definition (1.3) of $\mu_j(g)$.

We conclude with two open problems for future study.

1. Evaluating $|N_k(g)|$ via the path enumeration formula Theorem 1.12 involves a sum over integer partitions of g, and is therefore computationally intensive. In the special case $a = b = 1$, Theorem 1.15 provides a significantly simpler expression than does Theorem 1.12. Is there a closed form expression for $|N_k(g)|$ that is simpler than Theorem 1.12 for other special cases of a, b (or in general)?

2. We established the path enumeration formula by solving the recurrence relation given in Corollary 1.7 and making use of the known values stated in Corollary 1.11. These known values are in turn predicated on Theorem 1.10, which was proved by Bizley using generating functions [2]. Is there a direct combinatorial proof of Theorem 1.10?

4 Acknowledgments

The first and third authors collaborated with Takudzwa Marwendo in 2019 on a preliminary investigation of the material of this paper, resulting in Theorems 1.14 and 1.15, the results of various numerical experiments, and the conjecture that property P1 (constant on blocks) holds. This investigation was a major inspiration for this paper. We are grateful to Taku for his many contributions and helpful conversations.

References

[1] M. Aigner. *A Course in Enumeration*, volume 238 of *Graduate Texts in Mathematics*. Springer, Berlin, 2007.

[2] M.T.L. Bizley. Derivation of a new formula for the number of minimal lattice paths from $(0,0)$ to (km, kn) having just t contacts with the line $my = nx$ and having no points above this line; and a proof of Grossman's formula for the number of paths which may touch but do not rise above this line. *J. Inst. Actuar.*, 80:55–62, 1954.

[3] D. Callan. Pair them up! A visual approach to the Chung-Feller theorem. *Coll. Math. J.*, 26(3):196–198, 1995.

[4] Y.-M. Chen. The Chung-Feller theorem revisited. *Discrete Math.*, 308(7):1328–1329, 2008.

[5] K.L. Chung and W. Feller. On fluctuations in coin-tossing. *Proc. Natl. Acad. Sci. U.S.A.*, 35(10):605–608, 1949.

[6] F. Firoozi. Enumeration of lattice paths with respect to a linear boundary. Master's thesis, Simon Fraser University, 2023.

[7] H.D. Grossman. Paths in a lattice triangle. *Scr. Math.*, 16:207, 1950.

[8] V.J.W. Guo and X.-X. Wang. A Chung-Feller theorem for lattice paths with respect to cyclically shifting boundaries. *J. Algebraic Combin.*, 50(2):119–126, 2019.

[9] A. Huq. *Generalized Chung-Feller Theorems for Lattice Paths*. PhD thesis, Brandeis University, 2009.

[10] J. Ma and Y.-N. Yeh. Generalizations of Chung-Feller theorems. *Bull. Inst. Math. Acad. Sin. (N.S.)*, 4(3):299–332, 2009.

[11] I.G. MacDonald. *Symmetric Functions and Hall Polynomials* (2 ed.). Oxford University Press, 1979.

[12] J.O. Tirrell. *Orthogonal polynomials, lattice paths, and skew Young tableaux*. ProQuest LLC, Ann Arbor, MI, 2016. Thesis (Ph.D.)–Brandeis University.

[13] S.H.F. Yan and Y. Zhang. On lattice paths with four types of steps. *Graphs Combin.*, 31(4):1077–1084, 2015.