Counting lattice paths

An extended abstract of the PhD dissertation

Maciej Dziemiańczuk

A *lattice path* (or simply a path) is a finite sequence of points $p_0, p_1, \ldots, p_n$ in $\mathbb{Z} \times \mathbb{Z}$. A *step* of the path is the difference between two of its consecutive points, i.e., $p_i - p_{i-1}$. The lattice path can also be represented by the initial point p_0 and the sequence of its steps $s_1, s_2, \ldots, s_n$, which uniquely determine the remaining points of the path. For instance, the path from Figure 1 is

$$\big((0,0),(1,3),(2,1),(3,1),(4,0),(5,-1),(6,1),(7,1),(8,0)\big),$$

whose step representation is the initial point $(0,0)$ and the following sequence of steps:

$$\big((1,3),(1,-2),(1,0),(1,-1),(1,-1),(1,2),(1,0),(1,-1)\big).$$

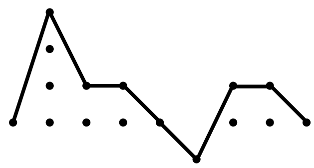

FIGURE 1: A lattice path running from $(0,0)$ to $(8,0)$.

The literature on lattice paths is very rich. Humphreys [13] refers to more than two hundred crucial articles. Most of them are related to path counting problems and relationships with other structures.

Some of the most well-known families of lattice paths are those that consist of two types of steps: $(1,1)$ and $(1,-1)$. These paths are called *Dyck paths*. In 1878, Whitworth [22] used them to describe various combinatorial problems. In 1887, Bertrand [3] formulated the famous ballot problem, which can be translated into a question about the number of Dyck paths running from $(0,0)$ to $(u+d, u-d)$, where $u > d$, and that do not touch the x-axis except at the initial point. André [1] solved this problem and showed that the number of such paths is equal to $\binom{u+d-1}{u}(u-d)/d$. Therefore, the number of Dyck paths running from $(0,0)$ to $(2n,0)$ and that never go below the x-axis is equal to the Catalan number $C_n = \binom{2n}{n}/(n+1)$.

Simple generalizations of Dyck paths are *Motzkin paths* [17], which consist of three types of steps: $(1,1)$, $(1,0)$, and $(1,-1)$. Donaghey and Shapiro [7]) provided a representative selection of 14 situations wherein Motzkin numbers occur in connection with other combinatorial structures such as Dyck paths, sequences of parentheses, trees with loops, and bipartite graphs. The number of Motzkin paths running from $(0,0)$ to $(n,0)$ that do not go below the x-axis is called the nth *Motzkin number* and it is equal to $\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} C_k$.

Łukasiewicz paths [11, Sec. I.5.3], named after the Polish logician Jan Łukasiewicz (1878-1956), represent another generalization of Dyck paths. A *Łukasiewicz path* is a lattice path in which the set of allowable steps contains all steps of the form $(1,k)$ for $k \geq -1$. The number of Łukasiewicz paths running from $(0,0)$ to $(n,0)$ that never go below the x-axis is equal to the Catalan number C_n (see Flajolet and Sedgewick [11]). These paths are useful in the analysis of algorithms because of their close relationship with plane trees, which will be discussed in Chapter 3 of this thesis.

Weighted (or colored) paths are used in several applications. Weighted Motzkin paths were studied by, among others, Deutsch and Shapiro [6], Chen et al. [4, 21]. They established several connections between weighted Motzkin paths, Schröder paths, partitions of sets, and other graph counting problems. Weighted Łukasiewicz paths were studied by Varvak [20]. She showed that there are several one-to-one correspondences between appropriate weighted Łukasiewicz paths and multipermutations, partitions of sets, idempotent functions, and trees. The main tool that she used was a connection between weighted Łukasiewicz paths and continued fractions (see also Flajolet [10]). Hennessy [12, Sec. 5.3] showed that there are bijections between certain families of Łukasiewicz, Schröder, and Motzkin paths.

The structure that generalizes all the above-mentioned paths is lattice paths that consist of steps $(1,k)$ for any $k \in \mathbb{Z}$. They are called *simple directed paths* (see, e.g., Banderier and Flajolet [2]). They are used as models in the analysis of algorithms and dynamic data structures [18]. A unified approach to simple directed paths was developed by Banderier and Flajolet [2]. They showed that the counting generating functions of such paths are algebraic functions. They also described the asymptotic behavior of these numbers using the method of singularity analysis.

In Chapter 3, we use lattice paths to study the combinatorial and statistical properties of plane multitrees. Plane trees are well-known structures in combinatorics [5, 6, 14, 16]. There are several equivalent definitions of plane trees, but they are mostly defined as unlabeled rooted trees in which the order of sons is significant. For instance, all plane trees with four nodes are given in Figure 2.

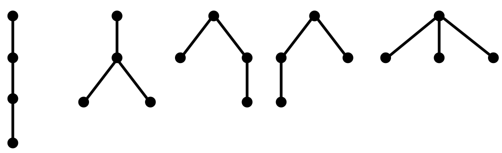

FIGURE 2: All plane trees with four nodes.

There is a bijection between the family of all plane trees with n nodes and the set of Łukasiewicz paths running from $(0,0)$ to $(n-1,0)$ that do not go below the x-axis (see Flajolet and Sedgewick [11, Sec. I.5.1]). This bijection implies that the number of plane trees with n nodes is equal to the Catalan number C_{n-1}. It is also well known that the number of plane trees with n nodes and k leaves is equal to the Narayana number $\binom{n-1}{k}\binom{n-1}{k-1}/(n-1)$ (see Dershowitz and Zaks [5]). Dershowitz and Zaks [5] also showed that the expected number of leaves in a plane tree with n nodes is $n/2$ and that the expected number of outcoming edges from the root in this tree is $3(n-1)/(n+1)$. Plane trees are used in computer science due to their close relationship with Łukasiewicz codes and polish prefix notation (see, e.g., Sedgewick i Flajolet [11, Sec. I.5.3]). Computer compilers use such trees as structures to parse expressions (see Knuth [15, Sec. 2.3]).

A plane tree in which every edge has assigned a positive integer, called *weight*, is called the *plane multitree*. The first mention of such trees is credited to R. Bacher (see the description of the sequence A002212 in OEIS [19]). He showed that the number of plane multitrees with n edges is equal to $\sum_{k=0}^{n} C_k \binom{n-1}{k-1}$. A plane multitree, in which outdegree (the sum of weights of outgoing edges from a node to its sons) of every internel node is not greater than N will be called the *N-ary plane multitree*. For instance, all 2-ary plane multitrees with four nodes are given in Figure 3, where the weighted edges are represented by multiple edges.

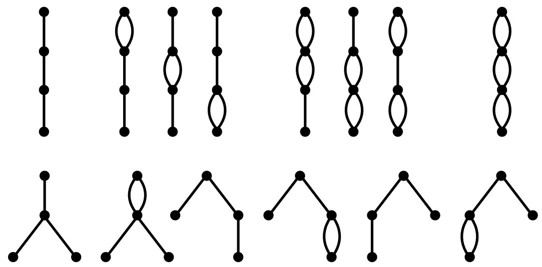

FIGURE 3: All 2-ary plane multitrees with four nodes.

Basic definitions

Throughout the thesis, we will consider lattice paths that consist of steps $V = (0, -1)$ and $S_k = (1, k)$ for $k \in \mathbb{Z}$. Let $N \geq 0$ and $\Sigma \subseteq \{V, S_N, S_{N-1}, \ldots\}$, then a Σ-*path* is a finite sequence of points $p_0, p_1, \ldots, p_n$ in $\mathbb{Z} \times \mathbb{Z}$ such that $p_i - p_{i-1} \in \Sigma$ for $i \in \{1, \ldots, n\}$. We will consider the following three types of paths:

- An *m-primary Σ-path* is a Σ-path running from $(0,0)$ to some $(n, -m)$, with $n \geq 0$ and $m \geq 0$, whose all points, except the possibly last one, lie on or above the horizontal axis. We will denote by $\mathcal{P}_\Sigma(n, -m)$ the family of all m-primary Σ-paths running from $(0,0)$ to $(n, -m)$.
- A *free* Σ-path is a Σ-path running from $(0,0)$ to some $(n, -m)$ with $n \geq 1$, $m \in \mathbb{Z}$. We will denote by $\mathcal{F}_\Sigma(n, -m)$ the family of all free Σ-paths running from $(0,0)$ to $(n, -m)$.
- An *N-Raney path* of length n is a lattice path running from $(0,1)$ to $(n, 0)$ in which the set of allowable steps is $\{S_N, S_{N-1}, \ldots\}$ for $N \geq 0$, and only the ending point of the path lies below the line $y = 1$. We will denote by $\mathcal{R}_N(n)$ the family of all N-Raney paths of length n.

Results of the first part of the thesis

The first part of the thesis (Chapter 2) is devoted to the study of two families of primary and free Λ-paths, where Λ is an arbitrary set of lattice steps satisfying $\Lambda \subseteq \{V, S_N, S_{N-1}, \ldots\}$ and $S_N, V \in \Lambda$ for any fixed $N \geq 0$. The results of this part originate from the paper [9].

The main result of this part of the thesis is to show (Theorem 2.10) that there is a bijection between the family $\mathcal{P}_\Lambda(n, -m)$ of m-primary Λ-paths and the family of properly weighted m-primary Γ-paths from $\mathcal{P}_\Gamma(n, -m)$, where $\Gamma = (\Lambda \setminus \{V\}) \cup \{S_N, S_{N-1}, \ldots, S_{-1}\}$. This implies that the additional vertical steps V in the paths of $\mathcal{P}_\Lambda(n, -m)$ can be coded using the weights of steps in paths without V. From the combinatorial point of view, Γ-paths from $\mathcal{P}_\Gamma(n, -m)$ have a simpler structure than do the paths in $\mathcal{P}_\Lambda(n, -m)$. Recall that the paths of $\mathcal{P}_\Gamma(n, -m)$ are simple directed paths, and note that they are essentially one-dimensional objects. There are many results for simple directed paths that have already been described in the literature. The classical work here is the paper of Banderier and Flajolet [2]. Therefore, using the bijection mentioned above, we can apply some of these results to Λ-paths. For instance, simple directed paths can be easily decomposed into shorter subpaths, and this decomposition property provides a straightforward method of calculating generating functions that count these paths. In Section 2.6, we derive a functional equation for the generating function $P_{\Lambda,m}(x) = \sum_{n \geq 0} |\mathcal{P}_\Lambda(n, -m)| x^n$. We show (Theorem 2.27) that for $m \geq 1$,

$$P_{\Lambda,0}(x) = 1 + \delta_{\Lambda,0} x P_{\Lambda,0}(x) + x P_{\Lambda,0}(x) \sum_{k=1}^{N} \sum_{d=1}^{k} |\mathcal{H}_\Lambda(0, d, k)| \sum_{M} \prod_{j=1}^{d} P_{\Lambda,m_j}(x),$$

$$P_{\Lambda,m}(x) = \delta_{\Lambda,m} x + x \sum_{k=0}^{N} \sum_{d=1}^{k+1} |\mathcal{H}_\Lambda(m, d, k)| \sum_{M} \prod_{j=1}^{d} P_{\Lambda,m_j}(x),$$

for some constants $\delta_{\Lambda,m}$ and $|\mathcal{H}_\Lambda(m, d, k)|$ depending on Λ.

In Section 2.4, we establish the following connections between 1-primary Λ-paths and free Λ-paths. We prove (Theorem 2.17) that for $n \geq 1$, we have

$$|\mathcal{P}_\Lambda(n, -1)| = \frac{1}{n} \Big(\big|\mathcal{F}_\Lambda(n, -1)\big| - \big|\mathcal{F}_\Lambda(n, 0)\big| \Big).$$

Further, we show (Theorem 2.19 and 2.20) that for $n \geq 1$,

$$\begin{aligned} \#\mathit{Steps}(V \in \mathcal{P}_\Lambda(n, -1)) &= |\mathcal{F}_\Lambda(n, 0)|, \\ \#\mathit{Steps}(S_k \in \mathcal{P}_\Lambda(n, -1)) &= |\mathcal{F}_\Lambda(n-1, -k-1)|, \qquad (S_k \in \Lambda), \\ \#\mathit{Steps}(\mathcal{P}_\Lambda(n, -1)) &= |\mathcal{F}_\Lambda(n, -1)|, \end{aligned}$$

where $\#\mathit{Steps}(S \in \mathcal{P}_\Sigma(n, -1))$ denotes the number of all occurrences of the step S in all paths of $\mathcal{P}_\Sigma(n, -1)$ and $\#\mathit{Steps}(\mathcal{P}_\Sigma(n, -1))$ denotes the number of all steps in all paths of $\mathcal{P}_\Sigma(n, -1)$.

In Section 2.5, we provide various results for Λ-paths. We show (Theorem 2.22) that the numbers of free and 1-primary Λ-paths are given by the following formulas:

$$|\mathcal{F}_\Lambda(n,m)| = [x^{Nn-m}]\frac{1}{(1-x)^{n+1}}\Big(\sum_{S_k\in\Lambda} x^{N-k}\Big)^n, \qquad (n\geq 1, m\in\mathbb{Z}),$$

$$|\mathcal{P}_\Lambda(n,-1)| = \frac{1}{n}[x^{Nn+1}]\frac{1}{(1-x)^{n}}\Big(\sum_{S_k\in\Lambda} x^{N-k}\Big)^n, \qquad (n\geq 1).$$

We also derive certain statistical properties of Λ-paths. Any Γ-path running from $(0,0)$ to (n,m) has exactly n steps. The number of steps in a Λ-path running between the same points is equal to or greater than n. We show (Corollary 2.23) that the expected number of steps in a 1-primary Λ-path running from $(0,0)$ to $(n,-1)$ is equal to

$$n\frac{|\mathcal{F}_\Lambda(n,-1)|}{|\mathcal{F}_\Lambda(n,-1)|-|\mathcal{F}_\Lambda(n,0)|} \qquad (n\geq 1).$$

In Sections 2.7 – 2.9, we consider three cases for the set of steps Λ. These examples are connected with the well-known families of lattice paths from the literature. We apply to them the general results from the previous sections and see that several of the examples take on a simple form. Namely, for fixed $N, K\geq 0$, we consider

- $\Lambda_1 = \{V, S_N, S_{N-1}, \dots, S_{-1}\}$ and $\Gamma_1 = \{S_N, S_{N-1}, \dots, S_{-1}\}$ (Łukasiewicz paths),
- $\Lambda_2 = \{V, S_N, S_{N-1}, \dots\}$ and $\Gamma_2 = \{S_N, S_{N-1}, \dots\}$ (Raney paths),
- $\Lambda_3 = \{V, S_N, S_{-K}\}$ (generalized Dyck paths) and $\Gamma_3 = \{S_N, S_{N-1}, \dots, S_{-1}, S_{-K}\}$.

In particular, we show (Theorem 2.30), that for $m\in\{0,1\}$, the equation for the generating function $P_{\Lambda_1,m}(x)$ that counts paths in $\mathcal{P}_{\Lambda_1}(n,-m)$ according to n is

$$P_{\Lambda_1,0}(x) = 1 + xP_{\Lambda_1,0}(x)\sum_{k=0}^{N}(1+P_{\Lambda_1,1}(x))^k,$$

$$P_{\Lambda_1,1}(x) = x\sum_{k=0}^{N+1}(1+P_{\Lambda_1,1}(x))^k.$$

We show (Theorem 2.34), that for $n\geq 1$, we have

$$|\mathcal{P}_{\Lambda_1}(n,0)| = (-1)^n + \sum_{j=1}^{n}\sum_{k=0}^{\lfloor\frac{Nj+1}{N+2}\rfloor}\frac{(-1)^{k+n-j}}{j}\binom{j}{k}\binom{(N+2)(j-k)}{2j-1},$$

$$|\mathcal{P}_{\Lambda_1}(n,-1)| = \frac{1}{n}\sum_{k=0}^{\lfloor\frac{Nn+1}{N+2}\rfloor}(-1)^k\binom{n}{k}\binom{(N+2)(n-k)}{2n-1}.$$

In Section 2.8 we show (Theorem 2.37), that

$$|\mathcal{P}_{\Lambda_2}(n,-1)| = \frac{1}{n}\binom{(N+2)n}{2n-1} \qquad (n\geq 1).$$

We show that the expected number of vertical steps V in a path in $\mathcal{P}_{\Lambda_2}(n,-1)$ is equal to $(Nn+1)/2$ and that the expected number of all steps in a path in $\mathcal{P}_{\Lambda_2}(n,-1)$ is equal to $((N+2)n+1)/2$. In Section 2.9, we show (Theorem 2.40) that

$$|\mathcal{P}_{\Lambda_3}(n,-1)| = \frac{1}{n}\sum_{k=0}^{\lfloor\frac{Nn+1}{N+K}\rfloor}\binom{n}{k}\binom{n(N+1)-k(N+K)}{n-1} \qquad (n\geq 1).$$

Results of the second part of the thesis

The second part of the thesis (Chapter 3) is devoted to the study of the combinatorial and statistical properties of plane multitrees. The results of this part originate from the paper [8]. The main tool that will be used in this chapter is a bijection between plane multitrees and Raney paths. Namely, in Section 3.2, we show (Theorem 3.8) that for all $N\geq 0$ and $n\geq 1$, there is a bijection between the family $\mathcal{R}_N(n)$ of N-Raney paths of length n and the family $\mathcal{T}_{N+1}(n)$ of $(N+1)$-ary plane multitrees with n nodes. From the combinatorial view of point, Raney paths have a simpler structure than do plane multitrees. Therefore, the obtained results for Raney paths will be translated to the corresponding properties of plane multitrees.

In Section 3.3, we define a bijection between N-Raney paths of length n and the so-called $(N-1,n,1)$-Raney sequences (Lemma 3.23). Using this bijection, we show (Theorem 3.24) that the number of N-ary plane multitrees with n nodes is equal to

$$|\mathcal{T}_N(n)| = \frac{1}{n}\binom{Nn}{n-1} \qquad (N\geq 1, n\geq 1).$$

In Sections 3.4 – 3.7, we use the two above-mentioned bijections to obtain certain results for plane multitrees. Namely, for the family $\mathcal{T}_N(n)$ of all N-ary plane multitrees with n nodes, we consider the following numbers:

1. $L_N(n,k)$ = number of such trees with exactly k leaves,
2. $E_N(n,k)$ = number of such trees with exactly k edges,
3. $G_N(n,d)$ = number of such trees in which the root has d outgoing edges, and
4. $M_N(n,d)$ = total number of nodes that have d outgoing edges in all such trees.

We show that

$$L_N(n,k) = \frac{1}{n}\binom{n}{k}\sum_{s=0}^{n-k}(-1)^s\binom{n-k}{s}\binom{N(n-k-s)}{n-1}, \qquad \text{(Th. 3.26)}$$

$$E_N(n,k) = \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=0}^{i}(-1)^j\binom{n}{i}\binom{i}{j}\binom{k-i}{n-i-1}\binom{k-jN-1}{i-1}, \qquad \text{(Th. 3.28)}$$

$$G_N(n,d) = \frac{N-1+d}{(N-1)(n-1)+d}\binom{N(n-1)+d-2}{n-2}, \qquad \text{(Th. 3.30)}$$

$$M_N(n,d) = \binom{N(n-1)+d-1+0^d}{n-2+0^d}, \qquad \text{(Th. 3.31)}$$

where $0^d = 0$ for $d > 0$, and $0^0 = 1$.

In Section 3.7, we study the statistical properties of plane multitrees. We prove (Theorem 3.34) that

$$\lim_{N\to\infty} \lim_{n\to\infty} \frac{M_N(n,0)}{nT_N(n)} = \lim_{n\to\infty} \lim_{N\to\infty} \frac{M_N(n,0)}{nT_N(n)} = \frac{1}{e}.$$

References

[1] D. André. Solution directe du problème résolu par M. Bertrand. *Comptes Rendus de l'Académie des Sciences*, 105:436–437, 1887. Paris.

[2] C. Banderier and P. Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science*, 281:37–80, 2002.

[3] J. Bertrand. Solution d'un probleme. *Comptes Rendus de l'Academie des Science Paris 105*, page 369, 1887.

[4] W. Y. C. Chen, N. Y. Li, Louis W. Shapiro, and S. H. F. Yan. Matrix identities on weighted partial Motzkin paths. *European Journal of Combinatorics*, 28:1196–1207, 2007.

[5] N. Dershowitz and S. Zaks. Enumerations of ordered trees. *Discrete Mathematics*, 31(1):9–28, 1980.

[6] E. Deutsch and L. W. Shapiro. A bijection between ordered trees and 2-Motzkin paths and its many consequences. *Discrete Mathematics*, 256:655–670, 2002.

[7] R. Donaghey and L. W. Shapiro. Motzkin numbers. *Journal of Combinatorial Theory Series A*, 23:291–301, 1977.

[8] M. Dziemiańczuk. Enumerations of plane trees with multiple edges and raney lattice paths. *Discrete Mathematics*, 337:9–24, 28 Dec. 2014.

[9] M. Dziemiańczuk. On directed lattice paths with additional vertical steps. Submitted for publication, arXiv:1410.5747, 2014.

[10] P. Flajolet. Combinatorial aspects of continued fractions. *Discrete Mathematics*, 32:125–161, 1980.

[11] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[12] A. Hennessy. *A Study of Riordan Arrays with Applications to Continued Fractions, Orthogonal Polynomials and Lattice Paths*. PhD thesis, Waterford Institute of Technology, 2011.

[13] K. Humphreys. A history and a survey of lattice path enumeration. *Journal of Statistical Planning and Inference*, 140(8):2237–2254, 2010.

[14] D. A. Klarner. Correspondences between plane trees and binary sequences. *Journal of Combinatorial Theory*, 9:401–411, 1970.

[15] D. E. Knuth. *The Art of Computer Programming*, volume vol. 1: Fundamental Algorithms. Addison-Wesley, third edition, 1997.

[16] W. Młotkowski and K. A. Penson. The probability measure coresponding to 2–plane trees. *Probability and Mathematical Statistics*, 33:255–264, 2013.

[17] T. Motzkin. Relations between hypersurface cross ratios, and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products. *Bulletin of the American Mathematical Society*, 54(4):352–360, 1948.

[18] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms*. Addison-Wesley Publishing Company, 1996.

[19] N. J. A. Sloane. The on-line encyclopedia of integer sequences. Published electronically at `http://oeis.org`.

[20] A. L. Varvak. *Encoding Properties of Lattice Paths*. PhD thesis, Brandeis University, 2004.

[21] C. J. Wang W. Y.C. Chen. Noncrossing linked partitions and large (3,2)-Motzkin paths. *Discrete Mathematics*, 312(11):1918–1922, 2012.

[22] W. A. Whitworth. Arrangements of m things of one sort and n things of another sort, under certain conditions of priority. *Messenger of Math.*, 8:105–114, 1879.