Secondary structures, plane trees and Motzkin numbers

TOMISLAV DOŠLIĆ* AND DARKO VELJAN†

Abstract. *A bijective correspondence is established between secondary structures of a given rank and size and plane trees satisfying certain additional conditions. The correspondence is then used to obtain new combinatorial interpretations of Motzkin numbers in terms of plane trees and Dyck paths.*

Key words: *Motzkin numbers, plane trees, secondary structures, Dyck paths*

AMS subject classifications: 05A15, 05C05, 05C30

Received May 17, 2006 Accepted May 19, 2007

1. Introduction and preliminaries

The sequence of Catalan numbers, C_n, is one of the most ubiquitous sequences in combinatorics; more than one hundred sixty combinatorial interpretations are listed in Exercise 6.19 of [17], along with 11 algebraic interpretations of Exercise 6.25. (Here we counted also the interpretations given in "Catalan addendum", [18].) In a similar way, 13 instances of Motzkin numbers are given in Exercise 6.38 of the same book. The aim of this paper is to present a couple of combinatorial interpretations of Motzkin numbers that do not appear in [17] and that originate in the context of secondary structures.

Secondary structures are mathematical objects obtained by abstracting topologically non-relevant properties of planar foldings of single-stranded nucleic acids. There are many papers concerning their combinatorial properties (see, e.g. [20], [13], [14]) and their enumerating sequences ([19], [11], [12]).

We list some basic definitions and relations that will be used in subsequent sections.

The n-th **Catalan number**, C_n, $n \geq 0$, is defined by $C_n = \frac{1}{n+1}\binom{2n}{n}$. The well-known recurrence relation $C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k}$, $C_0 = 1$, is frequently used as an alternative definition of the Catalan sequence.

*Faculty of Civil Engineering, University of Zagreb, Kačićeva 26, HR-10 000 Zagreb, Croatia, e-mail: `doslic@grad.hr`

†Department of Mathematics, University of Zagreb, Bijenička 30, HR-10 000 Zagreb,

Perhaps the best known combinatorial objects enumerated by Catalan numbers are Dyck paths. A **Dyck path** (or a "mountain path") on $2n$ steps is a lattice path in the (x, y)-coordinate plane from $(0,0)$ to $(2n, 0)$ with steps $(1,1)$ (Up) and $(1,-1)$ ($Down$) never falling below the x-axis. The set of all Dyck paths on $2n$ steps is denoted by $\mathcal{D}(n)$.

A **slope** of a Dyck path is any sequence of consecutive steps of the same type. The number of steps in a slope is its length. A **long slope** is a slope of length at least three. A **peak** of a Dyck path is a place where an Up step is immediately followed by a $Down$ step. If both slopes that meet at a peak are longer than one, we say that the peak is **lonely**. The set of all Dyck paths on $2n$ steps with exactly k peaks is denoted by $\mathcal{D}_k(n)$. The cardinality of this set is given by

$$|\mathcal{D}_k(n)| = N(n,k) = \frac{1}{n}\binom{n}{k}\binom{n}{k-1} = \frac{1}{k}\binom{n}{k-1}\binom{n-1}{k-1}.$$

The numbers $N(n,k)$, $n,k \geq 1$, are called **Narayana numbers**. Obviously, $\sum_{k=1}^{n} N(n,k) = C_n$.

The Motzkin numbers, M_n, satisfy a convolutional recurrence, $M_{n+2} = M_{n+1} + \sum_{k=0}^{n} M_k M_{n-k}$, with $M_0 = M_1 = 1$. The most popular combinatorial objects counted by Motzkin numbers are Motzkin paths. A **Motzkin path** on n steps is a lattice path in the (x,y)-coordinate plane from $(0,0)$ to $(n,0)$ using only steps of the type $(1,1)$ (Up), $(1,-1)$ ($Down$), and $(1,0)$ ($Level$), never falling below the x-axis. A classical reference for Motzkin numbers is [10].

The two sequences are connected by the well-known relations (see, e.g. [1])

$$M_n = \sum_{k=0}^{n} \binom{n}{2k} C_k, \quad C_{n+1} = \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k} M_k.$$

Assume, for the moment, that n and l are integers, $n \geq 1$, $l \geq 0$. A **secondary structure** of **size** n and **rank** l is a labeled non-oriented graph S on the vertex set $V(S) = [n] = \{1, 2, ..., n\}$ whose edge set $E(S)$ consists of two disjoint subsets, $P(S)$ and $H(S)$, satisfying the following conditions:
(a) $\{i, i+1\} \in P(S)$, for all $1 \leq i \leq n-1$;
(b) $\{i,j\} \in H(S)$ and $\{i,k\} \in H(S) \Longrightarrow j = k$;
(c) $\{i,j\} \in H(S) \Longrightarrow |i-j| > l$;
(d) $\{i,j\} \in H(S), \{p,q\} \in H(S)$ and $i < p < j \Longrightarrow i < q < j$.

The set $H(S)$ may be empty, in which case we say that the secondary structure S is **trivial**. The cardinality of $H(S)$ is called the **order** of S. The only secondary structure of size 1 is the graph K_1 (one point).

In the language of molecular biology, the vertices of S are called **bases**, the edges of $P(S)$ are **p-bonds** and the edges of $H(S)$ are **h-bonds**. If an edge of $H(S)$ connects the vertices i and j, we say that the bases i and j are paired by an h-bond; vertices not incident to an edge from $H(S)$ are said to be unpaired. We shall resort to this terminology whenever it proves itself more concise.

We denote the set of all secondary structures of size n and rank l by $\mathcal{S}^{(l)}(n)$. The set of all such structures of order k is denoted by $\mathcal{S}_k^{(l)}(n)$. The cardinalities of these sets will be denoted by $S^{(l)}(n)$ and $S_k^{(l)}(n)$, respectively. By definition, we put $S^{(l)}(0) = 1$, for all l.

There are many ways to represent secondary structures graphically. The most convenient for our purpose is the **arc diagram**, in which the edges from $H(S)$ are depicted as circular arcs in the upper half-plane. From the definition of secondary structures, it is clear that these arcs may not intersect, that they may not share an endpoint and that each of them must leap over at least l consecutive vertices of S.

An arc diagram representing a secondary structure from $\mathcal{S}^{(2)}(12)$ is shown in *Figure 1*.

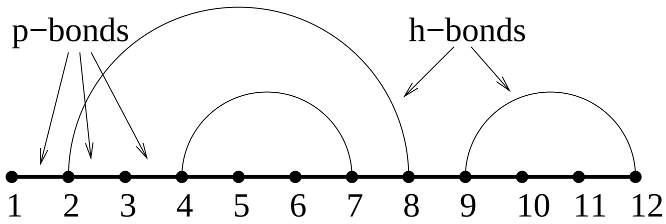

Figure 1. *An example of a secondary structure*

It has been known for long that the secondary structures of rank 0 are counted by the Motzkin numbers ([19]). Hence, we have $S^{(0)}(n) = M_n$.

A secondary structure of size n is **closed** if there is an h-bond connecting bases 1 and n. For given integers $n \geq 2$ and $l \geq 0$, there are $S^{(l)}(n-2)$ closed secondary structures of size n and rank l. We denote the set of all closed secondary structures of size n and rank l by $\mathcal{C}^{(l)}(n)$.

Starting from the arc diagram representation it is easy to obtain various correspondences between the secondary structures and certain classes of lattice paths. The case $l = 0$ leads to the Motzkin paths, and for $l = 1$ one obtains the Motzkin paths without peaks [19]. By relaxing the condition of non-negativity of l and allowing $l = -1$, one obtains a correspondence between degenerate secondary structures $\mathcal{S}^{(-1)}(n)$ and bi-colored Motzkin paths (or 2-Motzkin paths), much investigated recently by Chen *et al.* ([3, 8]). Another class of objects related to secondary structures are non-crossing partitions ([4, 5, 6, 15]). Secondary structures are simply non-crossing partitions with blocks whose size does not exceed 2. The non-crossing paradigm has been further extended toward non-crossing trees and graphs [7].

The results from the papers on the non-crossing partitions and lattice paths do not specialize easily to the case of secondary structures. It seems that there are considerable difficulties in taking into account the additional constraints imposed on them that convert them into secondary structures. Hence, it makes sense to look for a direct way to relate general secondary structures and plane trees.

2. Closed secondary structures and plane trees

Assume that n and l are integers, $n \geq 1$, $l \geq 0$. Denote by $\mathcal{T}^{(l)}(n)$ the set of all plane trees with exactly n leaves satisfying the following constraints:
(i) no vertex of a tree from $\mathcal{T}^{(l)}(n)$, has exactly one successor;

(ii) the leftmost and the rightmost successors of an internal vertex are leaves;
(iii) if all successors of a vertex v are leaves, there must be at least $l+2$ of them.

We say that a base k is **visible** from an arc $h = \{i,j\} \in H(S)$ if k is either an endpoint of h, or if $i < k < j$ and there are no arcs $\{p,q\} \in H(S)$ with $i < p < k < q < j$.Similarly, an arc $\{p,q\} \in H(S)$ is visible from an arc $\{i,j\} \in H(S)$ if there is no arc $\{r,s\} \in H(S)$ such that $i < r < p \leq q < s < j$.

Theorem 1. *There is a bijective correspondence between the sets $\mathcal{C}^{(l)}(n)$ and $\mathcal{T}^{(l)}(n)$.*

Proof. Take a closed secondary structure $S \in \mathcal{C}^{(l)}(n)$ and construct a tree T as follows: Assign a leaf w_i of T to each base i of S, and an internal node v_h of T to each arc h of S. Connect each internal node v_h with the nodes that correspond to bases and arcs that are visible from h. Obviously, if an arc g is visible from h, then v_h is the predecessor of v_g. The tree T is obviously an element of $\mathcal{T}^{(l)}(n)$, and the correspondence is bijective. $\square$

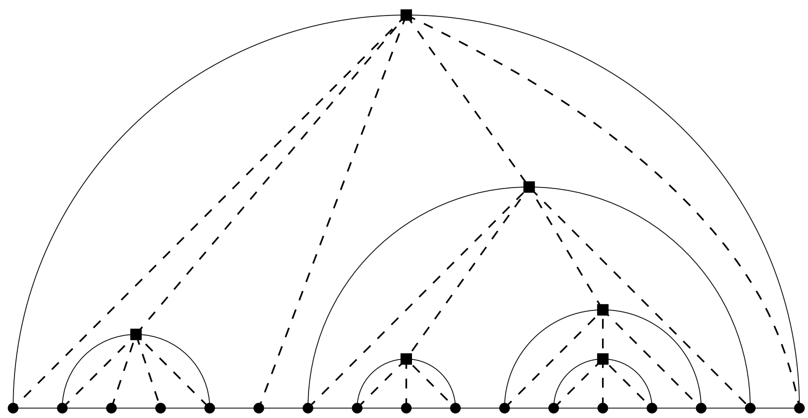

Figure 2. *The correspondence between secondary structures and plane trees*

An example of the correspondence just described is shown in *Figure 2*. The vertices of the secondary structure (and at the same time the leaves of the corresponding plane tree) are shown as circles; internal vertices of the tree are represented by black squares; the edges of the secondary structure are drawn with solid lines, and the edges of the tree are drawn in dashed lines.

For the case $l = -1$ the correspondence was obtained by Schmitt and Waterman [16], and also by Klazar in the context of non-crossing partitions [15].

3. Consequences

By setting the value of the parameter l equal to zero, we obtain some new combinatorial interpretations of Motzkin numbers.

Corollary 2. *For each integer $n \geq 1$, the n-th Motzkin number M_n counts the plane trees with $n+2$ leaves such that every internal vertex has at least 2 descendants, and the leftmost and the rightmost of them are leaves.*

The $m_3 = 4$ trees of $\mathcal{T}^{(0)}(5)$ are shown in *Figure 3*.

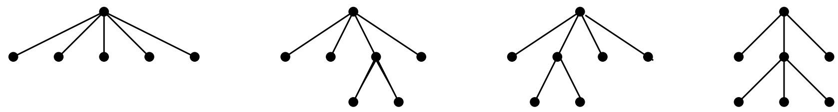

Figure 3. *All plane trees of* $\mathcal{T}^{(0)}(5)$

Corollary 3. *For each integer* $n \geq 1$*, the* n*-th Motzkin number* M_n *counts the plane forests with* n *leaves in which each component is a tree of* $\mathcal{T}^{(0)}(m)$*, for* $m \leq n$.

Proof. The claim follows by deleting the root vertex and all edges incident to it from the trees of $\mathcal{T}^{(0)}(n+2)$, and then discarding the leftmost and the rightmost trivial tree. An example is shown in *Figure 4*. □

Figure 4. *Plane forests with three leaves obtained according to Corollary 3*

By using the standard bijection between plane trees and lattice paths ([17], p. 256) we can translate the result of *Corollary 2* into the context of Dyck paths.

Corollary 4. *For each integer* $n \geq 1$*, the* n*-th Motzkin number* M_n *counts Dyck paths with* n *peaks that contain neither long slopes nor lonely peaks.*

Proof. Let T be a tree of $\mathcal{T}^{(0)}(n+2)$. We construct a Dyck path P by doing a depth-first (preorder) search through T, adding a U step when going down an edge of T, and a D step when going up an edge of T. After discarding the leftmost and the rightmost peak from P, we obtain a Dyck path with n peaks that has neither long slopes nor lonely peaks. As the construction can be reversed, the correspondence is bijective, and the claim follows. □

An interesting property of the interpretation of *Corollary 4* is that it is given in terms of peaks, and not in terms of total number of steps, as the interpretations in [17] and [10]. The nine Dyck paths with four peaks, none of them lonely, and without long slopes are shown in *Figure 5.*

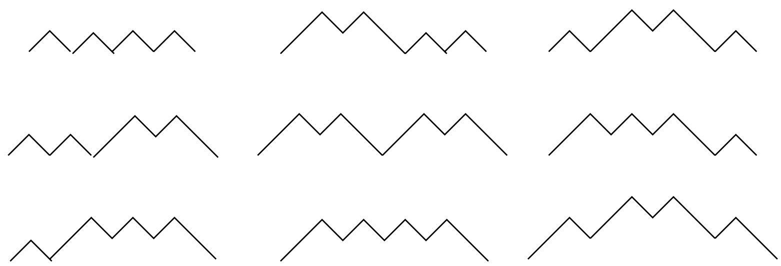

Figure 5. *Dyck paths without long slopes and lonely peaks with exactly four peaks*

Another way to arrive at the result of *Corollary 4* is to start from a secondary structure of size n and rank 0 and to construct a Dyck path as follows. We scan the structure from left to right. For a base where an h-bond starts, we add the sequence UUD of steps to the already constructed path. For a base where an h-bond ends, we add the sequence UDD. For an unpaired base, we add the sequence UD. It is clear from the defining properties of secondary structures that the resulting path will be a Dyck path without long slopes and lonely peaks.

It is clear from the construction of *Theorem 1* that the number of internal vertices of a tree from $\mathcal{T}^{(l)}(n)$ is equal to the order (i.e., to the number of h-bonds) of the corresponding secondary structure. By denoting the number of plane trees of $\mathcal{T}^{(l)}(n)$ with k internal vertices by $T_k^{(l)}(n)$, and using explicit formulas for $S_k^{(l)}(n)$ from, e.g., [11], we obtain the following result.

Corollary 5.
$T_k^{(0)}(n) = \binom{n-2}{2(k-1)} C_{k-1}$;
$T_k^{(1)}(n) = N(n-k-1, k)$. □

References

[1] M. Aigner, *Motzkin numbers*, European J. Combin. **19**(1998), 663–675.

[2] W. Y. C. Chen, *A general bijective algorithm for trees*, Proc. Natl. Acad. Sci. USA **87**(1990), 9635–9639.

[3] W. Y. C. Chen, L. Shapiro, L. Yang, *Parity reversing involutions on plane trees and 2-Motzkin paths*, European J. Combin. **27**(2006), 283–289.

[4] W. Y. C. Chen, E. Y. P. Deng, R. R. X. Du, *Reduction of m-regular noncrossing partitions*, European J. Combin. **26**(2005), 237–243.

[5] W. Y. C. Chen, E. Y. P. Deng, R. R. X. Du, R. P. Stanley, C. H. Yan, *Crossings and nestings of matchings and partitions*, Trans. Am. Math. Soc. **359**(2007), 1555–1575.

[6] W. Y. C. Chen, S. Y. J. Wu, C. H. Yan, *Linked partitions and linked cycles*, *European J. Combin.*, in press.

[7] W. Y. C. Chen, S. H. F. Yan, *Noncrossing trees and noncrossing graphs*, Electron. J. Combin. **13**(2006) #N12.

[8] W. Y. C. Chen, S. H. F. Yan, L. L. M. Yang, *Identities from weighted 2-Motzkin paths*, Adv. Appl. Math., in press.

[9] C. Cooker, *Enumerating a class of lattice paths*, Discrete Math. **271**(2004), 13–28.

[10] R. Donaghey, L. W. Shapiro, *Motzkin numbers*, J. Combin. Theory Ser A **23**(1977), 291–301.

[11] T. Došlić, D. Svrtan, D. Veljan, *Enumerative aspects of secondary structures*, Discrete Math. **285**(2004), 67–82.

[12] T. Došlić, D. Veljan, *Calculus proofs of some combinatorial inequalities*, Math. Inequal. Appl. **6**(2003), 197–209.

[13] C. Haslinger, P. F. Stadler, *RNA structures with pseudo-knots: Graph-theoretical, combinatorial and statistical properties*, Bull. Math. Biology **61**(1999), 437–467.

[14] I. L. Hofacker, P. Schuster, P. F. Stadler, *Combinatorics of RNA secondary structures*, Discrete Appl. Math. **88**(1998), 207–237.

[15] M. Klazar, *On trees and non-crossing partitions*, Discrete Math. **82**(1998), 263–269.

[16] W. R. Schmitt, M. S. Waterman, *Linear trees and RNA secondary structures*, Discrete Appl. Math. **51**(1994), 317–323.

[17] R. P. Stanley, *Enumerative Combinatorics vol. 2*, Cambridge Univ. Press, Cambridge, 1999.

[18] R. P. Stanley, *Catalan addendum*,
`http://www-math.mit.edu/~rstan/ec/catadd.ps.gz`

[19] P. R. Stein, M. S. Waterman, *On some new sequences generalizing the Catalan and Motzkin numbers*, Discrete Math. **26**(1979), 261–272.

[20] M. S. Waterman, *Secondary structures of single stranded nucleic acids*, In: *Studies on Foundations and Combinatorics. Advances in Mathematics Supplementary Studies, Vol. I*, (G.C. Rota, Ed.), Academic Press, New York, 1978, 167–212.