On the Prime Decomposition of Dyck Words

A. Panayotopoulos and A. Sapounakis
Department of Informatics
University of Piraeus
80 Karaoli & Dimitriou, 185 34 Piraeus
Greece

ABSTRACT. In this paper the decomposition of Dyck words into a product of Dyck prime subwords is studied. The set of Dyck words which are decomposed into k components is constructed and its cardinal number is evaluated.

1 Introduction

This work deals with the decomposition of Dyck words into prime factors and will be used for the analogous decomposition of Motzkin words [5]. On the other hand this decomposition is related to the decomposition of the set of all nested sets of $[2m] = \{1, 2, \ldots, 2m\}$ into subsets, the elements of which have a given number of outer pairs [7].

For the determination of the set of Dyck words of length $2m$ with certain number of factors we transform each Dyck word into a pair of partitions of m such that the first component of the pair dominates the second [3]. This relates Dyck words to graphs and matrices where the domination order appears [1], [6].

We recall some basic notions about Dyck words. A word $u \in \{\alpha, \bar{\alpha}\}^*$ is called Dyck if $|u|_\alpha = |u|_{\bar{\alpha}}$ and for every factorization $u = pq$ we have $|p|_\alpha \geq |p|_{\bar{\alpha}}$, where $|u|_\alpha$, $|p|_\alpha$, (resp. $|u|_{\bar{\alpha}}$, $|p|_{\bar{\alpha}}$) is the number of occurences of the letter α (resp. $\bar{\alpha}$) in u, p.

A Dyck path of length $2m$ is a path in the first quadrant, which begins at the origin, ends at $(2m, 0)$ and consists of steps $(+1, +1)$ (North-East) and $(+1, -1)$ (South-East). It is well known that the Dyck paths of length $2m$ are coded by the Dyck words $u = u_1 u_2 \ldots u_{2m}$ so that every North-East (resp. South-East) step corresponds to the letter $u_i = \alpha$ (resp. $u_i = \bar{\alpha}$)

It is also known that the cardinality of the set D_m of all Dyck words (or paths) of length $2m$ is equal to the number of Catalan $C_m = \frac{1}{m+1}\binom{2m}{m}$.

JCMCC **40** (2002), pp. 33-39

On the other hand the Dyck words (or paths) may by generated either by the context-free grammar $D \rightarrow \varepsilon + \alpha D\bar{\alpha}D$ where ε is the empty word [2] or by using binary trees [4].

Let $u = u_1u_2\ldots u_{2m}$ be a Dyck word of $\{\alpha, \bar{\alpha}\}^*$. Two indices $i, j \in [2m]$ such that $i < j$, $u_i = \alpha$ and $u_j = \bar{\alpha}$ are called *conjugates* with respect to u, if j is the smallest number in $[i+1, |u|]$ for which the subword $u_iu_{i+1}\ldots u_j$ is a Dyck word. Obviously the set $S = \{(i,j)\colon i < j \text{ and } i,j \text{ conjugates}\}$ is nested.

A non empty Dyck word that is not a product of two nonempty Dyck words is called *prime*. It is clear that the Dyck prime words are these that the only intersections of their coresponding Dyck paths with the x-axis are the initial and the final points of the paths.

It is evident that a Dyck word $u = u_1u_2\ldots u_{2m}$ is prime if and only if $1, 2m$ are conjugates with respect to u. Further, if the first and the last letter of a Dyck prime word are removed one gets a Dyck word. Therefore the number of Dyck primes of length $2m+2$ is equal to the number of Dyck words of length $2m$. Thus if we denote by P_m the set of all Dyck prime words of length $2m$ we obtain that

$$|P_1| = 1 \text{ and } |P_{m+1}| = C_m, \ m \in N^*.$$

Given a Dyck word $u = u_1u_2\ldots u_{2m}$ let $i_1 = 1$, j_1 the conjugate of i_1 and $u^1 = u_{i_1}\ldots u_{j_1}$. It is clear that u^1 is a Dyck prime word. Moreover if $u^1 \neq u$ let $i_2 = j_1 + 1$, j_2 the conjugate of i_2 and $u^2 = u_{i_2}\ldots u_{j_2}$. It is clear that u^2 is also a Dyck prime word. Proceeding in this way we define recursively a finite sequence $u^1, u^2, \ldots, u^k$ of Dyck prime words such that $u = u^1u^2\ldots u^k$. Moreover it is easy to check that this sequence is unique. From the above discussion we have the following proposition.

Proposition 1.1. *Every Dyck word is uniquely decomposed into a product of Dyck prime words.*

The Dyck words which are decomposed into k Dyck prime words (components) are those whose coresponding Dyck paths meet the x-axis at exactly $k-1$ points, apart from the points $(0,0)$ and $(0,2m)$.

Let $D_{m,k}$ be the set of all Dyck words of length $2m$ which are decomposed into k Dyck primes and $d(m,k)$ its cardinal number.

In section 2 the number $d(m,k)$ is evaluated with the aid of several recursion formulas. In section 3 the elements of $D_{m,k}$ are determined using a transformation between the Dyck words and some pairs of finite sequences.

2 The number $d(m, k)$

For the number $d(m, k)$ of all Dyck words of length $2m$ which are decomposed into k Dyck prime words we have the following formulas:

$$d(m,k) = 0, \quad \text{if } k > m.$$
$$d(m,m) = 1.$$
$$d(m,1) = C_{m-1}.$$
$$d(m,1) + d(m,2) + \cdots + d(m,m) = C_m.$$

Further by proposition 1.1 we obtain the following result.

Proposition 2.1. *For $m \geq k$, we have $d(m,k) = \sum C_{x_1} C_{x_2} \ldots C_{x_k}$ where the sum is taken over all $(x_1, x_2, \ldots, x_k)$ with $x_i \in N$ and $x_1 + x_2 + \cdots + x_k = m - k$.*

We now give the main result of this section

Proposition 2.2. *For $2 \leq k \leq m$ we have*

$$d(m,k) = \sum_{\lambda=k-1}^{m-1} d(m-1, \lambda).$$

Proof: We define a function

$$f\colon D_{m,k} \to \bigcup_{\lambda=k-1}^{m-1} D_{m-1,\lambda}$$

as follows: Given a word $u = u_1 u_2 \ldots u_{2m} \in D_{m,k}$ and $j \in [2m]$ the conjugate of 1 with respect to u we set,

$$f(u) = \begin{cases} u_3 u_4 \ldots u_{2m} & \text{if } j = 2 \\ u_2 u_3 \ldots u_{j-1} u_{j+1} \ldots u_{2m}, & \text{if } 2 < j. \end{cases}$$

It follows that $f(u)$ is a Dyck word of length $2m-2$. Moreover if $j = 2$ the word $f(u)$ is decomposed into $k-1$ components so that $f(u) \in D_{m-1,k-1}$. On the other hand if $j > 2$ the word $u_2 u_3 \ldots u_{j-1}$ is also Dyck. Let ϱ be the number of prime components into which this word is decomposed. Since the word $u_{j+1} u_{j+2} \ldots u_{2m}$ is decomposed into $k-1$ components it follows that $f(u)$ is decomposed into $\varrho + k - 1$ components and $f(u) \in D_{m-1,\varrho+k-1}$.

Further, since

$$1 \leq \varrho \leq \frac{j-2}{2} \text{ and } j + 2(k-1) \leq 2m$$

we deduce easily that

$$k \leq \varrho + k - 1 \leq m - 1 \text{ and } f(u) \in \bigcup_{\lambda=k}^{m-1} D_{m-1,\lambda}.$$

This shows that in general

$$f(u) \in \bigcup_{\lambda=k-1}^{m-1} D_{m-1,\lambda}$$

and the function f is well defined.

We will show that f is a bijection. Clearly since f is $1-1$ it is enough to show that f is onto. For this let

$$v = v_1 v_2 \dots v_{2m-2} \in \bigcup_{\lambda=k-1}^{m-1} D_{m-1,\lambda}.$$

We consider two cases:

If $v \in D_{m-1,k-1}$ we define $u = \alpha\bar{\alpha}v$. On the other hand if $v \in D_{m-1,\lambda}$ for some $\lambda \in [k, m-1]$ and i is the index corresponding to the last letter of the $(\lambda - k + l)$th component of v we define $u = \alpha v_1 \dots v_i \bar{\alpha} v_{i+1} \dots v_{2m-2}$.

In both cases we have $u \in D_{m,k}$ and $f(u) = v$. This shows that f is a bijection and

$$d(m,k) = |D_{m,k}| = \sum_{\lambda=k-1}^{m-1} |D_{m-1,\lambda}| = \sum_{\lambda=k-1}^{m-1} d(m-1,\lambda).$$

□

From the above proposition and the equality $d(m,1) = C_{m-1}$ we obtain the values of $d(m,k)$, $k \leq m$ (see Table 1).

Table 1: The values of $d(m,k)$

$m\backslash k$	1	2	3	4	5	6
1	1					
2	1	1				
3	2	2	1			
4	5	5	3	1		
5	14	14	9	4	1	
6	42	42	28	14	5	1

We conclude this section with some formulas for the number $d(m,k)$.

$$d(m,2) = C_{m-1}. \tag{1}$$

$$d(m,k) = d(m,k-1) - d(m-1,k-2). \tag{2}$$

$$\sum_{k=1}^{m}(k+1)d(m,k) = C_{m+1}. \tag{3}$$

$$d(m,k) = \sum_{\lambda=1}^{m-k+1} C_{\lambda-1}d(m-\lambda,k-1). \tag{4}$$

$$d(m,k) = \sum_{\lambda=0}^{[\frac{k-1}{2}]}(-1)^{\lambda}\binom{k-1-\lambda}{\lambda}C_{m-\lambda-1}. \tag{5}$$

$$d(m,k) = \frac{k}{m}\binom{2m-k-1}{m-1}. \tag{6}$$

The proofs of (1) and (2) are easy whereas the proof of (3) follows immediately from proposition 2.2. Finally, the equalities (4), (5) and (6) are deduced easily from (3) by induction.

3 The generation of $D_{m,k}$

We recall that a ν-partition of m is a finite sequence $x = (x_i)$, $i \in [\nu]$ of positive integers such that

$$\sum_{i=1}^{\nu} x_i = m.$$

For two ν-partitions $x = (x_i)$ and $y = (y_i)$, $i \in [\nu]$ of m we say that x *dominates* y if and only if

$$\sum_{i=1}^{\varrho} x_i \geq \sum_{i=1}^{\varrho} y_i \text{ for each } \varrho \in [\nu - 1]$$

$$\sum_{i=1}^{\nu} x_i = \sum_{i=1}^{\nu} y_i = m.$$

It is well known that this relation is a partial order [3].

Clearly every Dyck word $u \in D_m$ is written uniquely in the form $u = \alpha^{x_1}\bar{\alpha}^{y_1}\alpha^{x_2}\bar{\alpha}^{y_2}\ldots\alpha^{x_\nu}\bar{\alpha}^{y_\nu}$ so that the two sequences $x = (x_i)$ and $y = (y_i)$, $i \in [\nu]$ are ν-partitions of m and x dominates y.

Then the sequences $\eta = (\eta_i)$ and $\theta = (\theta_i)$ with

$$\eta_i = \sum_{j=1}^{i} x_j \text{ and } \theta_i = \sum_{j=1}^{i} y_j, i \in [\nu]$$

are increasing such that $\eta_i \geq \theta_i$ for each $i \in [\nu]$ and $\eta_\nu = \theta_\nu = m$.

For example for the Dyck word $u = \alpha\bar{\alpha}\alpha\alpha\bar{\alpha}\alpha\bar{\alpha}\bar{\alpha}\alpha\alpha\bar{\alpha}\alpha\alpha\bar{\alpha}\bar{\alpha}\bar{\alpha}$ we have that $u \in D_{8,3}$ and $x = (1,2,1,2,2)$, $y = (1,1,2,1,3)$, $\eta = (1,3,4,6,8)$, $\theta = (1,2,4,5,8)$.

If A_m is the set of all pairs (η, θ) as above we can easily check that the function $\gamma(u) = (\eta, \theta)$ defines a bijection between the sets D_m and A_m.

This discusion suggests the following.

Proposition 3.1. *There exists a bijection between the sets D_m and A_m.*

Clearly using this bijection the set $D_{m,k}$ may be regarded as the subset of A_m such that the underlying sequences coincide for exactly k indices.

For the generation of $D_{m,k}$ one may consider for each $\nu \geq k$ the sets $I_{m,\nu}$ of all increasing sequences $\eta = (\eta_i)$, $i \in [\nu]$, in $[m]$ with $\eta_\nu = m$ and compare their elements in order to find the pairs $(\eta, \theta) \in I_{m,\nu} \times I_{m,\nu}$ with $\eta_i \geq \theta_i$ for each $i \in [\nu]$ and $|\{i \in [\nu] \colon \eta_i = \theta_i\}| = k$.

In the following we present a recursive method for the generation of $D_{m,k}$. This method is based on the idea used in the proof of proposition 2.2 and shows how the set $D_{m,k}$ is generated by the set $D_{m-1,\lambda}$ where $\lambda \in [k-1, m-1]$.

Construction: Given a word $w \in \bigcup_{\lambda=k-1}^{m-1} D_{m-1,\lambda}$ with $\gamma(u) = (\eta, \theta)$, $\eta = (\eta_i)$, $\theta = (\theta_i)$, $i \in [\nu]$ we define a word u^* by $\gamma(u^*) = (\eta^*, \theta^*)$, $\eta^* = (\eta_i^*)$, $\theta^* = (\theta_i^*)$, as follows:

If $u \in D_{m-1,k-1}$ then $\eta_1^* = \theta_1^* = 1$, $\eta_{i+1}^* = \eta_i + 1$ and $\theta_{i+1}^* = \theta_i + 1$.

If $u \in D_{m-1,\lambda}$ for $\lambda \geq k$, then $\eta_i^* = \eta_i + 1$, $\theta_i^* = \theta_i$ for $i < p$, and $\theta_i^* = \theta_i + 1$ for $i \geq p$ where p is the $(\lambda - k + 1)$th element of $[\nu]$ for which $\eta_p = \theta_p$.

We can easily check that $u^* \in D_{m,k}$ and every element of $D_{m,k}$ may be constructed by some element of

$$\bigcup_{\lambda=k-1}^{m-1} D_{m-1,\lambda}$$

in this way.

We finally remark that since the set of all nested sets of $[2m]$ with exactly k outer pairs [7] may also be identified with $D_{m,k}$, the above method may be used for its construction.

References

[1] C. Berge, *Theorie des graphes et ses applications*, Dunod, 1958.

[2] A. Denise and R. Simion, Two combinatorial statistics on Dyck paths, *Discrete Math.* **137** (1995), 155–176.

[3] T.V. Narayana, A partial order and its applications to probability, *Sankhya* **21** (1959), 91–98.

[4] A. Panayotopoulos and A.Sapounakis, On binary trees and Dyck paths, *Math. Inf. Sci. Hum.* **131** (1995), 39–51.

[5] A. Panayotopoulos and A. Sapounakis, On Motzkin words and non-crossing partitions, (to be submitted).

[6] H.J. Ryser, Matrices of zeros and ones, *Bull. A.M.S.* **66** (1960), 442–464.

[7] A. Sapounakis and P. Tsikouras, Nested Sets of Pairs, *J. Infor. and Optim. Sciences* **16** (1995), 549–555.