

Around the cycle lemma

Dominique Perrin

June 10, 2009

Outline

- Words and trees
- The cycle lemma
- Enumeration of codes
- The non-commutative cycle lemma

CHAPTER 11

Words and Trees

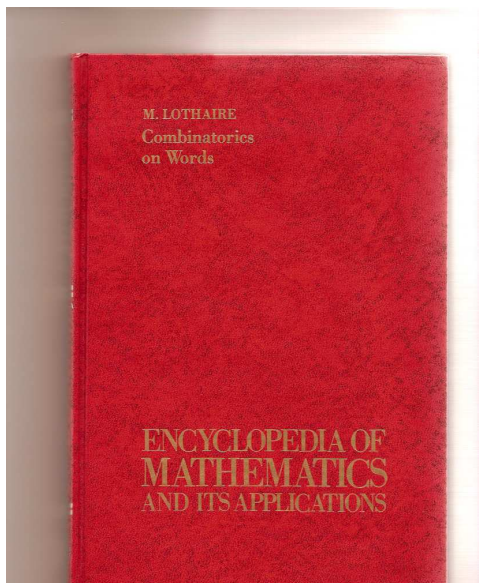
11.0. Introduction

The aim of this chapter is to give a detailed presentation of the relation between plane trees and special families of words: parenthesis systems and other families. The relation between trees and parenthesis notation is classical and has been known perhaps since Catalan 1838.

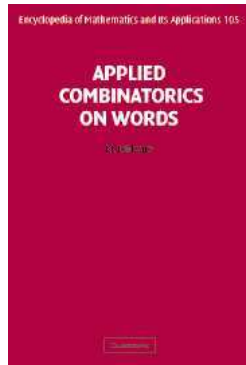
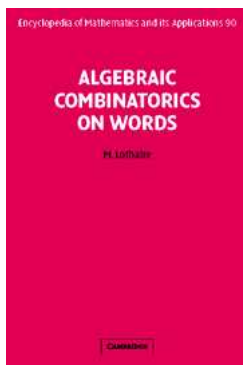
Because trees play a central role in the field of combinatorial algorithms (Knuth 1968), their coding by parenthesis notation has been investigated so very often that it is quite impossible to give a complete list of all the papers dealing with the topic. These subjects are also considered in enumeration theory and are known to combinatorialists (Comtet 1970) as being counted by Catalan numbers. Note that a generalization of the type of parenthesis system often called Dyck language is a central concept in formal language theory. These remarks give a good account of the main role played by trees and their coding in combinatorics on words.

Presented here are three ways to represent trees by words. The first one

The little red book



The Lothaire series



The cycle lemma

A word $w \in \{a, b\}^*$ is **k -dominating** if any prefix p of w is such that $|p|_a > k|p|_b$.

Lemma (Dvoretzky, Motzkin, 1947)

Let $k, m, n \geq 0$ be such that $m \geq kn$. Any word with m a and n b has $m - kn$ conjugates which are k -dominating.

Joseph Bertrand



The enumeration of paths in the plane was studied early in connection with probability theory. The first significant result, a

solution to the *ballot problem*, is due to **Bertrand** in 1887 . It says

Suppose that, in a ballot, candidate P scores p votes and candidate Q scores q votes, where $p > q$. The probability that throughout the counting there are always more votes for P than for Q equals $(p - q)/(p + q)$.



The k -ary Lukasiewicz language is

the solution of the algebraic equation $L = L^{k+1}b + a$ (for $k = 1$, postfix polish notation).

For a word $w \in \{a, b\}^*$, denote $\delta(w) = |w|_a - k|w|_b$. A word w is in L iff it is k dominating and $\delta(w) = 1$.

A word w has $\delta(w)$ conjugates in L^* .

Enumeration of trees

The number of

- planar trees with $n + 1$ vertices,
- k -ary Lukaziewicz words with $kn + 1$ letters,
- k -ary trees with n internal vertices

is

$$T_{n,k} = \frac{1}{(k-1)n+1} \binom{kn}{n}$$

The $T_{n,2}$ are the **Catatlan** numbers.

Enumeration of bifix codes

The number of maximal binary prefix codes with n elements is

$$\alpha_n = T_{n-1,2}.$$

Problem

What is the number β_n of maximal bifix codes with n elements?

For $k = 2$,

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
α_n	0	1	2	5	14	42	132	429	1430							
β_n	0	1	0	1	0	0	0	1	2	0	0	0	0	0	0	1

The 3/4 conjecture

Theorem (Kraft, McMillan, 1956)

There exists a binary prefix code with n_i words of length i if and only if $\sum_{i \geq 1} n_i 2^{-i} \leq 1$.

Conjecture (Alswhehde et al., 1996)

For any sequence of integers n_1, n_2, \dots such that $\sum_{i \geq 1} n_i 2^{-i} \leq 3/4$, there exists a binary bifix code X with n_i words of length i .

Best possible bound: no word of length $\leq i$ can be added to $X = a \cup b\{a, b\}^{i-2}b$ with Kraft sum $1/2 + 2^{i-2}2^{-i} = 3/4$.

Enumeration of codes

Problem

What is the number γ_n of maximal binary codes with n elements?

n	1	2	3	4	5	6	7	8	9
α_n	0	1	2	5	14	42	132	429	1430
β_n	0	1	0	1	0	0	0	1	2
γ_n	0	1	4	9	39	84			

Explanation: if X is a prefix code, then \tilde{X} is a suffix code. For $n \leq 4$, all maximal codes with n elements are prefix or suffix and thus $\gamma_n = 2\alpha_n - \beta_n$ where β_n is the number of maximal bifix codes with n elements.

$$\gamma_5^{(2)} = 2\alpha_5^{(2)} + \alpha_5^{(3)}\alpha_3^{(2)} + \alpha_3^{(2)}\alpha_5^{(3)} - 1$$

The non-commutative cycle lemma

A word w on the alphabet $A \cup \bar{A}$ has **good reduction** if

- (i) It has no prefix reducing to 1.
- (ii) Its reduced form is also cyclically reduced.

Generalization of 1-dominating words for $A = \{a, \bar{a}\}$, (condition (ii) is always satisfied).

Example: $a\bar{b}b$ has good reduction.

Theorem (Armstrong et al., 2009)

A word with a cyclic reduction of length p has p conjugates with good reduction.

$w = aa\bar{a}bb\bar{b}\bar{a}\bar{a}$ has cyclic reduction $b\bar{a}$ and conjugates

$$\bar{a}bb\bar{b}\bar{a}\bar{a}aa, \quad bb\bar{b}\bar{a}\bar{a}aa\bar{a}$$

with good reduction.

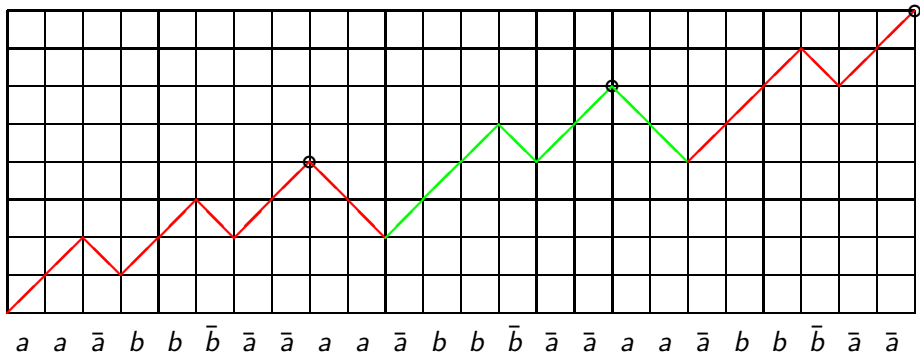


Figure: A path labeled $aa\bar{a}bb\bar{b}\bar{a}\bar{a}$

The **standard** cyclic reduction is the one corresponding to the least circular shift with good reduction.

Theorem (Armstrong et al., 2009)

The number of words of length n with given standard cyclic reduction of length k is

$$(2|A| - 1)^{(n-k)/2} \binom{n}{(n-k)/2}.$$

Thus, this number depends only on k .

Applications to random matrices.