TECHNION - Israel Institute of Technology
Computer Science Department

THE CYCLE LEMMA AND SOME APPLICATIONS

by

Nachum Dershowitz* and Shmuel Zaks

Technical Report #238

April 1982

* Dept. of Computer Science, University of Illinois at Urbana-Champaign Urbana, ILL. 61801, U.S.A. Current address: Dept. of Mathematics and Computer Science, Bar-Ilan University, Ramat Gan, Israel.
Research supported in part by the National Science Foundation under

ABSTRACT

Two proofs of a frequently rediscovered combinatorial lemma are presented. Using the lemma it is shown that the number of ordered trees with n_j nodes of degree j, $0 \leq j \leq d$, no restrictions on nodes of degree greater than d, and a total of n nodes is

$$\frac{1}{n}\binom{n-e-d(n-m)-2}{n-m-1}\binom{n}{n_0,\ldots,n_d,n-m},$$

where $m = \Sigma n_j \leq n-1$ and $e = \Sigma j n_j$. In the case when $m = \Sigma n_j = n$ the number of these trees is known to be

$$\frac{1}{n}\binom{n}{n_0,\ldots,n_d}.$$

Additional illustrations of the use of the lemma are given.

THE CYCLE LEMMA

A sequence $p_1 p_2, \dots, p_\ell$ of boxes and circles is called k-dominating if for every position i, $1 \leqslant i \leqslant \ell$, the number of boxes in $p_1 p_2 \cdots p_i$ is more than k times the number of circles (k is a positive integer). For example, the sequence □□□□□○○□□□○ is 2-dominating; the sequence □□□□○○□□○ is 1-dominating (or just dominating) but not 2-dominating; the sequences ○□□□□○○□□ and □□○○□□○□□ are not even 1-dominating.

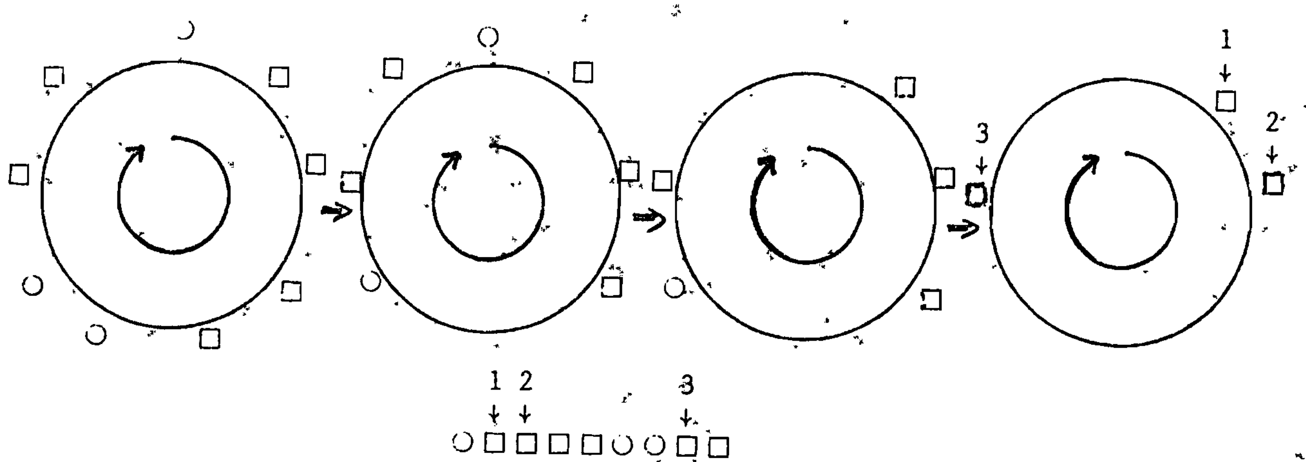
The following lemma has been rediscovered many times. Though it is not difficult to prove, it is a powerful tool in enumeration arguments.

Cycle Lemma (Dvoretzky and Motzkin [1947]): For any sequence $p_1 p_2 \cdots p_{m+n}$ of m boxes and n circles, $m \geqslant kn$, there exist exactly $m-kn$ cyclic permutations $p_j p_{j+1} \cdots p_{m+n}, p_1 \cdots p_{j-1}$, $1 \leqslant j \leqslant m+n$, that are k-dominating.

For example, of the nine cyclic permutations of the sequence ○□□□□○○□□ of six boxes and three circles, only three are dominating: □□○□□□□○○, □□□□○○□□○ and □□□○○□□○□. None are 2-dominating. As a special case of this lemma, if $m = n+1$, then there is a unique dominating permutation.

First Proof: For the first proof of the lemma, arrange the $m+n$ figures on a cycle. Removing a subsequence of k boxes followed by one circle from the cycle does not change the number of k-dominating permutations, since those $k+1$ figures have no net effect and no k-must be such a subsequence on the cycle; these subsequences are removed one-by-one until only boxes remain. The remaining m-kn boxes yield m-kn k-dominating cyclic permutations.

<u>Example</u>: We consider the sequence ○□□□□○○□□, with $k = 1$, and start by placing it on a cycle. After three removal steps we are left with three boxes, that correspond to the starting points of the three dominating cyclic permutations □□□□○○□□○, □□□○○□□○□ and □□○□□□□○○ (see Figure 1).



First Proof (k=1)

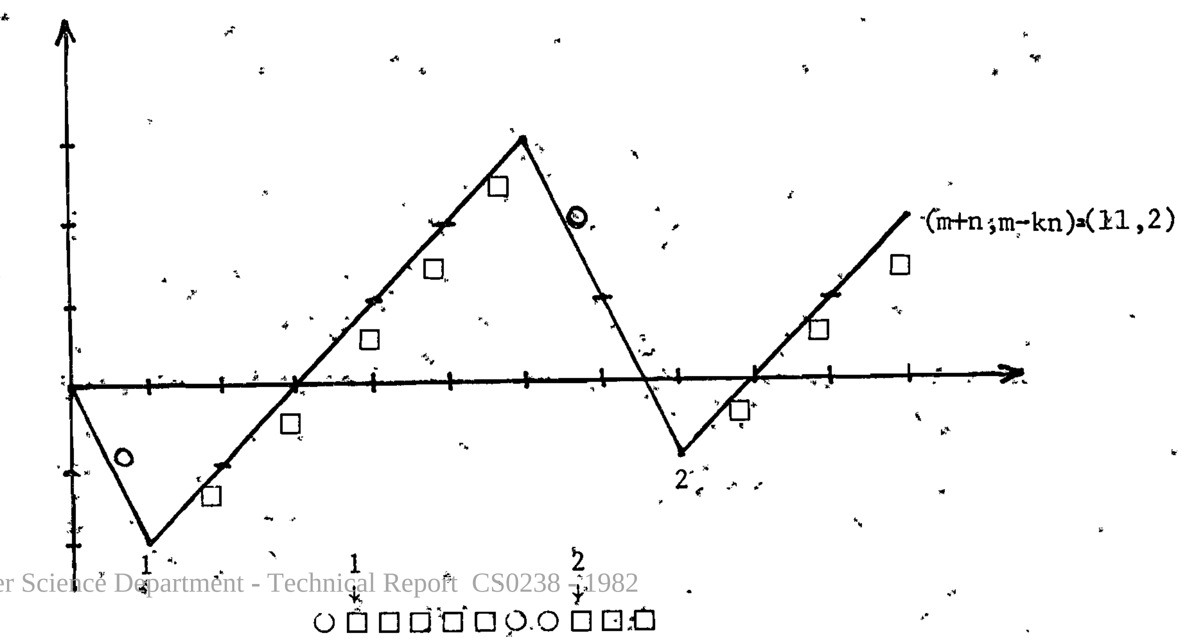
<u>Figure 1</u>

<u>Second Proof</u>: Another simple proof is the following: Given a sequence of figures, construct a "mountain range". Begin to the left at "sea level", slope upwards one unit for each box and slope downwards k units for each circle. The resultant range extends m+n units to the right and ends m-kn units above sea level.

By cyclically permuting the sequence, the origin is moved to a different point along the range. A k-dominating sequence corresponds equally low (or lower) valley to its right — otherwise, that valley would end up at (or below) sea-level, and (2) which is less than m-kn units above the deepest valley — otherwise, that valley would descend to (or past) sea-level. Any point that was to the right of the new origin was higher and is therefore above sea-level now; any point that was to its left was less than m-kn units lower and is therefore above sea-level now.

Clearly, there are exactly m-kn such points to choose a valid origin from.

<u>Example</u>: We consider the sequence ○□□□□□○○□□□, with k = 2, and start by constructing the corresponding "mountain-range", as described above. The two numbered points in that range (see Figure 2) correspond to the starting points of the two 2-dominating cyclic permutations □□□□□○○□□□○ and □□□○□□□□□○○.



Second Proof (k=2)

<u>Other Proofs</u>: Other proofs of varying degree of generality may be found in Dvoretzky and Motzkin [1947] (this paper is discussed in Grossman [1950]), Motzkin [1948] (two proofs), Hall [1958], Raney [1960], Yaglom and Yaglom [1964], Takács [1967], Silberger [1969], Bergman [1978] (three proofs), Sands [1978], and Singmaster [1978]. (The first paper is not credited by the other authors, but is referenced by Barton and Mallows [1965] and Mohanty [1979]). Dvoretzky and Motzkin [1947], Motzkin [1948], and Yaglom and Yaglom [1964] give the lemma in its general form; the others prove only the case $k = 1$ or $m-kn = 1$. Our first proof is a generalization of Silberger's, one of Bergman's, and Singmaster's; our second proof resembles Grossman's and Yaglom and Yaglom's.

APPLICATIONS

Pattern Matching

A sequence $x_0 x_1 \ldots x_n$ of n+1 nonnegative integers is called *subordinating* if the sum of all the x_j is n and the partial sums $\sum_{j=0}^{i} x_j$ are not greater than i, for all i $(0 \leq i \leq n)$. A *pattern* $p_1, p_2, \ldots, p_m$ of m distinct integers is said to *occur* in a sequence x if the p_i's all appear within x in any order. Using the Cycle Lemma, we can show that the total number of occurrences of such a pattern in the set of subordinating sequences of n+1 integers (there may be more than one occurrence of a pattern in a sequence) is

$$\frac{n!}{(n-m+1)!}\binom{2n-e-m}{n-m},$$

where $e = \sum_i p_i$. This result can be extended to patterns whose elements are subsequences of integers (by replacing n with n-d+m, where d is the total number of integers in all the patterns) and to nondistinct patterns; it generalizes Dershowitz and Zaks [1980] and Flajolet and Steyaert [1980] (in both m=1).

To prove the above formula, consider the correspondence between subordinating sequences of n+1 integers and dominating sequences of n+1 boxes and n circles obtained by replacing an integer x_j with one box followed by x_j circles. (The number of circles for i+1 boxes is $\sum_{j=0}^{i} x_j$ which is no greater than i.) By the Cycle Lemma, to each subordinating sequence of integers there corresponds a unique cycle of figures and, hence, a unique cycle of integers. Since there are $\frac{n!}{(n+1-m)!}$ ways of placing the m elements of the pattern on a cycle with n+1 positions and there are $\binom{2n-e-m}{n-e}$ ways of decomposing the remainder n-e of the sum into the n+1-m integers that must still be determined, it follows t-ary Trees

The number of t-ary trees with n internal nodes of (out-) degree t and tn+1-n leaves of degree 0 is $\frac{1}{tn+1}\binom{tn+1}{n}$ (see Klarner [1966]; Grunert [1841] gives the analogous result for polygons). In particular, the number of binary trees is $\frac{1}{2n+1}\binom{2n+1}{n} = \frac{1}{n+1}\binom{2n}{n}$, the well-known Catalan numbers, see Cayley [1859]; Silberger [1969], Sands [1978], and Singmaster [1978] derive the Catalan sequence using the Cycle Lemma.

To see this, note the correspondence between t-ary trees and (t-1)-dominating sequences obtained by traversing the tree in postorder, i.e. first each of its subtrees from left to right is traversed and then the node connecting them, and recording a circle for each internal node encountered and a box for each leaf encountered. By the Cycle Lemma, each (t-1)-dominating sequence corresponds to a cycle of tn+1-n boxes and n circles, of which there are exactly $\frac{1}{tn+1}\binom{tn+1}{n}$.

Ordered Trees

In an ordered (or plane-planted) tree the order in which subtrees of a node are arranged is significant. Every ordered tree corresponds to a cycle containing the (out-) degrees of the nodes listed in postorder, since a postorder list of degrees yields a subordinating sequence of integers and, as we have already seen, each subordinating sequence of integers corresponds to a unique cycle of integers. Using this correspondence between trees and cycles, one can show that the number of ordered trees with n_j nodes of degree j, $0 \leq j \leq d$, no restrictions on nodes of degree greater than d, and

$$\frac{1}{n}\binom{n-e-d(n-m)-2}{n-m-1}\binom{n}{n_0,\ldots,n_d,n-m},$$

where $m = \Sigma n_j \leq n-1$ is the total number of restricted nodes, $e = \Sigma jn_j$ is the total number of edges accounted for, and the last factor is the multinomial coefficient $\frac{n!}{n_0!\ldots n_d!(n-m)!}$. In the case when $m = \Sigma n_j = n$ (all the degrees are specified) this simplifies to $\frac{1}{n}\binom{n}{n_0,\ldots,n_d}$.

To count the number of cycles with these restrictions, note that there are $\frac{1}{n}\binom{n}{n_0,\ldots,n_d,n-m}$ ways of placing the degrees of the restricted nodes on a cycle with n positions. The remaining $n-m$ unrestricted nodes must have degrees greater than $d+1$ and summing to $n-e-1$. The number of ways to place these degrees is the same as the number of ways of decomposing the integer $(n-e-1) - (d+1)(n-m)$ into $n-m$ integers greater than 0, which is $\binom{n-e-d(n-m)-2}{n-m-1}$.

This formula generalizes the enumeration of unrestricted $(d = -1)$ ordered trees in Harary et al. [1964], the multinomial formula for the fully restricted $(d = n)$ case in Erdélyi and Etherington [1940] (in the context of subdivisions of polygons), and the result in Narayana [1959] for $d = 0$ (in the context of orderings on partitions). The Cycle Lemma can also be used to solve the two color case considered by Tutte [1964]. Our proof is a generalization of those in Raney [1960] and Sands [1978] (both for the case $d = n$) and Dershowitz and Zaks [1980] (for $d = 0$).

Using the above formula, we get that the total number of ordered trees with no unary nodes and k leaves $(d = 1, n_0 = k, n_1 = 0,$ hence $m = k$ and $e = 0)$, is

$$\sum_n \frac{1}{n+1} \binom{k-2}{n-k} \binom{n+1}{k} = \frac{1}{k} \sum_n \binom{k-2}{n-k} \binom{n}{k-1} .$$

For k = 1 through 7, the corresponding numbers of trees are 1, 1, 3, 11, 45, 197, and 903. (This sequence was investigated in Schröder [1870]; its relation to polygon partitions is discussed in Motzkin [1948]; its relation to trees appears in Knuth [1968].)

REFERENCES

Barton, D.E. and Mallows, C.L. [1965], Some aspects of the random sequence, Annals of Math. Statistics, Vol. 36, pp. 236-260.

Bergman, G.M. [1978], Terms and cyclic permutations, Algebra Universalis, Vol. 8, pp. 129-130.

Cayley, A. [1859], On the analytical forms called trees, Phil. Mag., Vol.20, pp. 374-378. Also, Coll. Math. Papers, Vol. 4 [1891], pp. 112-115.

Dershowitz, N. and Zaks, S. [July 1980], Enumerations of ordered trees, Discrete Math., Vol. 31, No. 1, pp. 9-28.

Dvoretzky, A. and Motzkin, Th. [1947], A problem of arrangements, Duke Math. J., Vol. 14, pp. 305-313.

Erdélyi, A. and Etherington, I.M.H. [1940], Some problems of non-associative combinations (2), Edinburgh Math. Notes, Vol. 32, pp. 7-12.

Flajolet, Ph. and Steyaert, J.M. [July 1980], On the analysis of tree-matching algorithms, Proc. Seventh Colloq. on Automata, Languages and Programming, Noordwijkerhout, The Netherlands, pp. 208-219.

Grossman, H.D. [1950], Fun with lattice points - 21, Scripta Math., Vol. 16, pp. 120-124.

Grunert, J.A. [1841], Ueber die Bestimmung der Anzahl der verschiedenen Arten, auf welche sich ein n-eck durch Diagonalen in lauter m-ecke zerlegen lässt, Archiv der Math. und Physik, Vol. 1 pp. 193-203.

Hall, P. [1958], Some word problems, J. London Math. Soc., Vol. 33, pp. 482-496.

Harary, F., Prins, G., and Tutte, W.T. [1964], The number of plane trees, Indagationes Math., Vol. 26, pp. 319-329.

Klarner, D.A. [1966], Enumeration involving sums over compositions, Ph.D. thesis, University of Alberta, Edmonton, Canada.

Knuth, D.E. [1968], The Art of Computer Programming, Vol. 1: "Fundamental Algorithms", Addison-Wesley, Reading, MA.

Mohanty, S.G. [1979], Lattice Path Counting and Applications, Academic Press, New York, NY.

Motzkin, Th. [April 1948], Relations between hypersurface cross ratios and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products, Bulletin Amer. Math. Soc., Vol. 54, pp. 352-360.

Narayana, T.V. [1959], A partial order and its application to probability,

REFERENCES (cont'd)

Sands, A.D. [February 1978], On generalized Catalan numbers, Discrete Math., Vol. 21, No.2, pp. 219-221.

Schröder, E. [1870], Vier combinatorische Probleme, Zeit. fur Math. und Physik, Vol. 15, pp. 361-376.

Silberger, D.M. [1969], Occurrences of the integer (2n-2)!/n!(n-1)!, Roczniki Polskiego Towarzystwa Math. I, Vol. 13, pp. 91-96.

Singmaster, D. [May 1978], An elementary evaluation of the Catalan numbers, Amer. Math. Monthly, Vol. 85, No.5, pp. 366-368.

Takács, L. [1967], Combinatorial Methods in the Theory of Stochastic Processes, Wiley, New York, NY.

Tutte, W.T. [1964], The number of planted plane trees with a given partition, Amer. Math. Monthly, Vol. 71, pp. 272-277.

Yaglom, A.M. and Yaglom, I.M. [1964], Challenging Mathematical Problems with Elementary Solutions, Vol.1: "Combinatorial Analysis and Probability Theory", Holden Day, San Francisco, CA.