

FORMULAE FOR WELL FORMED FORMULAE AND THEIR ENUMERATION

Dedicated to the memory of Hanna Neumann

DOV TAMARI

(Received 10 July 1972)

Communicated by M. F. Newman

Introduction

1. *General Idea.* In this paper formal operators and bracketings devices are essentially the same; so are well formed formulae and correct bracketings.

A well known theorem of symbolic logic (Post languages, etc.) and universal algebra characterizes well formed formulae among (linear) strings of symbols for variables and operators in terms of a system of numerical inequalities, one of them an equality.¹

A better idea is to consider circular formulae. By bending formulae into oriented circles, or cycles, in particular well formed formulae into so-called unit cycles, one overcomes the asymmetry between head (left) and tail (right) end. Unit cycles have a simpler characterization than well formed formulae and are readily enumerated. The "welding point" can be recovered, as each unit cycle determines a unique cutting place such that the stretched out string is a well formed formula.

2. *Resume.* The basic results of this paper are

THEOREM 1. *Each string of symbols satisfying $L - a = 1$, or, what is the same, $a = m + n$ has one and only one cyclic permutation for which it is a well formed formula; that is $a = m + n$ is a necessary and sufficient condition for a cyclic formula to be that of a well determined well formed formula.*

Here $L = n + m + 1$ is the length of the string, n the number of variables x_1, \dots, x_n ; $m + 1$ the number of operators g_0, \dots, g_m of "arities" a_0, \dots, a_m respec-

¹ The so-called Polish or Lukasiewicz notation, often somewhat misleadingly said to be "bracketfree" or "to rid us of the nuisance of brackets", provides a redundancy-free method of bracketing.

tively; $a = \sum_{j=0}^m a_j$ the total arity of the string. The theorem assures that each unit cycle admits a unique cut determining a well formed formula. This gives a simple method of producing well formed formulae, indeed *all* well formed formulae. The essential tools for assuring the unity of the cut are Lemma 2, which asserts the fundamental bracketing property, and its corollary—the natural partial order of brackets by inclusion. As a further corollary one obtains

THEOREM 2. *The number of distinct well formed formulae*

$$(*) \quad N_{g_0 \dots g_m}^{(n)} = \prod_{j=1}^m (n+j) / \prod_{f \in F} m_f!$$

where F is the set of operations and m_f the number of times the operator f appears among the operators g_j .

Theorem 1 has two distinct proofs: The first one, based on building up well formed formulae from shorter ones has already been described above. The other one is essentially a proposition about sequences of integers and their partial sums (Proposition 2) which, by the way, has obvious generalizations to finite sums of reals (Proposition 3) and integrals (Proposition 4); this method effectively determines the unique place of the cut, that is, the cyclic permutation in Theorem 1, by using the existence of partial sums having an extremal property.

3. Interpretation and Extensions. Traditionally F in (*) would be interpreted as the set of “fundamental” operations (the f 's not listed among the g_j having multiplicity $m_f = 0$); or, as the set of fundamental operations appearing among the g_j (all $m_f \geq 1$). However, the limitation to fundamental operations is irrelevant—everything applies to all operations, fundamental or derived (composed) ones, without distinction. Moreover, the interpretation as operations, i.e. as mappings $f: A^a \rightarrow A$ of the a -th power of a carrier (support) A into A ($a = a_f$ is the arity of f), is irrelevant too. Our theory is one of abstract groupings, packings, or, synonymously, of bracketings as purely formal operations.

Extension of these methods to languages with richer logical content, admitting a greater variety of symbols (predicates, quantifiers, allowing for the special role of “=”, etc.) is desirable and seems feasible. Further combinatorial and number theoretical results and applications to systems of linear inequalities, convex polyhedra etc., may be implied. One may also venture the suggestion of an analogue in theoretical physics: particle formation, by packing of sub-particles, and splitting of particles, considering well formed formulae built up from shorter ones as analogues of stable particles composed from smaller ones.

4. Background.—Formula (*) and many of its numerous interpretations, e.g. enumerating trees, in particular some of its special cases (binary bracketings or trees, or, what is the same, non-associative “powers”) are not new. What

seems to be new is the surprisingly simple Theorem 1, its elegant conceptual proof, as well as the more general interpretation.

This paper was motivated by the search for a simpler argument to replace the ingenuous, but long and difficult proof of a formula for the number of faces of given form of polyhedra \mathcal{P}_n due to Mme de Fougères ([3], section 3, Theorem 2).

The idea of circular permutations was conceived independently and earlier, perhaps for the first time, by Silberger [7]² in connection with the special problem of enumeration of binary bracketings. The method of partial sums with an extremal property implying also cyclic permutations, appears, probably for the first time in Raney's [5]³ proof of formula (*). The history of special cases and interpretations of this problem goes back for more than 200 years, at least till Euler. Closed enumeration formulae were first proven, mostly by function-theoretical methods, in the 19th century (Binet, Catalan, Lamé, Rodriguez). Cayley and Schröder considered certain variations on this problem. Formula (*) was obtained for the first time by Erdélyi and Etherington in 1940 [2] (by function-theoretical means). Further references can be found in the books and papers listed in our bibliography and in standard reference works.

2. Cyclic formulae

1. *Notations.* One denotes: x, x_1, \dots, x_n variables (constants which are the same as nullary operations, included) which are symbols of "arity" 0 (or *valency* or *weight* 1), chosen from a set X ;

$\{\dots, f, \dots\} = F = \bigcup_{a \in N'} F_a$ the set of functions or operations, F_a the set of those of arity a ($N' = \{1, 2, \dots\}$)⁴; g_0, \dots, g_m are $m + 1$ not necessarily distinct $f \in F$;

$Y = X \cup F$ set of symbols y , from which y_1, \dots, y_L are chosen; a_y arity of y ; $s = s^L = s_1^L = y_1 \dots y_j \dots y_L \in Y^L$ a (linear) string of length L , y_1/y_L its head/end (first/last symbol) $S = \bigcup_{L \in N} Y^L$ set of strings⁵; $L = n + m + 1$; $a_j = a_{g_j}$, $\sum_{j=1}^L a_{y_j} = \sum_{j=0}^m a_j = a$ the total arity of s ;

$v : S \rightarrow Z$ (integers) is the *valency* of strings defined by $v(y) = 1 - a_y$, $v(s) = \sum_{j=1}^L v(y_j) = L - a$.

Operators act on symbols to their right; e.g. the distributive law for rings $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$ will be written $\cdot x + yz = + \cdot xy \cdot xz$.

2. *Some Preliminaries on Well Formed Formulae.* A well known theorem of symbolic logic (Post Languages) and universal algebra (for example [1],

² The author is obliged to Professor J. Isbell for calling his attention to this recent paper.

³ The author is obliged to Professor Don Knuth for this reference, as well as that to his recent book [4].

⁴ Putting $X = F_0$ one replaces N' by $N = [0, 1, 2, \dots]$ and obtains $Y = F$.

⁵ Including the empty string s_0 .

p. 118; [6], pp. 152–157, p. 207) characterizes the subset $W, W \subset S$, of *meaningful* words (derived functions, polynomials) or well formed formulae (expressions substitutable for variables) by

$$s \in W \Leftrightarrow v(s_j) \leq 0 \quad (1 \leq j < L), \quad v(s^L) = 1.$$

This is a characterization by *initial* segments. Dually, their complements or *final* segments satisfy $v(Cs^j) = v(y_{j+1} \cdots y_L) \geq 1$ for all $j = 0, 1, \dots, L - 1$ with $Cs^0 = s = s^L$ and $v(s) = 1$.

Defining a function $u = u_s : J^L \rightarrow Z (J^L = \langle 1, \dots, L \rangle)$ assigning $j \mapsto u_s(j) = v(s_j)$ one reformulates this characterization as

PROPOSITION 1. $s \in W \Leftrightarrow u_s(j)$ takes the maximum value 1, and this only once at the “end” $j = L$.

Call a sequence $z_1 \cdots, z_L$ of integers ≤ 1 with sum = 1 a *unit sequence*, and call it *well balanced* if all properly partial, i.e. initial sums are ≤ 0 . One reformulates

PROPOSITION 1’. $s \in W \Leftrightarrow v(y_1), \dots, v(y_L)$ is a well balanced unit sequence.

3. *Cycles.* Bend s into an oriented *cycle* t with addition and subtraction of indices in $J^L \text{ mod } L$. *Substrings* (segments or arcs) of length $l \leq L$ are strings $s_i^{i-1+l} = y_i, \dots, y_{i-1+l}$, the “full” arcs $s_i^{i-1} (Vi \leq L)$ having length L .

The common valence of $v(s_i^{i-1})$ for all i is the valency $v(t)$ of t . All cycles considered shall have *positive valency* $v(t) > 0$. If $v(t) = 1$ call t a *unit cycle*.

If $s_i^{i-1+l} \in W$ call s_i^{i-1+l} an *elementary unit (e.u.)*.

LEMMA 1. Each y_i of t is head of a uniquely determined elementary unit $s_i^{i'}$ (of length $l_i = i' - i + 1 \geq 1$).

PROOF. a) $y_i \in X \Rightarrow y_i = s_i^i$ is the desired elementary unit with $l_i = 1$.

b) $y_i \in F \Leftrightarrow v(y_i) < 1, v(s_i^{i-1}) \geq 1 \Rightarrow \exists l (1 < l \leq L) | v(s_i^{i-1+l}) = 1;$

$$l_i = \min \{l | v(s_i^{i-1+l}) = 1\} > 1.$$

COROLLARY 1. t is covered by elementary units.

COROLLARY 2. $i \mapsto i'$ defines a well defined function $J^L \rightarrow J^L$.

LEMMA 2. (Fundamental bracketing property): *Elementary units do not overlap (properly); that is two distinct elementary units are either disjoint or one is contained in the other.*

PROOF. Assume e.u. $s_i^{i'}, s_j^{j'} i < j$, overlap properly, that is $i < j \leq i' < j'$. This implies $v(s_i^{i-1}), v(s_j^{j-1}) \leq 0$ and, therefore, the contradiction $1 = v(s_i^{i'}) = v(s_i^{j-1}) + v(s_j^{j'}) \leq 0$.

The following reformulation of this proof is even simpler and more conceptual: An intersection of two properly overlapping elementary units is simultaneously a final and an initial segment, that is of weight ≥ 1 and ≤ 0 which is absurd.

REMARK. There can be two intersections at once.

COROLLARY 3. *Among the L distinct full arcs s_i^{i-1} there is at most one elementary unit.*

PROOF. Any two among them overlap (even twice).

A necessary condition for t to have a full elementary unit is $v(t) = 1$; by Corollary 3 there can be at most one. One proves that indeed there always is one. Consider the system of all elementary unit arcs of t ordered by inclusion.

COROLLARY 4. *t is partitioned by its maximal elementary units.*

COROLLARY 5. *$v(t) = L - a =$ number of maximal elementary units of t .*

COROLLARY 6. *$v(t) = L - a = 1 \Leftrightarrow \exists i \mid s_i^{i-1} \in W$.*

This proves

THEOREM 1. *$v(t) = 1 \Leftrightarrow t$ has a full elementary unit (of course, only one). As $L = n + m + 1$ and $L - a = 1 \Rightarrow a = n + m$, $v(t) = 1 \Leftrightarrow a = n + m$. This gives the version of this theorem in the introduction. Other versions are:*

THEOREM 1^A. *Each unit cycle admits one and only one cut that is, determines a unique head at i such that $s = s_i^{i-1}$ is a well formed formula. In other words:*

For each string $s = y_1 \cdots y_L$ with $v(s) = 1$ there exists one and only one cyclic permutation such that $s_i^{i-1} = y_i \cdots y_L y_1 \cdots y_{i-1} \in W$.

How to find i ? If the cut falls between $y_{j-1}y_j, y_Ly_1$ included, then $y_{j-1} \in X$ and $y_j \in F$. If there is only one such place this one is it. Otherwise one can try such places one after the other. Proposition 2 in the next section gives an independent characterization and direct determination of i .

4. *Arithmetical Version and Generalizations.* An elementary arithmetical version of theorem 1 is

THEOREM 1^B. *Every unit sequence admits one and only one cyclic permutation making it well balanced.*

Proposition 1 and version B of Theorem 1 suggest the stronger Proposition 2 with an even simpler proof.

PROPOSITION 2. Let $z = z_1, \dots, z_L$ be a unit sequence, $i - 1$ the lowest index at which its (initial) partial sums take their maximal value: then $z_i, \dots, z_L, z_1, \dots, z_{i-1}$ is the unique well balanced sequence among the cyclic permutations of z .

PROOF. Put $M = \sum_{j=1}^{i-1} z_j$. Then $\sum_{j=1}^k z_j < M$ for all $k < i - 1$, $\sum_{j=1}^k z_j \leq M$ for all k .

Therefore $\sum_{j=i}^k z_j = \sum_{j=1}^k z_j - M \leq 0 \quad i \leq k \leq L$

$$\sum_{j=i}^L z_j + \sum_{j=1}^k z_j = 1 - M + \sum_{j=1}^k z_j \begin{cases} \leq 0, & k < i - 1 \\ = 1, & k = i - 1. \end{cases}$$

One obtains the uniqueness, as above from the fundamental bracket property (Lemma 2), that is, by Corollary 3; or independently, as follows:

For $l \neq i$ the sequence $z_l, \dots, z_L, z_1, \dots, z_{l-1}$ will not be well balanced because

for $l < i$: $\sum_{j=l}^{i-1} z_j = \sum_{j=1}^{i-1} z_j - \sum_{j=1}^{l-1} z_j > 0$,

for $l > i$: $\sum_{j=l}^L z_j + \sum_{j=1}^{i-1} z_j = 1 - \sum_{j=1}^{l-1} z_j + \sum_{j=1}^{i-1} z_j \geq 1$.

An obvious, but weak generalization is

PROPOSITION 3. Let r_1, \dots, r_L be a sequence of real numbers with positive sum $s > 0$ and $i - 1$ the first index at which the partial sums take their greatest value. Then all properly partial sums of the cyclic permutation

$$r_i, \dots, r_L, r_1, \dots, r_{i-1}$$

are $< s$; that is, its sequence of partial sums takes the smallest possible greatest value namely s , and this only once namely as the total sum.

REMARKS. The first $L - i + 1$ partial sums are even ≤ 0 . The i of the proposition is well defined, but there is no point in emphasizing uniqueness as no generalization of well balanced has been formulated. If all r_i are integral multiples of s and $\leq s$; one falls back to Proposition 2. Otherwise new, possibly richer possibilities arise.

An even wider generalization is

PROPOSITION 4. Let v be a real function in the interval $a \leq x \leq a + L$ with well defined integral

$$V(x) = \int_a^x v \text{ for all } x \ a \leq x \leq a + L$$

and $s = V(a + L) > 0$; let $b, a < b \leq a + L$, be the first (= smallest) x at which $V(x)$ reaches its greatest value. Put

$$u(x) = \begin{cases} v(x) & \text{for } b \leq x \leq a + L \\ v(x - L) & \text{for } a + L < x \leq b + L \end{cases}$$

and

$$U(x) = \int_b^x u \quad b \leq x \leq b + L.$$

$$\text{Then } U(x) = \begin{cases} < s & \text{for } x < b + L \text{ (even)} \\ \leq 0 & \text{for } x \leq a + L \\ = s & \text{for } x = b + L. \end{cases}$$

PROOF. Obvious.

3. Enumeration of well formed formulae

1. *First derivation.*—Consider x_1, \dots, x_n as n distinct points in cyclic order determining n places, place p being the open arc from x_{p-1} to x_p (for $p = 1$ from x_n to x_1). The operators g_0, \dots, g_m are points in these places: g_0 is in one of n possible places, dividing in into two new places; g_1 in one of the $n + 1$ places determined by x_1, \dots, x_n and g_0 , dividing again one of them into two; etc.; finally g_m can be placed in any of $n + m$ places determined by x_1, \dots, x_n and g_0, \dots, g_{m-1} . One has therefore

$$(1) \quad n(n + 1) \cdots (n + m) = \prod_{j=0}^m (n + j)$$

possible arrangements of cyclic formulae with $m + 1$ operators with respect to the fixed cycle (x_1, \dots, x_n) .

Cycles, distinct only by permutations of places of repeated appearances of the same operator, are identical. If $f \in F$ appears m_f times among the g ($\sum_{f \in F} m_f = m + 1, \sum_{f \in F} a_f m_f = a$), the number of distinct cycles is

$$(2) \quad \prod_{j=0}^m (n + j) / \prod_{f \in F} m_f!$$

If $a = n + m$, the cycles are unit cycles, each determining a unique well formed formula with a certain $x_i (i = 1, \dots, n)$ as its first variable $\in X$. Therefore $N_{g_0 \dots g_m}^{(n)}$, the number of well formed formulae with indeterminates x_1, \dots, x_n in this order, i.e. with $i = 1$, is $1/n$ of (2):

THEOREM 2.

$$(*) \quad N_{g_0 \dots g_m}^{(n)} = \prod_{j=1}^m (n + j) / \prod_{f \in F} m_f!$$

2. *Alternative derivation of formula (*)*. The number of strings in $L = n + m + 1$ distinct symbols $x_1, \dots, x_n, g_0, \dots, g_m = (n + m + 1)!$

The n variables are interchangeable as their notation or enumeration is indifferent; so are the m_f operators g_j equal f . Therefore there are only $(n + m + 1)! / n! \prod m_f!$ distinct strings. One and only one among the $n + m + 1$ cyclic permutations of each string is a well formed formula. Therefore one has precisely

$$N_{g_0 \dots g_m}^{(n)} = (n + m)! / n! \prod m_f! = \prod_{j=1}^m (n + j) / \prod_{f \in F} m_f!$$

distinct well formed formulae as above.

3. *Applications of (*)*. Formula (3) is nearly identical with that of Theorem 2 of [3];

$$(**) \quad f_{d_0 \dots d_i}^{(n)} = \prod_{j=1}^m (n + 2 + j) / \prod_{j=0}^{\infty} t_j!$$

where t_j is the number of $d_k = j$, that is, $\sum_{j=0}^{\infty} t_j = i + 1$ and $\sum_{j=1}^{\infty} t_j = n - i$. (**) is (*) with $F = \{f \mid j = 2, 3, \dots\}$, f just the ‘‘universal’’ j -ary operation (ignoring the unary identity map f_1). In (**) n is replaced by $n + 2$, m by i , m_{j+2} by t_j , and g_k by d_k with $a_k = d_k + 2$.

REMARK. One interprets the *universal* j -ary operation f_j as ‘‘ j -ary brackets’’ (this, indeed, is meant in [3]), i.e. the formation of j -uples. One extends the carrier A to $A \cup A \times A \cup \dots = \bigcup_{j=1}^{\infty} A^j = \mathfrak{A}$; f_1 is extended from A to \mathfrak{A} , and so are all f_j .

One obtains the number of binary bracketings k_n by putting $F = \{f\}$, $f = f_2$ the universal binary operation,

$$\sum m_f = m_f = m + 1, a = 2m_f = n + m \rightarrow n + m = 2n - 2,$$

that is

$$k_n = N_{\underbrace{2 \dots 2}_{m+1 = n-1 \text{ times}}}^{(n)} = (n + m)! / n! \prod m_f! = \frac{(2n - 2)!}{(n - 1)! n!} = \binom{2(n - 1)}{n - 1} / n.$$

Bibliography

- [0] A. Cayley: ‘On the analytical forms called trees’, 2nd part, *Coll. Math. Papers*, vol. 4, Cambridge 1891, 114–5.
- [1] P. M. Cohn: *Universal Algebra*, (Harper and Row, 1965).
- [2] A. Erdélyi and I. H. M. Etherington: ‘Some problems of non-associative combinations (2)’, *Edinburgh Math. Notes*, 32 (1940), VII–XII.
- [3] D. de Fougères and D. Tamari: ‘La structure polyédrale des complexes de parenthésages’, (submitted).

- [4] Don Knuth: *Fundamental Algorithms*, (Addison-Wesley, 1969.)
- [5] G. N. Raney: 'Functional composition patterns and power series reversion', *Trans. Amer. Math. Soc.* 94 (1960), 441–51.
- [6] P. C. Rosenbloom: *The elements of mathematical logic*. (Dover 1950.)
- [7] D. M. Silberger: 'Occurrence of the integer $(2n-1)!/n!(n-1)!$ '; *Prace Mat.* 13 (1969), 91–96; M. R. 40 # 2556.
- [8] D. Tamari: *Monoides préordonnés et chaînes de Malcev*; Thèse, (Paris 1951) (printed part in *Bull. Soc. Math. France* 82 (1954), 53–96).
- [9] D. Tamari 'The algebra of bracketings and their enumeration', *Nieuw Arch. Wisk.* X (1962), 131–146.

Department of Mathematics
State University of New York at Buffalo
4246 Ridge Lea Road, Amherst, N.Y. 14226
U.S.A.