COMBINATORIAL THEORY, OLD AND NEW

by Gian-Carlo ROTA

1. Introduction.

Combinatorial analysis, or combinatorial theory, as it has come to be called, is currently enjoying an outburst of activity. This can be partly attributed to the abundance of new and highly relevant problems brought to the fore by advances in discrete applied mathematics, and partly to the fact that only lately has the field ceased to be the private preserve of mathematical acrobats, and attempts have been made to develop coherent theories, thereby bringing it closer to the mainstream of mathematics. By way of example of this second trend I shall sketch the outlines and prospects of perhaps one of the most promising of such infant theories: the theory of combinatorial geometries. This theory is historically rooted in the four-color conjecture much like algebraic number theory was born out of Fermat's conjecture. Indeed, its main achievement to-date is not only a substantial body of results with disparate applications, but, we hazard to state, its success in displaying the four-color problem in a new light, where it appears as merely one case of a general problem leading to a host of analogous problems of varying difficulty (the critical problem, §4). The interaction of techniques and insights which results from the simultaneous study of these problems offers—if we are to believe the so-called "lessons of the past"—the brightest hope for the future of the theory.

2. Basic Notions.

Combinatorial geometries are the creation of several mathematicians. Because of limitations of space, we omit all names, but we cannot pass under silence those of the two workers on whose pioneering shoulders the body of the theory rests: Hassler Whitney and W.T. Tutte. A combinatorial geometry (c.g.) \( G = G(S) \) is a set \( S \) (hereafter assumed finite for simplicity) together with a closure relation \( A \rightarrow \overline{A} \) defined on all subsets \( A \) of \( S \) (i.e., \( \overline{\overline{A}} = A, A \subseteq B \) implies \( \overline{A} \subseteq B \) and \( A \subseteq \overline{\overline{A}} \), but not \( \overline{A} \cup \overline{B} = \overline{A \cup B} \)) satisfying (1) \( \emptyset = \emptyset \) for the empty set, (2) \( \overline{p} = p \) for every element \( p \in S \), (3) the exchange property: if \( p, r \in S \) and \( A \subseteq S \), and if \( r \in \overline{A \cup p} \) but \( r \notin A \), then \( p \in \overline{A \cup r} \). The closed sets or flats \( (A=\overline{A}) \) form a geometric lattice (g.l.). A subset \( I \subseteq S \) is independent if \( p \notin I - p \) for all \( p \in I \); all maximal independent sets, or bases, have the same cardinality (hereafter denoted by \( n \)). There are equivalent axioms for a c.g. in terms of the family of independent sets and bases.

Dropping conditions (1) and (2) one obtains a pregeometry or matroid. To every pregeometry one can canonically associate a geometry, and we often use
the two terms interchangeably. The **direct sum** of two geometries is defined in the obvious way, as is the **restriction** of a geometry to a subset $A$; the cardinality of a basis of the restriction to $A$ is the **rank** $r(A)$ of $A$, satisfying

$$r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$$

and

$$r(A \cup p) = r(A) + 0 \quad \text{or} \quad r(A) + 1$$

(this gives another axiom system for c.g. 's). The numbers $W^{(k)}$ of closed sets of rank $k$ are the **Whitney numbers (of the second kind)**. An open problem is whether the Whitney numbers always form a **unimodal sequence**. Let $\mu$ be the Möbius function of the geometric lattice : the numbers

$$W_k = \{ \Sigma \mu(0, \bar A) : A = \bar A \text{ and } r(A) = k \}$$

are the **Whitney numbers (of the first kind)**. $p_G(\lambda) = \sum_{k=0}^{n} W_k \lambda^{n-k}$ is the characteristic (or Birkhoff) polynomial of the geometry. The **critical problem** (see §4) is the problem of locating the positive integer zeros (especially the largest) of the characteristic polynomial. A **strong map** $f$ of $G(S)$ to $G(T)$ is a function from $S$ to $T \oplus 0$ (where $0$ is a "dummy element" added to $T$) such that the inverse image of a closed set is a closed set. Strong maps are the morphisms in the category of c.g. 's. A **contraction** by a subset $A$ is the (pre-) c.g. on $S - A$ whose closed sets $C$ are those for which $A \cup C$ is closed in $G(S)$. Every strong map can be factored into a monomorphism followed by a contraction. A **modular flat** $A$ satisfies

$$r(A \cup B) + r(A \cap B) = r(A) + r(B)$$

for all flats $B$. The **orthogonal geometry** $G^\perp(S)$ of $G(S)$ is the unique geometry whose bases are those subsets of $S$ whose complement is a basis of $G(S)$.

3. **Typical Examples.**

The notion of c.g. is abstracted from the properties of linear dependence in projective space $P$. In fact, the typical examples are obtained by taking a subset $S$ of $P$ and setting $\bar A = \text{sp}(A) \cap S$ for $A \subseteq S$, where sp$(A)$ is the linear variety in $P$ spanned by the subset $A$. A host of notions of synthetic projective geometry can be carried over to c.g. 's by this analogy. Conversely, the **coordinatization** or **representation problem** (v. §5) asks for conditions on a c.g. that it be representable as a subset of projective space over a given field (or even more generally) in the manner just described. Most geometries are not representable over any field; e.g., take the direct sum of a projective space over $GF(2)$ and one over $GF(3)$.

Given an Abelian group $A$, a subgroup $V$ of the Abelian group of all functions from a set $S$ to $A$ can be used to define the **function space geometry** of $V$ : for $A \subseteq S$, set $\bar A = \{ p \in S : f(p) = 0 \text{ for all } f \in V \text{ s.t. } f(a) = 0 \text{ for all } a \in A \}$. For example, if $S$ is the set of edges of an oriented linear graph, one can take $V$ as the set of all 1-coboundaries ; in the resulting **bond geometry** independent sets are trees and bases are spanning trees. The orthogonal c.g. to the bond g. (the **circuit** g.) is obtained by taking $V$ to be all 1-cycles. Much of classical graph theory can be obtained by applying general results on c.g. 's to these two special cases. If one starts with the complete graph on $j$ vertices, the bond-geometric...
lattice is the lattice of partitions of a \( j \)-element set, whose Whitney numbers are the Stirling numbers. (of the first and second kind).

Passing from graphs to finite simplicial complexes, let \( S = T_k \) be the set of all \( k \)-element subsets of a set \( T \). For \( A \subseteq T_k \), set \( \Sigma(A) = T_0 \cup T_1 \cup \ldots \cup T_{k-1} \cup A \), and let \( \beta_i \) be the \((i-1)\)-dimensional Betti no. of the simplicial complex \( \Sigma(A) \). Setting \( r(A) = \binom{n-1}{k-1} - \beta_{k-1}(\Sigma(A)) = \nu(A) - \beta_k(\Sigma(A)) \), where \( \nu(A) \) is the cardinality of \( A \), we obtain the rank function of a c.g., the simplicial geometry of dimension \( k \) on \( T \). It can be used to extend notions of graph theory to higher dimensions; its orthogonal geometry is connected with Alexander duality. A slight extension of this construction leads to a class of geometries on triangulable manifolds whose Whitney nos. give a new set of topological invariants, not invariant under homotopy.

A submodular set function \( \mu \) on \( S \) satisfies: (1) \( \mu(A) \) is an integer for all \( A \subseteq S \); (2) \( \mu(A \cup B) + \mu(A \cap B) \leq \mu(A) + \mu(B) \); (3) If \( A \subseteq B \), then \( \mu(A) \leq \mu(B) \).

The family of all sets \( I \) s.t. \( \nu(A) \leq \lambda(A) \) for all non-empty subsets \( A \subseteq I \) is the family of all independent set of a c.g. In this way, one can construct a wide variety of c.g. ’s which are a long way from being representable as subsets of projective space. One can for example set \( \mu \) as a linear combination with positive integer coefficients of rank functions of several geometries defined on the same set. These c.g. ’s can be used to give a unified and widely extended treatment of matching, covering, minimax theorems, etc.

A connection with integer programming is obtained through unimodular geometries, which are integer-valued function space geometries \( V \) on \( S \) s.t. if \( f \in V \) has minimal support, then there is a \( g \in V \) with the same support as \( f \) and taking only the values \( \pm 1 \) and 0.

4. The Critical Problem.

For c.g. ’s \( G(S) \) which are subsets \( S \) of projective space \( P \) of dim \( n \) over \( GF(q) \), the smallest integer \( c \) (the critical exponent) such that \( p_G(q^c) > 0 \) equals the smallest number of hyperplanes \( H_1, \ldots, H_c \) in \( P \) such that every \( p \in S \) does not belong to at least one such \( H_i \). If \( S \) is the set of all points with \( d \) or fewer non-zero coordinates, this is the classical problem of finding linear codes. Taking \( G(S) \) to be all vectors in a vector space over \( GF(5) \) whose coordinates are \( \pm 1 \) or 0, the critical exponent is the capacity of the 5-graph. A host of other classical problems are thus seen to be critical problems. For a function space geometry with values in \( A = GF(q) \), the critical exponent is the smallest number of a set of functions \( f_1, \ldots, f_c \) such that for every \( p \in S \), at least one \( f_i(p) \neq 0 \). For the bond geometry, the critical exponent is \( c \) whenever the graph is colorable in \( q^c \) colors. For a unimodular geometry, the smallest positive integer \( n \) such that \( p_G(n) > 0 \) is the smallest \( n \) for which there is a non-vanishing function \( f \in V \) taking \( n - 1 \) values. For the circuit geometry of a graph with integer values, this reduces to the problem of minimum flows; it is conjectured that \( n = 6 \) for all graphs (a far more interesting conjecture than the four-color conjecture).

A study of the characteristic polynomial for general c.g. ’s has already yielded a crop of results, of which we shall quote only two. A c.g. is supersolvable when
there exist modular flats $A_0 \subseteq A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n = S$ s.t. $r(A_i) = i$. The roots of the characteristic polynomial of a supersolvable c.g. are positive integers, equal to $v(A_i + 1 - A_i)
A loop of a pregeometry $G(S)$ is a point $p$ s.t. $p \in \bar{\Omega}$; a link is a point $p$ s.t. $G(S)$ is the direct sum of $S - p$ and $p$. Take the free polynomial ring generated by all isomorphism classes of pregeometries, divide it by the ideal generated by:
(a) $G(S) - G(T)G(V)$ whenever $S$ is the direct sum of $T$ and $V$;
(b) $G(S) - G(S - p) - G(S/p)$ whenever $p$ is not a loop or a link, where $G(S - p)$ is the subgeometry on $S - p$ and $G(S/p)$ is the contraction by $p$. The resulting Tutte-Grothendieck ring has the remarkable property of being isomorphic to a polynomial ring in two generators without constant terms. By evaluating the polynomial $t_G(z, x)$ associated in this way to $G(S)$, one can compute all “invariants”, for example $t_G(1 - \lambda, 0) = \pm p_G(\lambda)$ (the characteristic polynomial); $t(1, 1) =$ no. of bases; $t(2, 1) =$ no. of independent sets $t(1, 0) =$ Möbius function $|\mu(\bar{\Omega}, S)|$, etc. It appears that classical “reductions” in the coloring problem can be systematically studied—and extended to the critical problem generally—by algebraic methods through the $T$-G-ring (and other categorical structures currently being developed). This leads to the conjecture (supported by the results in §5) that the critical exponent is related to the absence of certain “forbidden segments”, or “obstructions” in the g.l. of $G(S)$ or in the g.l. of $G^1(S)$. Hadwiger’s conjecture can be restated in these terms.

The Bose-Segre problem of finding in $P$ over $GF(q)$ maximum-sized subsets $B$ with the property that no $k$-subset of $B$ is linearly dependent is also closely related to a suitable critical problem and opens interesting connections with algebraic geometry over finite fields.

5. Representation.

A c.g. is representable as a subset of a projective space over $GF(2)$ iff no segment $[A, B]$ of length 2 (i.e., $r(B) - r(A) = 2$) in the g.l. contains more than five flats. We abbreviate this by “the 4-point line is the only obstruction for representation over $GF(2)$”.

In the same vein, the five-point line and its orthogonal c.g. and the Fano plane and its orthogonal c.g. are the only obstructions for representation over $GF(3)$.

There probably is a finite number of obstructions for representations over any given finite field, although their classification is far from complete. The obstructions for representation as a unimodular c.g. are the 4-pt. line, the Fano plane and its orthogonal c.g., and for representation as a bond-g. they are the same plus the circuit g. 's of the two Kuratowski graphs; finally, for representation as the bond g. of a planar graph they are the above, plus the bond-g. 's of the two Kuratowski graphs (a crowning achievement of Tutte). Thus Kuratowski’s theorem, which has not been satisfactorily explained by topology, finds itself in pleasing company in c.g.

The preceding results lead to the question: is there a “universal” algebraic structure which can “represent” every c.g.? An algebra of syzygies of rank $n$
over an Abelian group $A$ consists of a set $S$, together with a map $T$ of pairs of ordered $n$-tuples of $S$ into $A$, satisfying the following identities:

1. If $\sigma$ and $\tau$ are permutations of $(1, 2, \ldots, n)$, then

$$T(x_{\sigma 1}, x_{\sigma 2}, \ldots, x_{\sigma n} | y_{\tau 1}, y_{\tau 2}, \ldots, y_{\tau n}) = \pm T(x_1, \ldots, x_n | y_1, \ldots, y_n)$$

according as the product $\sigma\tau$ is even or odd ($x_i \in S$ and $y_i \in S$).

2. $T(x_1, x_2, \ldots, x_n | y_1, y_2, \ldots, y_n)$

$$= \sum_{i=1}^{n} T(x_1, x_2, \ldots, x_{i-1}, y_1, x_{i+1}, \ldots, x_n | x_i, y_2, y_3, \ldots, y_n)$$

Every algebra of syzygies defines a standard c.g. by taking the bases to be those $n$-tuples $(x_1, x_2, \ldots, x_n)$ for which $T(x_1, \ldots, x_n | y_1, \ldots, y) \neq 0$ or

$$T(y_1, \ldots, y_n | x_1, \ldots, x_n) \neq 0$$

for some $n$-tuple of $y$'s. Every c.g. representable in projective space is standard: set $T(x_1, \ldots, x_n | y_1, \ldots, y_n) = \det (x_1, \ldots, x_n) \det (y_1, \ldots, y_n)$ relative to a fixed basis. More generally, to every c.g. one can canonically associate a standard c.g. which is, in a precise sense, universal relative to all representations. Several classical theorems of projective geometry, such as Desargues and Pappus, can be interpreted as conditions on the algebra of syzygies, allowing stronger representation theorems. The algebra of syzygies makes contact with the classical invariant theory of finite point sets in projective space, of which c.g. is an abstraction. The interaction and confluence of such manifold trends and problems is perhaps the most promising feature of this field of investigation.

**BIBLIOGRAPHY**

All bibliographical and historical references will be found in


In addition, I have drawn from unpublished work of T. Brylawski, R. Reid, W. Whiteley, N. White, R. Stanley, T. Helgason and my own.

2-347/M.I.T

Cambridge, Massachusetts 02139

U.S.A