The geometric approach to matroid theory

JOSEPH P. S. KUNG

1. Preamble

The aim of this article is to describe Rota’s approach to matroid theory and to outline the developments directly inspired by this special approach. It is not intended as a balanced or exhaustive account of the subject; indeed, in Rota’s spirit, it advocates a particular philosophical standpoint. A more conventional survey can be found in [Kung 1995].

2. Three Ways of Doing Matroid Theory

At the risk of oversimplification, there are three “ways” of doing matroid theory. In no particular order, these are the graph-theoretic way, the algorithmic way, and the geometric way. The graph-theoretic way approaches matroid theory as a generalization of graph theory and uses ideas, such as circuits and connectivity, from graph theory. It has its origins in Whitney’s attempt [1932a] to extend the construction of the dual graph of a planar graph to arbitrary graphs. He found that a dual does exist, only that for non-planar graphs, the dual is a matroid which cannot be represented by a graph. The algorithmic way is founded on the phenomenon (see, for example, [Faigle 1987]) that optimization problems for which the “greedy” algorithm produces an optimal solution have underlying combinatorial structures similar to a matroid.

The geometric way, the main subject of this article, views matroid theory as an abstract version, having the “correct” generality, of linear algebra and synthetic geometry. This approach can be traced back to three early papers:

1. Whitney’s 1935 paper, in which a matroid is defined for the first time,
2. Birkhoff’s 1935 paper, in which a geometric lattice is identified as the primary object of study in geometry, and
3. Mac Lane’s 1936 paper, in which parts of the dividing line between classical projective geometry with coordinates over a skew field and the combinatorial geometry of matroid theory is marked out through several examples.

The geometric way was eclipsed in the 1950’s and early 1960’s by the spectacular achievements of Tutte who used graph-theoretic ideas to prove several hard results, such as forbidden-minor theorems for binary, regular, and graphic matroids [Tutte 1958 and 1959]. It was only in the late 1960’s and 1970’s that geometry came again to share the forefront of matroid theory. Much of this is due to the advocacy
of Gian-Carlo Rota who renewed the geometric way with insights from nineteenth century invariant theory and synthesized it with the lattice-theoretic approach.

To Rota, matroid theory is the twentieth-century combinatorial continuation of the related nineteenth-century subjects of projective invariant theory and synthetic geometry. Thus, it is a subject in the central tradition of mathematics, rather than, as many still think, a narrow and eccentric subject. His advocacy of this point of view takes its most visible form in the book “Combinatorial geometries” written with Henry Crapo [Crapo and Rota 1970] and two brief survey papers [Rota 1971a, Kelly and Rota 1973]). His lectures at Bowdoin College in 1971 are equally important and we have, although in a rough and unfinished form, notes of those lectures [Rota 1971b]. These informal notes preserved the lively spirit of that conference and we shall quote from these notes later in this article.

Rota was also unhappy with the neologism “matroid” and proposed the alternative of “combinatorial (pre)geometry.” This proposal has not gained common acceptance. In this article, we adopt the sensible compromise of retaining “matroid” but using “geometry” to mean a simple matroid.

3. Influences

Much of Rota’s influence on matroid theory lies in his conjectures and problems. Many of Rota’s conjectures are broad and sweeping, seemingly made with only the slenderest of evidence. He proposed them to identify central problem areas, to guide research along the important channels, and to provide signposts by which progress in “understanding” can be measured. The last motivation is particularly important in combinatorics, where the problem-solving approach often obscures the structure of the subject. We shall attempt to describe briefly this aspect of his influence in this section.

3.1. The unimodality conjecture.

The $k$th Whitney number $W_k$ (of the second kind) of a geometry $G$ is the number of rank-$k$ flats in $G$. In [Harper and Rota 1970] and [Rota 1971], Rota made the following conjecture.

3.1. CONJECTURE. Let $G$ be a rank-$n$ geometry. Then the sequence $W_0, W_1, W_2, \ldots, W_n$ is unimodal, that is, there is a rank $s$ such that

$$W_0 \leq W_1 \leq W_2 \leq \ldots \leq W_s \text{ and } W_s \geq W_{s+1} \geq W_{s+2} \geq \ldots \geq W_n.$$  

Equivalently, for all $k$, $1 \leq k \leq n - 1$,

$$W_k \geq \min\{W_{k-1}, W_{k+1}\}.$$  

This conjecture follows from a stronger conjecture – the logarithmic convexity conjecture – made by Mason in [1972]:

$$W_k^2 \geq W_{k-1}W_{k+1}.$$  

605
In making his conjecture, Rota was motivated partly by three classical examples of geometric lattices, Boolean algebras, subspace lattices and partition lattices. But, less obviously, he was also motivated by inequalities for *quermassintegrals* in convex geometry. As this is unfamiliar territory for matroid theorists, we give a brief non-technical description (see [Sangwine-Yager 1993] for a definitive account). The *Minkowski sum* and *scalar product* is defined on convex sets as follows. If $A$ and $B$ are convex sets in Euclidean $n$-space $\mathbb{R}^n$ and $\lambda$ is a real number, then

$$A + B = \{ x + y : x \in A \text{ and } y \in B \}$$

$$\lambda A = \{ \lambda x : x \in A \}.$$

If $A$ is a convex set and $B$ is the unit ball, then the volume $V(A + \lambda B)$ is a polynomial of degree $n$ in $\lambda$:

$$V(A + \lambda B) = \sum_{k=0}^{n} \binom{n}{k} W_k(A) \lambda^k.$$

The (normalized) coefficients $W_k(A)$ are the quermassintegrals of $A$. The $k$th quermassintegral is a measure of the "$k$-dimensional volume" of a convex set and thus, $W_k(A)$ is somewhat analogous to the $k$th Whitney number $W_k$. It is known that the quermassintegrals form a logarithmically convex sequence.

Not much progress has been made on these conjectures. For both conjectures, only partial results are known. For the logarithmic convexity conjecture, the first nontrivial case, $k = 2$, is known as the "points-lines-planes conjecture". Currently the best result known is due to Seymour [1982]. He proved that

$$W_2^2 \geq \frac{3}{2} \left( \frac{W_1 - 1}{W_1 - 2} \right) W_1 W_3$$

for a geometry in which every line contains at most four points. Work on this conjecture is surveyed in the article by Griggs in Chapter 8. Results on logarithmic concavity of Whitney numbers for supersolvable upper homogeneous geometric lattices can be found in [Damiani et al. 1994].

For the unimodality conjecture, the two smallest non-trivial cases, $n = 3$ and $n = 4$, follow from the classical inequality $W_1 \leq W_2$. This inequality is almost classical and has been rediscovered several times. See [Kung 1986] for an extensive list of references. The next case, $n = 5$, would follow from the inequality $W_2 \leq W_3$ in any geometry of rank at least 5. I have proved this inequality for any geometry in which all lines have the same number of points.

There is at present a similar lack of progress in almost all the unimodality conjectures involving general classes of combinatorial objects, although there has been much progress on unimodality of specific sequences in combinatorics (see, for example, the survey [Stanley 1991]). Some doubts have been cast on these conjectures by Björner's counterexamples [1980] showing that the $f$-vector $(f_0, f_1, \ldots, f_n)$ of an $n$-dimensional convex polytope need not be unimodal. Here, $f_k$ is the number of $k$-dimensional faces.
3.2. Finite forbidden minors.

Another of Rota's conjectures is that the minor-closed class \( \mathcal{L}(q) \) of all geometries representable over the finite field \( GF(q) \) can be characterized by a finite set of forbidden minors.

3.2. CONJECTURE. There exists a finite set of geometries \( \{E_1, E_2, \ldots, E_m\} \) with the following property: if a geometry \( G \) does not contain any minor isomorphic to any of the geometries \( E_k \), then \( G \) can be represented over \( GF(q) \).

When Rota conjectured this in [1971] and [1972], the only value of \( q \) for which the conjecture is proved is \( q = 2 \). This case is the theorem of Tutte in [1957] that a geometry is binary if and only if it has no 4-point-line minor. In the same paper, Tutte proves a much harder theorem which says that a matroid is representable over every field if and only if it contains no minors isomorphic to the 4-point line, the Fano plane \( F_7 \), or its dual \( F_7^\perp \). This provides a finite forbidden minor characterization for the intersection \( \cap \mathcal{L}(p) \) of \( \mathcal{L}(p) \) over all primes.

Motivated by Rota's conjecture, R. Reid found the following set of forbidden minors for the class \( \mathcal{L}(3) \) of ternary geometries: the 5-point line, its dual \( U_{3,5} \) consisting of five points in general position in a plane, \( F_7 \), and \( F_7^\perp \). Reid never published his proof, but it is probable that it was based on Tutte's homotopy theorem and similar to Bixby's proof [1975]. The case of ternary geometries was independently pursued by matroid theorists in Britain. Seymour [1979] found another proof and his deep and powerful methods revolutionized matroid theory in the 1980's. There are now several proofs for the theorem on ternary geometries. See [Kahn 1984, Kahn and Seymour 1988, Truemper 1982].

3.3. The Laplace expansion and the bracket ring.

Rota's program for tackling Conjecture 3.2 is radically different. His starting point is classical projective invariant theory. According to Felix Klein's Erlanger Programm, projective geometry is the study of "invariant algebraic properties" of a vector space. To be more specific, let \( V \) be a \( d \)-dimensional vector space over a field \( F \). Fix a coordinate system of \( V \). A function \( p(x_1, x_2, \ldots, x_n) \) on \( n \) vector variables taking values in \( F \) is said to be polynomial if it can be written as a polynomial function in the coordinate variables \( x_{11}, x_{12}, \ldots, x_{1d}, x_{21}, x_{22}, \ldots, x_{2d}, \ldots, x_{n1}, x_{n2}, \ldots, x_{nd} \) corresponding to the vector variables \( x_1, x_2, \ldots, x_n \). The property of being polynomial is not dependent on the choice of coordinate system. A property \( P \) is said to be algebraic if it can be expressed in terms of the vanishing or non-vanishing of polynomial functions, that is, if there exists a first-order sentence \( S \) made up of atomic formulas of the form:

\[
p_i(x_1, x_2, \ldots, x_n) = 0
\]

\[
q_j(x_1, x_2, \ldots, x_n) \neq 0
\]

such that for any \( n \)-tuple of vectors \( v_1, v_2, \ldots, v_n \), the property \( P \) is true if and only if the sentence \( S \) is true under the substitution \( x_k = v_k \).
A property $P$ is said to be invariant if its validity is not dependent upon the choice of coordinate systems, or, more precisely, if for any non-singular linear transformation $A$, $P$ is true for $v_1, v_2, \ldots, v_n$ if and only if $P$ is true for $Av_1, Av_2, \ldots, Av_n$. A polynomial $p(x_1, x_2, \ldots, x_m)$ is said to be a (relative) invariant if for any non-singular linear transformation $A$,

$$p(Ax_1, Ax_2, \ldots, Ax_m) = (\det A)^g p(x_1, x_2, \ldots, x_m)$$

for some non-negative integer $g$. Because $\det A \neq 0$ for any non-singular linear transformation $A$, invariant algebraic properties are precisely those properties expressible by first-order sentences with invariant polynomials $p_i$ and $q_j$. The two fundamental theorems of classical projective invariant theory yield many of the key properties of invariants. The first fundamental theorem says that every invariant polynomial can be expressed as a homogeneous polynomial in the brackets or determinants:

$$[x_{i_1}, x_{i_2}, \ldots, x_{i_d}] = \det(x_{i,j,k}),$$

where $x_{i,j} = (x_{i_1}, x_{i_2}, \ldots, x_{i_d})$ in some coordinate system. The brackets satisfy two types of relations or syzygies. The first is the familiar alternating property of determinants. It has the form:

$$[x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(d)}] = \text{sgn} (\sigma) [x_1, x_2, \ldots, x_d],$$

where $\sigma$ is a permutation of $\{1, 2, \ldots, d\}$. The second is the Laplace expansion and has the form:

$$[x_1, x_2, \ldots, x_d][y_1, y_2, \ldots, y_d] = \sum_{i=1}^{d} [x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d][x_i, y_2, \ldots, y_d]$$

The second fundamental theorem says that all the algebraic relations between brackets are algebraic consequences of syzygies of type (1) and (2).

Rota described the importance of the Laplace expansion for matroid theory in his Bowdoin lectures in the following way:

Conceive of an abstract set of "vectors" $S$ and a collection of subsets of that set called "bases," each containing exactly $d$ "vectors." This is a typical setup for the definition of a combinatorial geometry. What does (2) say in this context? Note that for two "bases" $x_1, \ldots, x_d$ and $y_1, \ldots, y_d$, the left-hand side will be nonzero. Thus, the right-hand side is also nonzero and this means that at least one of the summands is nonzero. But then this says that there exists some $i$ such that $x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_d$ and $x_i, y_2, \ldots, y_d$ are also "bases" of $S$. This, of course, is nothing else but a statement of the exchange property. So we have seen that (2) is a coordinatized version of the exchange property — a fact which suggests that (2) must play a crucial rôle in the whole question of representation (i.e., coordinatization) of geometries.

---

1This working hypothesis is sometimes known as Gram's theorem.
One way of formalizing this insight is through the bracket ring, first defined in the Bowdoin lectures. Let $G$ be a rank-\(d\) matroid on the set $S$. The bracket ring $B(G)$ (over the integers) is constructed as follows. Take the polynomial ring over $\mathbb{Z}$ in the formal variables $[x_1, x_2, \ldots, x_d]$ where $(x_1, x_2, \ldots, x_d)$ ranges over all $d$-tuples of elements from $S$. Let $I$ be the ideal generated by relations of the form (1), (2), and

$$[x_1, x_2, \ldots, x_d] = 0 \quad \text{if } \{x_1, x_2, \ldots, x_d\} \text{ is not a basis of } G.$$

The bracket ring is studied in [White 1975]. (See also [White 1977, Sturmfels 1989].) He proved that those ideals $I$ in $B(G)$ such that $R/I$ is an integral domain are in one-to-one correspondence with representations of weak-map images or specializations of $G$. Thus, knowing the algebraic structure of bracket rings would yield information about representations. In particular, Rota hoped that for a given prime power $q$, the bracket ring $B(G)$ of a geometry $G$ in $\mathcal{L}(q)$ would have reasonable properties — for example, $B(G)$ might have tractable resolutions. If so, then information can be derived using commutative algebra. Before this program can be realized, methods for explicit computations have to be developed. Work in the past twenty years, most of it based on Rota's work on the straightening algorithm and standard bases, have progressed to the point where it is worthwhile to take up this program again (see, for example, [Anick and Rota 1991, Crapo 1993]).

3.4. Basis exchange properties.

Rota's derivation of the basis exchange axiom from the Laplace expansion can be extended to yield a procedure for deriving a (basis) exchange property from any identity between brackets in the following way. If

$$[X][Y] \ldots [Z] = [x_1, x_2, \ldots, x_d][y_1, y_2, \ldots, y_d] \ldots [z_1, z_2, \ldots, z_d]$$

is a bracket monomial, then $[X][Y] \ldots [Z] \neq 0$ if and only if all the sets in the collection $\{\{x_1, x_2, \ldots, x_d\}, \{y_1, y_2, \ldots, y_d\}, \ldots, \{z_1, z_2, \ldots, z_d\}\}$ are bases. We say that such a collection is basic. If

$$a_1[X_1][Y_1] \ldots [Z_1] + a_2[X_2][Y_2] \ldots [Z_2] + \ldots + a_m[X_m][Y_m] \ldots [Z_m] = 0$$

is a bracket identity with nonzero coefficients $a_i$, then none or at least two of the monomials are nonzero. This translates into the exchange property: Either none of the collections $\{X_1, Y_1, \ldots, Z_1\}, \{X_2, Y_2, \ldots, Z_2\}, \ldots, \{X_m, Y_m, \ldots, Z_m\}$ is basic or at least two are basic. In this way, every bracket identity translates into a exchange property which may or may not be true in all matroids. The question then arise. Which exchange properties hold for all matroids? The same question can be asked for classes of matroids, such as binary or regular matroids.

Rota's program for answering these questions, modelled on mathematical logic and proof theory, is to set up a system of deduction rules for bracket identities, a subsystem of which will yield all the exchange properties valid in matroids. Such a deduction system would break new ground in proof theory, because the exchange axioms are different in kind from the axioms for algebraic structures such as groups.
or fields. Some work on this has been done by Whiteley [1973, 1977, 1979] and Haiman [1984].

Recent work on supersymmetric straightening by Rota and others has led to extensions of the Laplace expansion and new bracket identities. One such novel identity led to the following conjecture due to Rota (see [Huang and Rota 1994]).

3.3. Conjecture. Let $B_1, B_2, \ldots, B_d$ be $d$ bases in a rank-$d$ matroid. Then the elements in these bases can be arranged in a square

$$[a_{ij}]_{1 \leq i, j \leq d}$$

so that

(1) the elements in $B_i$ are in the $i$th row, that is, $B_i = \{a_{i1}, a_{i2}, \ldots, a_{id}\}$, and

(2) the elements $\{a_{1j}, a_{2j}, \ldots, a_{dj}\}$ from every column form a basis.

3.5. The Bowdoin program.

In his Bowdoin lectures, Rota observed that the process of abstracting the Laplace expansion to give the basis exchange axiom can also be applied to invariants of the orthogonal group to yield an "orthogonal matroid". This lay the groundwork for what I have called - with tongue in cheek - the Bowdoin program for combinatorial geometries. I carried out this analogue of Klein's Erlanger Programm in my thesis [1978a] by defining bimatroids, orthogonal matroids, and Pfaffian structures. Parts of [Kung 1978a] appeared in [1978b].

The notion of a bimatroid was also discovered independently and earlier by Schrijver [1978]. His motivation is from matching and optimization theory and he called them "linking systems". A different notion of an "orthogonal matroid" - a "metroid" or metric matroid - can be found in [Dress and Havel 1986]. Gelfand and Serganova in [1987] have implemented another matroidal Erlanger Programm using greedoids.

3.6. Other topics.

Rota also influenced much work by suggesting research areas. We shall give several examples of this.

In 1947, Tutte defined a ring by taking the ring of polynomials with isomorphism classes of graphs as indeterminates modulo the ideal generated by relations, the most important of which are the contraction-and-deletion relations: for a graph $G$ and an edge $e$ in $G$ which is neither a loop nor an isthmus,

$$G - G/e - G\backslash e = 0.$$ 

This relation was suggested by a recursion of R. G. Foster for the chromatic polynomial which was reported in [Whitney 1932b]. Rota realized that Tutte's ring is in fact the Grothendieck ring - the Grothendieck ring was defined later - of the category of graphs with contraction-and-deletion relations performing the same rôle as exact sequences. (A detailed explanation can be found in [Kung 1986].) This observation suggested that a general theory can be developed and that the
contraction-and-deletion relation should be a useful tool in combinatorics. The general theory was developed by Brylawski [1972a and 1972b]. The use of contraction-and-deletion relations in counting and proving theorems is now a standard method (see, for example, [Brylawski and Oxley 1992]). The rather unspecific problem of finding an interesting higher \( k \)-theory for matroids or graphs remains open.

Rota's interest in \( q \)-analogues lead him to ask for a \( q \)-analogue of the partition lattice. This question inspired the study of Dowling geometries, which are group-labelled versions of graphs. See [Dowling 1973a and 1973b] and also [Doubilet, Rota and Stanley 1972]. Dowling geometries form an important class of examples of geometries. Rota also suggested studying "continuous" analogues of partition lattices. Two responses to this are [Björner 1987] and Haiman [1994]. The former has led to the study of pseudomodular lattices and continuous matroids [Björner and Lovász 1987].

Rota was also interested in extending the notion of a matroid. For example, in analogy to schemes in algebraic geometry,\(^2\) he proposed the study of geometries defined on partially ordered sets and more generally, on categories. For work in this area, see, for example, [Faigle 1978, Pezzoli 1981, Barnabei et al. pre].

Another problem posed by Rota is to study functor on categories of matroids. This is the motivation behind [Nguyen 1977]. A possible development from studying functors is a homology theory for matroids using derived functors.

3.7. What is a matroid?

One of the philosophical problems Rota often poses is "What is a matroid?"\(^3\) Part of the reason for asking this question is that, alone among mathematical structures, matroids can and must be axiomatized in many essentially different ways. All these ways are natural in some context and all of them provide the "correct" axioms to prove some theorem. There should be an explanation of this.

One of my answers to this question is to prove an analogue of the classification theorem for finite simple groups. This analogue is based on the intuition that there are only two types of well-behaved "infinite families" of finite geometries: geometries representable over a finite field and group-labelled gain-graphic geometries labelled by a finite group. This conjecture was made precise and proved in [Kahn and Kung 1982]. Just as the classification theorem for finite simple groups says that finite group theory is basically the study of automorphism groups of geometric structures, the analogous theorem for matroids says that matroid theory is the intersection of synthetic projective geometry and (generalized) graph theory.

4. The Critical Problem

To many matroid theorists, Rota's most significant contribution is his formu-
lation of the critical problem [Crapo and Rota 1970, Chapter 16]. Let $S$ be a set of points or nonzero vectors in a $n$-dimensional vector space $V$ over the finite field $GF(q)$. A $k$-tuple $(L_1, L_2, \ldots, L_k)$ of (linear) functionals $V \to GF(q)$ is said to distinguish $S$ if for every point $x \in S$, there exists at least one functional $L_i$ such that $L_i(x) \neq 0$. The critical problem is to find the minimum number $c$ such that there exists a $c$-tuple of functionals distinguishing $S$. Because the kernel $L = \{x : L(x) = 0\}$ is a hyperplane (that is, a projective subspace of codimension 1), the critical problem can also be formulated geometrically:

- Find the minimum number $c$ of projective hyperplanes or codimensional-1 subspaces $H_1, H_2, \ldots, H_c$ such that for every point $p$ in $S$, there exists at least one hyperplane $H_i$ such that $x \notin H_i$.
- Find the minimum number $c$ of affine subsets $A_1, A_2, \ldots, A_c$ in $S$ such that $S$ equals the union $A_1 \cup A_2 \cup \ldots \cup A_c$.

The minimum number $c$ is called the critical exponent of $S$ (over $GF(q)$).

4.1. Theorem (Crapo and Rota). Let $S$ be a set of points in an $n$-dimensional vector space over $GF(q)$. The number of $k$-tuples distinguishing $S$ equals

$$q^{k(n-r)}p(G; q^k) = \sum_{X \in L(G)} \mu(\hat{0}, X)q^{k(n-\text{rank}(X))},$$

where $G$ is the matroid defined by linear dependence on the vectors in $S$, $G$ has rank $r$, $L(G)$ is the lattice of flats of $G$, $\mu$ is the Möbius function in $L(G)$, and $p(G; \lambda) = \sum_{X \in L(G)} \mu(\hat{0}, X)\lambda^{r-\text{rank}(X)}$ is the characteristic polynomial of $L(G)$.

Because the characteristic polynomial of $S$ depends only on the lattice of flats $L(S)$, Theorem 4.1 implies that the critical exponent of $S$ depends only the order $q$ of the field and its matroid structure. This result is the first to indicate the primary importance of the matroid or combinatorial structure of geometric configurations. Shortly after this, Zaslavsky [1975] discovered a formula involving characteristic polynomials for the number of bounded regions in the complement of an arrangement of hyperplanes in $\mathbb{R}^n$ which shows that this number is dependent only on the matroid structure. A currently active area in the theory of arrangements of hyperplanes is to decide whether a given topological property depends only on the matroid structure.

The critical problem originated in an attempt to find a geometric version of the coloring problem. An account of this relation can be found in Rota's Bowdoin lectures (Chapter 8).

Suppose we are given an oriented graph $G$ and a field $F$. Let $V$ and $E$ denote respectively the sets of vertices and edges of $G$. We now consider the vector space $V$ of all 1-coboundaries, i.e., those functions from the set of edges $E$ to the field $F$ which satisfy: if $C$ is any circuit of $G$,

$$\sum_{e \in C} [\text{sgn } e]C f(e) = 0,$$

612
where \([\text{sgn} e]_C\) is positive if the orientation of \(e\) coincides with the direction around \(C\), and is negative otherwise. Given this vector space, we can recover the bond geometry on the set of edges, by setting, for \(A \subseteq E\),
\[
\overline{A} = \{x \in E : f(x) = 0 \text{ for all } f \text{ such that } f|_A \equiv 0\}.
\]

Conversely, of course, we can regard edges of \(G\) as linear functionals on the space \(V\) and thus, we can embed \(E\) into the dual space \(V^*\).

As we know from topology, each 1-coboundary is obtained from a “potential” or 0-chain, i.e., a function from the set of vertices to the field \(F\), by applying the coboundary operator \(\partial\). Thus, we can write \(f = \partial p\), where \(p\) is a 0-chain. We note that for an edge \(e\),
\[
f(e) = p(\text{head of } e) - p(\text{tail of } e).
\]

We now observe that the critical problem for the bond geometry of \(G\) is the following: find the minimum number \(c\) of 1-coboundaries \(f_1, f_2, \ldots, f_c\) such that for each edge \(e \in E\), there exists an \(f_i\), \(1 \leq i \leq c\), such that \(f_i(e) \neq 0\). Since each 1-coboundary comes from a 0-chain, we can rephrase this problem as follows: find the minimum number of 0-chains \(p_1, p_2, \ldots, p_c\) such that for each edge \(e \in E\), there exists a \(p_i\), \(1 \leq i \leq c\), such that
\[
p(\text{head of } e) \neq p(\text{tail of } e).
\]

Suppose our field is \(\text{GF}(q)\). Then with each vertex \(v\), we assign the vector \((p_1(v), p_2(v), \ldots, p_c(v))\). We clearly have a total of \(q^c\) such distinct vectors. But the condition for the critical problem translates precisely to this, that the same vector should not be assigned to any two vertices of \(G\) connected by an edge. Thinking of these vectors as “colors”, we see that the critical problem can be stated thus: find the minimum \(c\) so that \(G\) can be colored with no more than \(q^c\) colors.

The reason for reformulating the coloring problem in geometric terms is to embed it into a family of problems which present the same difficulties for their solution. We believe that the critical problem is the central problem of “extremal” combinatorial theory, and that the study of such problems in the full generality [of matroid theory] will eventually lead to an understanding of the classical coloring problems.

[Crapo and Rota 1970, Chapter 16].

A good example of the power of correct generality is Jaeger’s 8-flow theorem [Jaeger 1979]. An \(n\)-flow \(f\) in a graph \(\Gamma\) is a function from the edge set of \(\Gamma\) to the set of integers modulo \(n\) so that for every vertex \(v\) in \(\Gamma\),
\[
\sum_{e: e \text{ is incident on } v} f(e) \equiv 0 \pmod{n}
\]
A \(n\)-flow is non-zero if \(f(e) \neq 0 \pmod{n}\) for every edge \(e\). Tutte conjectured that every graph has a non-zero 5-flow. From Tutte’s work, it is known that the number
of nowhere-zero \( n \)-flows on a graph \( \Gamma \) equals \( p(M^-(\Gamma); n) \), the characteristic polynomial of the orthogonal dual of the cycle matroid of \( \Gamma \). Thus, to show the existence of an 8-flow on every graph, it suffices to show that every cographic geometry has critical exponent 3 over GF(2). To do this, Jaeger proved that a rank-\( n \) cographic matroid has at most \( 3n - 3 \) points, and hence, by the matroid partition theorem [Edmonds 1965], every cographic matroid can be partitioned into three independent sets. Because independent sets are affine, this proves the theorem. Currently, the best result is the 6-flow theorem of Seymour [1981]. Seymour proved the existence of a 6-flow by combining a 2-flow and a 3-flow. His methods suggest an extension of the critical problem over products of finite fields. Note that Jaeger's method combined with Euler's formula for the number of edges in a triangulation of the sphere yield a matroid-theoretic proof of an 8-color theorem for planar graphs.

The idea behind Jaeger's proof has been developed further. In particular, it shows that the "growth rate" and "uniform critical exponent" of a subgeometry-closed class are related. See [Kung 1993] for a survey of this area of extremal matroid theory.

Another important example [Dowling 1971] of the critical problem is the fundamental problem of coding theory. This is the problem of finding the maximum dimension of a linear code over GF(q) with given length \( n \) and given minimum distance \( k + 1 \). Stated geometrically, this is equivalent to the critical problem for the set \( T_k \) consisting of all the nonzero vectors in \([GF(q)]^n\) with \( k \) or fewer nonzero coordinates. Thus, explicit calculations of characteristic polynomials of the sets \( T_k \) are of particular interest. Only two non-trivial calculations have been done. Dowling [1973a] did the case \( k = 2 \) and Bonin [1994] did the case \( k = n - 2 \).

We conclude this section with two problems posed by Rota. The first is to find an interesting "continuous" version of the critical problem over the real or complex numbers. The essence of the second is in his Bowdoin lectures: Let \( S \) be a set of vectors in a vector space over GF(q). Find the minimum number \( c \) of codimensional-\( k \) subspaces \( K_1, K_2, \ldots, K_c \) such that for every rank-\( k \)-flat \( X \) in \( S \), there exists at least one subspace \( K_i \) such that \( X \cap K_i = S \). The second problem is a higher-dimensional analogue of the critical problem.

5. Simplicial matroids

Another elucidation of the geometry underlying graph theory appears in the paper [SG] on simplicial geometries by Crapo and Rota reprinted in this chapter. Crapo and Rota defined simplicial geometries using Betti numbers. There is also an elementary definition using boundary operators. Let \( T = \{a_1, a_2, \ldots, a_n\} \) be the set of "vertices", \( T_i \) be the collection of \( l \)-element subsets or \( l \)-simplices of \( T \), and \( V_i \) be the vector space over a given field or integral domain \( F \) of all formal linear combinations of \( l \)-simplices. For \( k \geq 1 \), the boundary operator from \( V_k \) to \( V_{k-1} \) is the linear operator defined on the basis elements of \( V_k \) in the following way: when \( k = 1 \),

\[
\partial(\{a_i\}) = 0;
\]
for general \( k \),

\[
\partial(\{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\}) = \sum_{j=1}^{k} (-1)^{k-1} \{a_{i_1}, \ldots, a_{i_{j-1}}, a_{i_{j+1}}, \ldots, a_{i_k}\},
\]

where the indices \( i_1, i_2, \ldots, i_k \) are in increasing order. The (full) simplicial matroid \( G(T_k; F) \) over the field \( F \) is the matroid on \( T_k \) with independent sets specified as follows: the set \( \{ X_1, X_2, \ldots, X_m \} \) of \( k \)-element subsets is independent if and only if their boundaries \( \partial(X_1), \partial(X_2), \ldots, \partial(X_m) \) form a linearly independent set. When \( k \geq 2 \), \( G(T_k; F) \) is a geometry. The idea behind the definition of simplicial geometries is to make all the reduced homology groups trivial with the exception of the highest one.

For any integral domain \( F \), the simplicial geometry \( G(T_2; F) \) is the cycle matroid of the complete graph \( K_n \). Thus, simplicial geometries are \( k \)-dimensional analogues of complete graphs. By taking subgeometry, this gives a way to extend graph theory to higher dimensions. Much of this extension remains to be done. Another aim of [SG] is to show "how the homology structure of any finite simplicial complex may be embodied in a sequence of geometric structures." For example, a theorem in Section 4 shows the connection between orthogonal duality of simplicial geometries and Alexander duality. (Elementary proofs of this theorem are given in [Crapo 1986] and [Cordovil and Lindström 1992].) In view of recent developments which showed the centrality of combinatorial structures in algebraic and differential geometry (see, for example, [Budach 1993, Gelfand et al. 1987, Gelfand and MacPherson 1992]), the objective of [SG], "to free combinatorial topology from its coefficients," seems remarkably prescient.

Most of the work done on simplicial geometries is about their representations. For example, Cordovil [1980] showed that the full simplicial geometry is uniquely representable. A survey can be found in [Cordovil and Lindström 1987].

6. Zeta Functions in Combinatorics

The objective of the paper [Rédei] is to establish connections between combinatorics and older branches of mathematics by (a) importing into combinatorics a central tool of classical mathematics, the zeta function, and (b) exporting from combinatorics the idea of a critical problem. Not much research has been done since the paper. Thus, the thrust of this section will be backwards. We shall describe in more detail Hajós' theorem, a proof of which motivated Rédei to define his zeta function. We shall also discuss one more example of a combinatorial zeta function, Philip Hall's Eulerian function for a group, which was mentioned only in passing in the abstract.

6.1. Hajós' theorem for abelian groups.

Hajós' theorem for abelian groups was motivated by his famous and difficult theorem in the geometry of numbers [1942].
6.1. Hajós' Theorem in the Geometry of Numbers. Let $K$ be a parallelootope in $\mathbb{R}^n$ defined by a system of $n$ linear inequalities

$$|a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n| \leq 1, \quad 1 \leq i \leq n,$$

where $|\text{det}(a_{ij})| = 1$. Then $K$ contains no lattice point (that is, a point with integer coordinates) in its interior if and only if there is an integer matrix $U$ with determinant equal to $\pm 1$ such that under the change of coordinates $x = Uy$ and a permutation of the indices $i$, the system (4) is transformed into a unidiagonal lower triangular system

$$|b_{i1}y_1 + b_{i2}y_2 + \ldots + b_{i,i-1}y_{i-1} + y_i| \leq 1, \quad 1 \leq i \leq n.$$

Theorem 6.1 is a consequence of the following theorem for abelian groups [Hajós 1942].

6.2. Hajós' Theorem for Abelian Groups. Let $G$ be a finite abelian group. Suppose that $X_1, X_2, \ldots, X_r$ are subsets of $G$ of satisfying the following two properties:

1. Each $X_i$ is of the form

$$\{1, a, a^2, a^3, \ldots, a^{r-1}\}$$

for some element $a$ in $G$. (Note that $r$ need not be the order of $a$.)

2. Every element $g \in G$ can be written uniquely as a product

$$g = x_1x_2\cdots x_r,$$

where $x_i \in X_i$.

Then at least one of the subsets $X_i$ forms a (cyclic) subgroup of $G$.

Theorem 6.2 can be viewed as a "converse" of the well-known structure theorem for finite abelian groups due to Frobenius and Stickelberger [1879]. This theorem says that if $G$ is a finite abelian group, then there exist cyclic subgroups $X_1, X_2, \ldots, X_r$ such that every element $g \in G$ can be written uniquely as a product $g = x_1x_2\cdots x_r$, where $x_i \in X_i$.

Rédei gave an equivalent formulation of Theorem 6.2 using his definition of zeta functions for algebraic structures.

6.3. Theorem. The following theorem is equivalent to Hajós' theorem for abelian groups. Let $X_1, X_2, \ldots, X_r$ be cyclic subgroups of a finite abelian group $G$, let $p_1, p_2, \ldots, p_r$ be prime numbers, and let $Y_i = X_i^{p_i} = \{x^{p_i} : x \in X_i\}$ be the subgroup of $p_i$th powers of elements in $X_i$. Suppose that

1. the $p_1p_2\cdots p_r$ possible products of the form $x_1x_2\cdots x_r$, where $x_i \in X_i$, are distinct, and

2. for every subset $M = \{i_1, i_2, \ldots, i_k\}$ of $\{1, 2, \ldots, r\}$,

$$\rho(1; Y_{j_1}x_M/X_M, Y_{j_2}x_M/X_M, \ldots, Y_{j_1}x_M/X_M) = 0,$$

616
where \( \{j_1, j_2, \ldots, j_t\} \) is the complementary subset \( \{1, 2, \ldots, r\} \setminus M \), \( X_M \) is the subgroup \( X_M = X_{i_1} X_{i_2} \cdots X_{i_k} \), the subgroup generated by \( X_{i_1}, X_{i_2}, \ldots, X_{i_k} \), and the Rédei function \( \rho \) is computed in the subgroup lattice of the quotient \( G/X_M \).

Then for all integers \( s \) greater than 1,

\[
\rho(s : \{Y_1, Y_2, \ldots, Y_r\}) = 0,
\]

or, equivalently, one of the subgroups \( Y_i \) is the identity subgroup.

6.2. Eulerian functions.

Let \( G \) be a finite group and let \( \phi(G; s) \) be the number of \( s \)-tuples \((x_1, x_2, \ldots, x_s)\) of elements of \( G \) such that \( x_1, x_2, \ldots, x_s \) form a set of generators of \( G \). The function \( \phi(G; s) \) was defined by P. Hall in [1936] and is called the Eulerian function of \( G \). (Another definition with \( s \)-element subsets instead of \( s \)-tuples is possible, but the present definition yields more interesting theorems.) Since every \( s \)-tuple of elements generates some subgroup,

\[
|H|^s = \sum_{K \leq H} \phi(K; s)
\]

for every subgroup \( H \) of \( G \). Hence, by Möbius inversion,

\[
\phi(G; s) = \sum_{H \in L(G)} \mu(H, G)|H|^s,
\]

where \( \mu \) is the Möbius function in the lattice \( L(G) \) of all subgroups of \( G \). In particular, the probability \( \tilde{\phi}(G; s) \) that an \( s \)-tuple would generate \( G \) equals

\[
\frac{1}{|G|^s} \phi(G; s) = \sum_{H \in L(G)} \mu(H, G) \left( \frac{|G|}{|H|} \right)^s
\]

The probability function \( \tilde{\phi}(G; s) \) is a Rédei function on the order dual of the subgroup lattice with the order function: for \( H \geq K \),

\[
\nu(H, K) = \left( \frac{|H|}{|K|} \right)^s
\]

The critical exponent, that is, the minimum positive integer \( s \) such that \( \tilde{\phi}(G; s) \neq 0 \), is the minimum number of generators for the group \( G \).

[One might note that this argument applies to any finite structure with a binary operation and, indeed, to matroids. For example, let \( M \) be a rank-\( n \) matroid on the set of elements \( S \) and let \( L(M) \) be its lattice of flats. Then the number \( \phi(M; s) \) of ordered \( s \)-tuple \((x_1, x_2, \ldots, x_s)\) such that \( \{x_1, x_2, \ldots, x_s\} \) is a spanning set is given by

\[
\phi(M; s) = \sum_{X \in L(M)} \mu(X, S)|X|^s.
\]

However, there is no order function on \( L(M) \) such that \( \phi(M; s) \) is a Rédei function. The critical problem is trivial in this case, as the critical exponent is clearly the rank \( n \) of \( M \) and \( \phi(M; n)/n! \) is the number of bases in \( M \).]
The problem Hall started with is to find the number of ordered pairs of generators of the alternating group $A_5$ of order 60. Hall laconically stated that if the "subgroups are sufficiently known," then the Eulerian function can be computed and the problem solved. But except in the simplest cases, such knowledge about subgroups is extremely difficult to come by. By a detailed analysis of the subgroup structure, Hall calculated the Eulerian function of $\operatorname{PSL}(2, p)$ for $p$ a prime in [1936]. With much more work, Downs [1991] computed the Eulerian function for $\operatorname{PSL}(2, q)$ for $q$ a prime power.

Delsarte [1948] calculated the Eulerian functions of abelian groups. Using Crapo's complementation theorem [Crapo 1966] and facts about complements in subgroup lattices of soluble groups, Kratzer and Thévenaz [1985] calculated certain values of the Möbius function of a soluble group. An elementary account of recent work on Möbius functions in subgroup lattices can be found in [Hawkes et al. 1989].

We end with an account of a factorization theorem due to Gaschütz [1959]. Let $N$ be a normal subgroup of a finite group $G$. Then
\[
\phi(G; s) = \phi(G/N; s)\phi(G \downarrow N; s),
\]
where
\[
\phi(G \downarrow N; s) = \sum_{H: H \leq G \text{ and } HG = G} \mu(H, G)|N \cap H|^s.
\]
The function $\phi(G \downarrow N; s)$ can be thought of as the Eulerian function of the order ideal of the subgroups $H$ in the dual subgroup lattice such that $NH = G$. (Gaschütz's formula in [1959] was a sum involving maximal subgroups which is equivalent to the one given here.)

Gaschütz gave a counting proof of this identity using the following fact. Let $(Ng_1, Ng_2, \ldots, Ng_s)$ be an $s$-tuple generating the quotient group $G/N$. Then, the number of $s$-tuples $(x_1, x_2, \ldots, x_s)$ in $G$ generating $G$ such that $x_i$ is in the coset $Ng_i$ depends only on $N$ and equals $\phi(G \downarrow N; s)$. Gaschütz's factorization theorem is very similar to Stanley's modular factorization theorem [Stanley 1971] for characteristic polynomials of geometric lattices. Indeed, it is straight-forward to obtain an abstract version for Rédei zeta functions which generalizes both.

6.3. Other extensions of the critical problem.

Zeta functions can also be defined for relations [Kung 1980]. Let $R \subseteq S \times T$ be a relation between the finite sets $S$ and $T$. If $C$ is a subset of $S$ or $T$, its perpendicular $C^\perp$ is the set of all elements in $T$ or $S$ not related to any of the elements in $C$. An $s$-tuple $(u_1, u_2, \ldots, u_s)$ of elements from $T$ is said to distinguish $S$ if for every element $a$ in $S$, there exists an element $u_i$ such that $a$ is not related to $u_i$. The Rédéi function of the relation $R$ is the function defined by the following formula:
\[
\zeta_R(s) = \sum_{B: B \subseteq S} (-1)^{|B|}|B^\perp|^s.
\]
When $s$ is a positive integer, $\zeta_R(s)$ equals the number of $s$-tuples distinguishing $S$. Although this extension is sufficiently general to include all known critical problems,
analagues of the basic theorems hold. For example, one has the following variation on the contraction-and-deletion recursion:

\[ \zeta_{R:S|T}(s) = \zeta_{R:S\setminus\{a\}|T}(s) - \zeta_{R:S\setminus\{a\} \setminus \{a\} \setminus T}(s) \]

where \( R: A\mid B \) is the relation \( R \) restricted to the subsets \( A \subseteq S \) and \( B \subseteq T \). Another extension of the critical problem, to “weighted” lattices, can be found in [Whittle 1994].

7. Conclusion

An indicator of how matroid theory has changed since the 1960’s can be found in the “rewriting” of the two classic textbooks. Crapo and Rota’s “Combinatorial geometries” [1970] has been rewritten in an extended project that took almost twenty years and resulted in four volumes [White 1985, 1986, and 1992; Björner et al. 1992]. Welsh’s “Matroid theory” [1976] was rewritten in a more focused way by Oxley [1994]. Both rewritings resulted in books which blend the graph-theoretic and geometric approach, so that the subject is once again unified. Rota’s work and influence have had major impact in bringing this about and thereby making certain that the geometric way of doing matroid theory will be a permanent voice in the continuing polyphony that is matroid theory.

REFERENCES

Papers reprinted in this chapter.

[SG]
H. H. Crapo and G.-C. Rota, *Simplicial geometries*.

[Rédei]

Rota’s work in matroid theory and related areas


References on matroids and simplicial matroids

R. Cordovil, Representation over a field of full simplicial matroids, European J. Combin. 1 (1980), 201–205.


References on zeta functions


