ON THE FOUNDATIONS OF COMBINATORIAL THEORY I-X

Gian-Carlo Rota *et al.*

Table of Contents

Z. Wahrscheinlichkeitstheorie 2, 340–368 (1964)

On the Foundations of Combinatorial Theory
I. Theory of Möbius Functions

By

GIAN-CARLO ROTA

Contents

1. Introduction

One of the most useful principles of enumeration in discrete probability and combinatorial theory is the celebrated *principle of inclusion-exclusion* (cf. FELLER ⋆, FRÉCHET, RIORDAN, RYSER). When skillfully applied, this principle has yielded the solution to many a combinatorial problem. Its mathematical foundations were thoroughly investigated not long ago in a monograph by FRÉCHET, and it might at first appear that, after such exhaustive work, little else could be said on the subject.

One frequently notices, however, a wide gap between the bare statement of the principle and the skill required in recognizing that it applies to a particular combinatorial problem. It has often taken the combined efforts of many a combinatorial analyst over long periods to recognize an inclusion-exclusion pattern. For example, for the ménage problem it took fifty-five years, since CAYLEY's attempts, before JACQUES TOUCHARD in 1934 could recognize a pattern, and thence readily obtain the solution as an explicit binomial formula. The situation becomes bewildering in problems requiring an enumeration of any of the numerous collections of combinatorial objects which are nowadays coming to the fore. The counting of trees, graphs, partially ordered sets, complexes, finite sets on which groups act, not to mention more difficult problems relating to permutations with restricted position, such as Latin squares and the coloring of maps, seem to lie beyond present-day methods of enumeration. The lack of a systematic

---

This work was begun under contract NSF-GP-149, continued under contract with the Office of Naval Research, and concluded while the author was a Fellow of the Sloan Foundation.

⋆ Author's names refer to the bibliography at the end.

theory is hardly matched by the consummate skill of a few individuals with a natural gift for enumeration.

This work begins the study of a very general principle of enumeration, of which the inclusion-exclusion principle is the simplest, but also the typical case. It often happens that a set of objects to be counted possesses a natural ordering, in general only a partial order. It may be unnatural to fit the enumeration of such a set into a linear order such as the integers: instead, it turns out in a great many cases that a more effective technique is to work with the natural order of the set. One is led in this way to set up a "difference calculus" relative to an arbitrary partially ordered set.

Looked at in this way, a surprising variety of problems of enumeration reveal themselves to be instances of the general problem of inverting an "indefinite sum" ranging over a partially ordered set. The inversion can be carried out by defining an analog of the "difference operator" relative to a partial ordering. Such an operator is the Möbius function, and the analog of the "fundamental theorem of the calculus" thus obtained is the Möbius inversion formula on a partially ordered set. This formula is here expressed in a language close to that of number theory, where it appears as the well-known inverse relation between the Riemann zeta function and the Dirichlet generating function of the classical Möbius function. In fact, the algebra of formal Dirichlet series turns out to be the simplest non-trivial instance of such a "difference calculus", relative to the order relation of divisibility.

Once the importance of the Möbius function in enumeration problems is realized, interest will naturally center upon relating the properties of this function to the structure of the ordering. This is the subject of the first paper of this series; we hope to have at least begun the systematic study of the remarkable properties of this most natural invariant of an order relation.

We begin in Section 3 with a brief study of the incidence algebra of a locally finite partially ordered set and of the invariants associated with it: the zeta function, Möbius function, incidence function, and Euler characteristic. The language of number theory is kept, rather than that of the calculus of finite differences, and the results here are quite simple.

The next section contains the main theorems: Theorem 1 relates the Möbius functions of two sets related by a Galois connection. By suitably varying one of the sets while keeping the other fixed one can derive much information. Theorem 2 of this section is suggested by a technique that apparently goes back to RAMANUJAN. These two basic results are applied in the next section to a variety of special cases; although a number of applications and special cases have been left out, we hope thereby to have given an idea of the techniques involved.

The results of Section 6 stem from an "Ideenkreis" that can be traced back to Whitney's early work on linear graphs. Theorem 3 relates the Möbius function to certain very simple invariants of "cross-cuts" of a finite lattice, and the analogy with the Euler characteristic of combinatorial topology is inevitable. Pursuing this analogy, we were led to set up a series of homology theories, whose Euler characteristic does indeed coincide with the Euler characteristic which we had introduced by purely combinatorial devices.

Some of the work in lattice theory that was carried out in the thirties is useful in this investigation; it turns out, however, that modular lattices are not combinatorially as interesting as a type of structure first studied by WHITNEY, which we have called geometric lattices following BIRKHOFF and the French school. The remarkable property of such lattices is that their Möbius function alternates in sign (Section 7).

To prevent the length of this paper from growing beyond bounds, we have omitted applications of the theory. Some elementary but typical applications will be found in the author's expository paper in the American Mathematical Monthly. Towards the end, however, the temptation to give some typical examples became irresistible, and Sections 9 and 10 were added. These by no means exhaust the range of applications, it is our conviction that the Möbius inversion formula on a partially ordered set is a fundamental principle of enumeration, and we hope to implement this conviction in the successive papers of this series. One of them will deal with structures in which the Möbius function is multiplicative, —-that is, has the analog of the number-theoretic property $\mu(mn) = \mu(m)\,\mu(n)$ if m and n are coprime — and another will give a systematic development of the Ideenkreis centering around POLYA's Hauptsatz, which can be significantly extended by a suitable Möbius inversion.

A few words about the history of the subject. The statement of the Möbius inversion formula does not appear here for the first time: the first coherent version—with some redundant assumptions—is due to WEISNER, and was independently rediscovered shortly afterwards by PHILIP HALL. Ward gave the statement in full generality. Strangely enough, however, these authors did not pursue the combinatorial implications of their work; nor was an attempt made to systematically investigate the properties of Möbius functions. Aside from HALL's applications to p-groups, and from some applications to statistical mechanics by M. S. GREEN and NETTLETON, little has been done; we give a hopefully complete bibliography at the end.

It is a pleasure to acknowledge the encouragement of G. BIRKHOFF and A. GLEASON, who spotted an error in the definition of a cross-cut, as well as of SEYMOUR SHERMAN and KAI-LAI CHUNG. My colleagues D. KAN, G. WHITEHEAD, and especially F. PETERSON gave me essential help in setting up the homological interpretation of the cross-cut theorem.

2. Preliminaries

Little knowledge is required to read this work. The two notions we shall not define are those of a *partially ordered set* (whose order relation is denoted by $\leqq$) and a *lattice*, which is a partially ordered set where max and min of two elements (we call them join and meet, as usual, and write them $\vee$ and $\wedge$) are defined. We shall use instead the symbols $\cup$ and $\cap$ to denote union and intersection *of sets* only. A *segment* $[x, y]$, for x and y in a partially ordered set P, is the set of all elements z between x and y, that is, such that $x \leqq z \leqq y$. We shall occasionally use open or half-open segments such as $[x, y)$, where one of the endpoints is to be omitted. A segment is endowed with the induced order structure; thus, a segment of a lattice is again a lattice. A partially ordered set is *locally finite* if every segment is finite. We shall only deal with locally finite partially ordered sets.

The *product* $P \times Q$ of partially ordered sets P and Q is the set of all ordered pairs (p, q), where $p \in P$ and $q \in Q$, endowed with the order $(p, q) \geqq (r, s)$ whenever $p \geqq r$ and $q \geqq s$. The product of any number of partially ordered sets is defined similarly. The *cardinal power* Hom (P, Q) is the set of all monotonic functions from P to Q, endowed with the partial order structure $f \geqq g$ whenever $f(p) \geqq g(p)$ for every p in P.

In a partially ordered set, an element p *covers* an element q when the segment $[q, p]$ contains two elements. An *atom* in P is an element that covers a minimal element, and a *dual atom* is an element that is covered by a maximal element.

If P is a partially ordered set, we shall denote by P^* the partially ordered set obtained from P by inverting the order relation.

A *closure relation* in a partially ordered set P is a function $p \to \bar{p}$ of P into itself with the properties (1) $\bar{p} \geqq p$; (2) $\bar{\bar{p}} = \bar{p}$; (3) $p \geqq q$ implies $\bar{p} \geqq \bar{q}$. An element is *closed* if $p = \bar{p}$. If P is a finite Boolean algebra of sets, then a closure relation on P defines a lattice structure on the closed elements by the rules $p \wedge q = p \cap q$ and $p \vee q = \overline{p \cup q}$, and it is easy to see that every finite lattice is isomorphic to one that is obtained in this way. A *Galois connection* (cf. ORE, p. 182ff.) between two partially ordered sets P and Q is a pair of functions $\zeta : P \to Q$ and $\pi : Q \to P$ with the properties: (1) both ζ and π are order-inverting; (2) for p in P, $\pi(\zeta(p)) \geqq p$, and for q in Q, $\zeta(\pi(q)) \geqq q$. Under these circumstances the mappings $p \to \pi(\zeta(p))$ and $q \to \zeta(\pi(q))$ are closure relations, and the two partially ordered sets formed by the closed sets are isomorphic.

In Section 7, the notion of a closure relation with the *MacLane-Steinitz exchange property* will be used. Such a closure relation is defined on the Boolean algebra P of subsets of a finite set E and satisfies the following property: if p and q are points of E, and S a subset of E, and if $p \notin \bar{S}$ but $p \in \overline{S \cup q}$, then $q \in \overline{S \cup p}$. Such a closure relation can be made the basis of WHITNEY's theory of independence, as well as of the theory of geometric lattices. The closed sets of a closure relation satisfying the MACLANE-STEINITZ exchange property where every point is a closed set form a geometric (= matroid) lattice in the sense of BIRKHOFF (Lattice Theory, Chapter IX).

A partially ordered set P is said to have a 0 or a I if it has a unique minimal or maximal element. We shall always assume $0 \neq I$. A partially ordered set P having a 0 and a I satisfies the *chain condition* (also called the JORDAN-DEDEKIND chain condition) when all totally ordered subsets of P having a maximal number of elements have the same number of elements. Under these circumstances one introduces the *rank* $r(p)$ of an element p of P as the length of a maximal chain in the segment $[o, p]$, minus one. The rank of 0 is 0, and the rank of an atom is 1. The height of P is the rank of any maximal element, plus one.

Let P be a finite partially ordered set satisfying the chain condition and of height $n+1$. The *characteristic polynomial* of P is the polynomial $\sum_{x \in P} \mu(0, x) \lambda^{n-r(x)}$, where r is the rank function (see the def. of μ below).

If A is a finite set, we shall write $n(A)$ for the number of elements of A.

3. The incidence algebra

Let P be a locally finite partially ordered set. The *incidence algebra* of P is defined as follows. Consider the set of all real-valued functions of two variables $f(x, y)$, defined for x and y ranging over P, and with the property that $f(x, y) = 0$ if $x \nleqq y$. The sum of two such functions f and g, as well as multiplication by scalars, are defined as usual. The product $h = fg$ is defined as follows:

$$h(x, y) = \sum_{x \leqq z \leqq y} f(x, z) g(z, y).$$

In view of the assumption that P is locally finite, the sum on the right is well-defined. It is immediately verified that this is an associative algebra over the real field (any other associative ring could do). The incidence algebra has an identity element which we write $\delta(x, y)$, the Kronecker delta.

The *zeta function* $\zeta(x, y)$ of the partially ordered set P is the element of the incidence algebra of P such that $\zeta(x, y) = 1$ if $x \leqq y$ and $\zeta(x, y) = 0$ otherwise. The function $n(x, y) = \zeta(x, y) - \delta(x, y)$ is called the *incidence function*.

The idea of the incidence algebra is not new. The incidence algebra is a special case of a semigroup algebra relative to a semigroup which is easily associated with the partially ordered set. The idea of taking "interval functions" goes back to DEDEKIND and E. T. BELL; see also WARD.

Proposition 1. *The zeta function of a locally finite partially ordered set is invertible in the incidence algebra.*

Proof. We define the inverse $\mu(x, y)$ of the zeta function by induction over the number of elements in the segment $[x, y]$. First, set $\mu(x, x) = 1$ for all x in P. Suppose now that $\mu(x, z)$ has been defined for all z in the open segment $[x, y)$. Then set

$$\mu(x, y) = -\sum_{x \leqq z < y} \mu(x, z).$$

Clearly μ is an inverse of ζ.

The function μ, inverse to ζ, is called the *Möbius function* of the partially ordered set P.

The following result, simple though it is, is fundamental:

Proposition 2. (Möbius inversion formula). *Let $f(x)$ be a real-valued function, defined for x ranging in a locally finite partially ordered set P. Let an element p exist with the property that $f(x) = 0$ unless $x \geqq p$.*

Suppose that

(*) $$g(x) = \sum_{y \leqq x} f(y).$$

Then

(**) $$f(x) = \sum_{y \leqq x} g(y) \mu(y, x).$$

Proof. The function g is well-defined. Indeed, the sum on the right can be written as $\sum_{p \leqq y \leqq x} f(y)$, which is finite for a locally finite ordered set.

Substituting the right side of (*) into the right side of (**) and simplifying, we get

$$\sum_{y \leqq x} g(y)\,\mu(y,x) = \sum_{y \leqq x}\sum_{z \leqq y} f(z)\,\mu(y,x) = \sum_{y \leqq x}\sum_{z} f(z)\,\zeta(z,y)\,\mu(y,x)\,.$$

Interchanging the order of summation, this becomes

$$\sum_{z} f(z) \sum_{y \leqq x} \zeta(z,y)\,\mu(y,x) = \sum_{z} f(z)\,\delta(z,x) = f(x)\,, \quad \text{q. e. d.}$$

Corollary 1. *Let $r(x)$ be a function defined for x in P. Suppose there is an element q such that $r(x)$ vanishes unless $x \leqq q$. Suppose that*

$$s(x) = \sum_{y \geqq x} r(y)\,.$$

Then

$$r(x) = \sum_{y \geqq x} \mu(x,y)\,s(y)\,.$$

The proof is analogous to the above and is omitted.

Proposition 3. (Duality). *Let P^* be the partially ordered set obtained by inverting the order of a locally finite partially ordered set P, and let μ^* and μ be the Möbius functions of P^* and P. Then $\mu^*(x,y) = \mu(y,x)$.*

Proof. We have, in virtue of Proposition 2 and Corollary 1,

$$\sum_{x \geqq^* y \geqq^* z} \mu^*(x,y) = \delta(x,z)\,.$$

Letting $q(x,y) = \mu^*(y,x)$, it follows that q is an inverse of ζ in the incidence algebra of P. Since the inverse is unique, $q = \mu$, q. e. d.

Proposition 4. *The Möbius function of any segment $[x,y]$ of P equals the restriction to $[x,y]$ of the Möbius function of P.*

The proof is omitted.

Proposition 5. *Let $P \times Q$ be the direct product of locally finite partially ordered sets P and Q. The Möbius function of $P \times Q$ is given by*

$$\mu((x,y),(u,v)) = \mu(x,u)\,\mu(y,v)\,,\ x,u \in P;\ y,v \in Q\,.$$

The proof is immediate and is omitted.

The same letter μ has been used for the Möbius functions of three partially ordered sets, and we shall take this liberty whenever it will not cause confusion.

Corollary (Principle of Inclusion-Exclusion). *Let P be the Boolean algebra of all subsets of a finite set of n elements. Then, for x and y in P,*

$$\mu(x,y) = (-1)^{n(y)-n(x)}\,, \qquad y \geqq x\,,$$

where $n(x)$ denotes the number of elements of the set x.

Indeed, a Boolean algebra is isomorphic to the product of n chains of two elements, and every segment $[x,y]$ in a Boolean algebra is isomorphic to a Boolean algebra.

Aside of the simple result of Proposition 5, little can be said in general about how the Möbius function varies by taking subsets and homomorphic images of a partially ordered set. We shall see that more sophisticated notions will be required to relate the Möbius functions of two partially ordered sets.

Let P be a finite partially ordered set with 0 and I. The *Euler characteristic* E of P is defined as

$$E = 1 + \mu(0,1).$$

The simplest result relating to the computation of the Euler characteristic was proved by PHILIP HALL by combinatorial methods. We reprove it below with a very simple proof which shows one of the uses of the incidence algebra:

Proposition 6. *Let P be a finite partially ordered set with 0 and I. For every k, let C_k be the number of chains with k elements stretched between 0 and I. Then*

$$E = 1 - C_2 + C_3 - C_4 + \cdots.$$

Proof. $\mu = \zeta^{-1} = (\delta + n)^{-1} = \delta - n + n^2 \ldots$. It is easily verified that $n^{k-1}(x, y)$ equals the number of chains of k elements stretched between x and y. Letting $x = 0$ and $y = I$, the result follows at once.

It will be seen in section 6 that the Euler characteristic of a partially ordered set can be related to the classical Euler characteristic in suitable homology theories built on the partially ordered set.

Proposition 6 is a typical application of the incidence algebra. Several other results relating the number of chains and subsets with specified properties can often be expressed in terms of identities for functions in the incidence algebra. In this way, one obtains generalizations to an arbitrary partially ordered set of some classical identities for binomial coefficients. We shall not pursue this line here further, since it lies out of the track of the present work.

Example 1. The classical Möbius function $\mu(n)$ is defined as $(-1)^k$ if n is the product of k distinct primes, and 0 otherwise. The classical inversion formula first derived by Möbius in 1832 is:

$$g(m) = \sum_{n \mid m} f(n); \quad f(m) = \sum_{n \mid m} g(n)\, \mu\left(\frac{m}{n}\right).$$

It is easy to see (and will follow trivally from later results) that $\mu\left(\frac{m}{n}\right)$ is the Möbius function of the set of positive integers, with divisibility as the partial order. In this case the incidence algebra has a distinguished subalgebra, formed by all functions $f(n, m)$ of the form $f(n, m) = G\left(\frac{m}{n}\right)$. The product $H = F\,G$ of two functions in this subalgebra can be written in the simpler form

$$(*) \qquad H(m) = \sum_{kn=m} F(k)\, G(n).$$

If we associate with the element F of this subalgebra the *formal Dirichlet series* $\hat{F}(s) = \sum_{n=1}^{\infty} F(n)/n^s$, then the product (*) corresponds to the product of two formal Dirichlet series considered as functions of s, $\hat{H}(s) = \hat{F}(s)\,\hat{G}(s)$. Under this representation, the zeta function of the partially ordered set is the classical *Riemann zeta function* $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$, and the statement that the Möbius function is the inverse of the zeta function reduces to the classical identity $1/\zeta(s) = \sum_{n=1}^{\infty} \mu(n)/n^s$. It is hoped this example justifies much of the terminology introduced above.

Example 2. If P is the set of ordinary integers, then $\mu(m, n) = -1$ if $m = n - 1$, $\mu(m, m) = 1$, and $\mu(m, n) = 0$ otherwise. The Möbius inversion formula reduces to a well known formula of the calculus of finite differences, which is the discrete analog of the fundamental theorem of calculus.

The Möbius function of a partially ordered set can be viewed as the analog of the classical difference operator $\Delta f(n) = f(n+1) - f(n)$, and the incidence algebra serves as a calculus of finite differences on an arbitrary partially ordered set.

4. Main results

It turns out that the Möbius functions of two partially ordered sets can be compared, when the sets are related by a Galois connection. By keeping one of the sets fixed, and varying the other from among sets with a simpler structure, such as Boolean algebras, subspaces of a finite vector space, partitions, etc., one can derive much information about a Möbius function. This is the program we shall develop. The basic result is the following:

Theorem 1. *Let P and Q be finite partially ordered sets, where P has a 0 and Q has a 0 and a 1. Let μ_P and μ be their Möbius functions. Let*

$$\pi : Q \to P; \quad \varrho : P \to Q$$

be a Galois connection such that

(1) $$\pi(x) = 0 \quad \text{if and only if} \quad x = 1\,.$$

(2) $$\varrho(0) = 1\,.$$

Then

$$\mu(0, 1) = \sum_{a>0} \mu_P(0, a)\,\zeta(\varrho(a), 0) = \sum_{[a:\varrho(a)=0]} \mu_P(0, a)\,.$$

One gets a significant summand on the right for every $a > 0$ in P which is mapped into 0 by ϱ. One therefore expects the right side to contain "few" terms. In general, μ_P is a known function and μ is the function to be determined.

Proof. We shall first establish the identity

(*) $$\sum_{a \geqq b} \delta(\pi(x), a) = \zeta(x, \varrho(b))$$

for every b in P. Here ζ on the right stands for the zeta function of Q. Equation (*) is equivalent to the following statement: $\pi(x) \geqq b$ if and only if $x \leqq \varrho(b)$. But this latter statement is immediate from the properties of a Galois connection. Indeed, if $\pi(x) \geqq b$, then $\varrho(\pi(x)) \leqq \varrho(b)$, but $x \leqq \varrho(\pi(x))$, hence $x \leqq \varrho(b)$, and similarly for the converse implication.

To identity (*) we apply the Möbius inversion formula relative to P, thereby obtaining the identity

(**) $$\delta(\pi(x), 0) = \sum_{a \geqq 0} \mu_P(0, a)\,\zeta(x, \varrho(a))\,.$$

Now, $\delta(\pi(x), 0)$ takes the value 1 if and only if $\pi(x) = 0$, that is, in view of assumption (1), if and only if $x = 1$. For all other values of x, we have $\delta(\pi(x), 0) = 0$. Therefore,

$$\delta(\pi(x), 0) = 1 - n(x, 1).$$

We can now rewrite equation (**) in the form

$$1 - n(x, 1) = \zeta(x, \varrho(0)) + \sum_{a>0} \mu_P(0, a) \zeta(x, \varrho(a))$$

However, in view of assumption (2), $\zeta(x, \varrho(0)) = \zeta(x, 1)$, and this is identically one for all x in Q. Therefore, simplifying,

$$-n(x, 1) = \sum_{a>0} \mu_P(0, a) \zeta(x, \varrho(a)).$$

Now, since $\zeta = \delta + n$, we have $\mu = \delta - \mu n$, hence, recalling that $0 \neq 1$,

$$\mu(0, 1) = -\sum_{0 \leqq x \leqq 1} \mu(0, x) n(x, 1) = \sum_{0 \leqq x \leqq 1} \sum_{a>0} \mu_P(0, a) \mu(0, x) \zeta(x, \varrho(a)).$$

Interchanging the order of summation, we get

$$\mu(0, 1) = \sum_{a>0} \mu_P(0, a) \sum_{0 \leqq x \leqq 1} \mu(0, x) \zeta(x, \varrho(a)).$$

The last sum on the right equals $\delta(0, \varrho(a))$, and this equals $\zeta(\varrho(a), 0)$. The proof is therefore complete.

For simplicity of application, we restate Theorem 1 inverting the order of P.

Corollary. *Let $p: Q \to P$; $q: P \to Q$ be order preserving functions between P and Q such that*

(1) If $p(x) = 1$ then $x = 1$, and conversely.

(2) $$q(1) = 1.$$

(3) $$p(q(x)) \leqq x \quad \text{and} \quad q(p(x)) \geqq x.$$

Then

$$\mu(0, 1) = \sum_{a<1} \mu_P(a, 1) \zeta(q(a), 0) = \sum_{[a:q(a)=0]} \mu_P(a, 1)$$

where μ is the Möbius function of Q.

The second result is suggested by a technique which apparently goes back to RAMANUJAN (cf. HARDY, RAMANUJAN, page 139).

Theorem 2. *Let Q be a finite partially ordered set with 0, and let P be a partially ordered set with 0. Let $p: Q \to P$ be a monotonic function of Q onto P. Assume that the inverse image of every interval $[0, a]$ in P is an interval $[0, x]$ in Q, and that the inverse image of 0 contains at least two points.*

Then

$$\sum_{[x:p(x)=a]} \mu(0, x) = 0$$

for every a in P.

The *proof* is by induction over the set P. Since $[0, 0]$ is an interval and its inverse image is an interval $[0, q]$ with $q > 0$, we have

$$\sum_{[x:p(x)=0]} \mu(0, x) = \sum_{0 \leqq x \leqq q} \mu(0, x) = 0.$$

Suppose now the statement is true for all b such that $b < a$ in P. Then

$$\sum_{b<a} \sum_{[x:p(x)=b]} \mu(0, x) = 0 .$$

It follows that

$$\sum_{[x:p(x)=a]} \mu(0, x) = \sum_{b \leqq a} \sum_{[x:p(x)=b]} \mu(0, x) .$$

The last sum equals the sum over some interval $[0, r]$ which is the inverse image of the segment $[0, a]$, that is

$$\sum_{b \leqq a} \sum_{[x:p(x)=b]} \mu(0, x) = \sum_{0 \leqq x \leqq r} \mu(0, x) = \delta(0, r) .$$

But $r > 0$ because a is strictly greater than 0. Hence $\delta(r, 0) = 0$, and this concludes the proof.

5. Applications

The simplest (and typical) application of Theorem 1 is the following:

Proposition 1. *Let R be a subset of a finite lattice L with the following properties: $1 \notin R$, and for every x of L, except $x = 1$, there is an element y of R such that $y \geqq x$.*

For $k \geqq 2$, let q_k be the number of subsets of R containing k elements whose meet is 0. *Then $\mu(0, 1) = q_2 - q_3 + q_4 + \cdots$.*

Proof. Let $B(R)$ be the Boolean algebra of subsets of R. We take $P = B(R)$ and $Q = L$ in Theorem 1, and establish a Galois connection as follows. For x in L, let $\pi(x)$ be the set of elements of R which dominate x. In particular, $\pi(1)$ is the empty set. For A in $B(R)$, set $\varrho(A) = \wedge A$, namely, the meet of all elements of A, an empty meet giving as usual the element 1. This is evidently a Galois connection. Conditions (1) and (2) of the Theorem are obviously satisfied.

The function μ_P is given by the Corollary of Proposition 5 of Section 3, and hence the conclusion is immediate.

Two noteworthy special cases are obtained by taking R to be the set of dual atoms of Q, or the set of all elements < 1 (cf. also WEISNER).

Closure relations. A useful application of Theorem 1 is the following:

Proposition 2. *Let $x \to \bar{x}$ be a closure relation on a partially ordered set Q having* 1, *with the property that $\bar{x} = 1$ only if $x = 1$. Let P be the partially ordered subset of all closed elements of Q. Then:* (a) *If $\bar{x} > x$, then $\mu(x, 1) = 0$;* (b) *If $\bar{x} = x$, then $\mu(x, 1) = \mu_P(x, 1)$, where μ_P is the Möbius function of P.*

Proof. Considering $[x, 1]$, it may be assumed that P has a 0 and $x = 0$. We apply Corollary 1 of Theorem 1, setting $p(x) = \bar{x}$ and letting q be the injection map of P into Q. It is then clear that the assumptions of the Corollary are satisfied, and the set of all a in P such that $q(a) = 0$ is either the empty set or the single element 0, q. e. d.

Corollary (Ph. Hall). *If* 0 *is not the meet of dual atoms of a finite lattice L, or if* 1 *is not the join of atoms, then $\mu(0, 1) = 0$.*

Proof. Set $\bar{x} = \wedge A(x)$, where $A(x)$ is the set of dual atoms of Q dominating x, and apply the preceding result. The second assertion is obtained by inverting the order.

Example 1. *Distributive lattices.* Let L be a locally finite distributive lattice. Using Proposition 2, we can easily compute its Möbius function. Taking an interval

$[x, y]$ and applying Proposition 4 of Section 3, we can assume that L is finite. For $a \in L$, define $\bar{a}$ to be the join of all atoms which a dominates. Then $a \to \bar{a}$ is a closure relation in the inverted lattice L^*. Furthermore, the subset of closed elements is easily seen to be isomorphic to a finite Boolean algebra (cf. BIRKHOFF Lattice Theory, Ch. IX) Applying Proposition 5 of Section 3, we find: $\mu(x, y) = 0$ if y is not the join of elements covering x, and $\mu(x, y) = (-1)^n$ if y is the join of n distinct elements covering x.

In the special case of the integers ordered by divisibility, we find the formula for the classical Möbius function (cf. Example 1 of Section 3.).

The Möbius function of cardinal products. Let P and Q be finite partially ordered sets. We shall determine the Möbius function of the partially ordered set Hom(P, Q) of monotonic functions from P to Q, in terms of the Möbius function of Q. It turns out that very little information is needed about P.

A few preliminaries are required for the statement.

Let R be a subset of a partially ordered set Q with 0, and let $\bar{R}$ be the ideal generated by R, that is, the set of all elements x in Q which are below ($<$) some element of R. We denote by Q/R the partially ordered set obtained by removing off all the elements of $\bar{R}$, and leaving the rest of the order relation unchanged. There is a natural order-preserving transformation of Q onto Q/R which is one-to-one for elements of Q not in $\bar{R}$. We shall call Q/R the *quotient* of Q by the ideal generated by R.

Lemma. *Let $f: P \to Q$ be monotonic with range $R \subset Q$. Then the segment $[f, 1]$ in* Hom (P, Q) *is isomorphic with* Hom $(P, Q/R)$.

Proof. For g in $[f, 1]$, set $g'(x) = g(x)$ to obtain a mapping $g \to g'$ of $[f, 1]$ to Hom $(P, Q/R)$. Since $g \geqq f$, the range of g lies above R, so the map is an isomorphism.

Proposition 3. *The Möbius function μ of the cardinal product* Hom (P, Q) *of the finite partially ordered set P with the partially ordered set Q with* 0 *and* 1 *is determined as follows:*

(a) *If $f(p) \neq 0$ for some element p of P which is not maximal, then $\mu(0, f) = 0$.*

(b) *In all other cases,*

$$\mu(0, f) = \prod_m \mu(0, f(m)), \qquad f \in P,$$

where the product ranges over all maximal elements of P, and where μ on the right stands for the Möbius function of Q.

(c) *For $f \leqq g$, $\mu(f, g) = \mu(0, g')$, where g' is the image of g under the canomial map of $[f, 1]$ onto Hom $(P, Q/R)$, provided Q/R has a 0.*

Proof. Define a closure relation in $[0, f]^*$, namely the segment $[0, f]$ with the inverted order relation, as follows. Set $\bar{g}(m) = g(m)$ if m is a maximal element of P, and $g(a) = 0$ if a is not a maximal element of P. If $\bar{g} = 0$, then $g(m) = 0$ for all maximal elements m, hence $g(a) = 0$ for all $a <$ some maximal element, since g is monotonic. Hence $g = 0$, and the assumption of Proposition 2 is satisfied. The set of closed elements is isomorphic to Hom (M, P), where M is a set of as many elements as there are maximal elements in P. Conclusion (a) now follows from Proposition 2, and conclusion (b) from Proposition 5 of Section 3. Conclusion (c) follows at once from the Lemma.

We pass now to some applications of Theorem 2.

Proposition 4. *Let $a \to \bar{a}$ be a closure relation on a finite lattice Q, with the property that $\overline{a \vee b} = \bar{a} \vee \bar{b}$ and $\bar{0} > 0$. Then for all $a \in Q$,*

$$\sum_{[x:\bar{x}=a]} \mu(0, x) = 0 .$$

Proof. Let P be a partially ordered set isomorphic to the set of closed elements of L. We define $p(x)$, for x in Q, to be the element of P corresponding to the closed element $\bar{x}$. Since $\bar{0} > 0$, any x between 0 and $\bar{0}$ is mapped into $\bar{0}$. Hence the inverse image of 0 in P under the homomorphism p is the nontrival interval $[0, \bar{0}]$.

Now consider an interval $[0, a]$ in P. Then $p^{-1}([0, a]) = [0, \bar{x}]$, where $\bar{x}$ is the closed element of L corresponding to a. Indeed, if $0 \leqq y \leqq \bar{x}$ then $\bar{y} \leqq \bar{\bar{x}} = \bar{x}$, hence $p(y) \leqq a$. Conversely, if $p(y) \leqq a$, then $\bar{y} \leqq \bar{x}$ but $y \leqq \bar{y}$, hence $y \leqq \bar{x}$. Therefore the condition of Theorem 2 is satisfied, and the conclusion follows at once.

Corollary (Weisner).

(a) *Let $a > 0$ in a finite lattice L. Then, for any b in L,*

$$\sum_{x \vee a = b} \mu(0, x) = 0$$

(b) *Let $a < 1$ in L. Then, for any b in L,*

$$\sum_{x \wedge a = b} \mu(x, 1) = 0 .$$

Proof. Take $\bar{x} = x \vee a$. Part (b) is obtained by inverting the order.

Example 2. Let V be a finite-dimensional vector space of dimension n over a finite field with q elements. We denote by $L(V)$ the lattice of subspaces of V. We shall use Proposition 4 to compute the Möbius function of $L(V)$.

In the lattice $L(V)$, every segment $[x, y]$, for $x \leqq y$, is isomorphic to the lattice $L(W)$, where W is the quotient space of the subspace y by the subspace x. If we denote by $\mu_n = \mu_n(q)$ the value of $\mu(0, 1)$ for $L(V)$, it follows that $\mu(x, y) = \mu_j$, when j is the dimension of the quotient space W. Therefore once μ_n is known for for every n, the entire Möbius function is known.

To determine μ_n, consider a subspace a of dimension $n - 1$. In view of the preceding Corollary, we have for all $a < 1$ (where 1 stands for the entire space V):

$$\sum_{x \wedge a = 0} \mu(x, 1) = 0$$

where 0 stands of course for the 0-subspace. Let a be a dual atom of $L(V)$, that is, a subspace of dimension $n - 1$. Which subspaces x have the property that $x \wedge a = 0$? x must be a line in V, and such a line must be disjoint except for 0 from a. A subspace of dimension $n - 1$ contains q^{n-1} distinct points, so there will be $q^n - q^{n-1}$ points outside of a. However, every line contains exactly $q - 1$ points. Therefore, for each subspace a of dimension $n - 1$ there are

$$\frac{q^n - q^{n-1}}{q - 1} = q^{n-1}$$

distinct lines x such that $x \wedge a = 0$. Since each interval $[x, 1]$ is isomorphic to a space of dimension $n-1$, we obtain

$$\mu_n = \mu(0, 1) = -\sum_{\substack{x \wedge a = 0 \\ x \neq 0}} \mu(x, 1) = -q^{n-1}\mu_{n-1}.$$

This is a difference equation for μ_n which is easily solved by iteration. We obtain the result, first established by PHILIP HALL (see also WEISNER and S. DELSARTE):

$$\mu_n(q) = (-1)^n q^{n(n-1)/2} = (-1)^n q^{\binom{n}{2}}.$$

6. The Euler characteristic

Sharper results relating $\mu(0, 1)$ to combinatorial invariants of a finite lattice can be obtained by application of Theorem 1, when the "comparison set" P remains a Boolean algebra.

A *cross-cut* C of a finite lattice L is a subset of L with the following properties:

(a) C does not contain 0 or 1.

(b) no two elements of C are comparable (that is, if x and y belong to C, then neither $x < y$ nor $x > y$ holds).

(c) Any maximal chain stretched between 0 and 1 meets the set C.

A *spanning subset* S of L is a subset such that $\vee S = 1$ and $\wedge S = 0$.

The main result is the following *Cross-cut Theorem*:

Theorem 3. *Let μ be the Möbius function and E the Euler characteristic of a nontrivial finite lattice L, and let C be a cross-cut of L. For every integer $k \geqq 2$, let q_k denote the number of spanning subsets of C containing k distinct elements. Then*

$$E - 1 = \mu(0, 1) = q_2 - q_3 + q_4 - q_5 + \cdots$$

The *proof* is by induction over the distance of a cross-cut C from the element 1.

Define the distance $d(x)$ of an element x from the element 1 as the maximum length of a chain stretched between x and 1. For example, the distance of a dual atom is two. If C is a cross-cut of L, define the distance $d(C)$ as $\max d(x)$ as x ranges over C. Thus, the distance of the cross-cut consisting of all dual atoms is two, and conversely, this is the only cross-cut having distance two.

It follows from Proposition 1 of Section 5 that the result holds when $d(C) = 2$ (take $R = C$ in the assertion of the Proposition). Thus, we shall assume the truth of the statement for all cross-cuts whose distance is less than n, and prove it for a cross-cut with $d(C) = n$.

If C is a subset of L, we shall write $x > C$ or $x \leqq C$ to mean that there is an element y or C such that $x > y$, or that there is an element y of C such that $x \leqq y$. For a general C, these possibilities may not be mutually exclusive; they are mutually exclusive when C is a cross-cut. We shall repeatedly make use of this remark below.

Define a modified lattice L' as follows. Let L' contain all the elements x such that $x \leqq C$ in the same order. On top of C, add an element 1 covering all the elements of C, but no others; this defines L'.

In L', consider the cross-cut C and apply Proposition 1 of section 5 again. If μ' is the Möbius function of L', then

$$\mu'(0, 1) = p_2 - p_3 + p_4 \ldots,$$

where p_k is the number of all subsets $A \subset C \subset L'$ of k elements, such that $\wedge A = 0$.

Comparing the lattices L and L', we have

$$0=\sum_{x\leqq C}\mu(0,x)+\sum_{x>C}\mu(0,x)=\sum_{x\leqq C}\mu'(0,x)+\mu'(0,1).$$

However, for $x\leqq C$, we have $\mu'(0,x)=\mu(0,x)$ by construction of L'. Hence

$$\sum_{x\leqq C}\mu(0,x)=-p_2+p_3-p_4+\cdots$$

Since the sets $(x/x\leqq C)$ and $(x/x>C)$ are disjoint, we can write

$$\mu(0,1)=-\sum_{x<1}\mu(0,x)=-[\sum_{x\leqq C}\mu(0,x)+\sum_{1>x>C}\mu(0,x)].$$

We now simplify the first summation on the right:

(*) $$\mu(0,1)=p_2-p_3+p_4\cdots-\sum_{1>x>C}\mu(0,x).$$

Now let $q_k(x)$ be the number of subsets of C having k elements, whose meet is 0 and whose join is x. In particular, $q_k(1)=q_k$. Then clearly

$$p_k=\sum_{x>C}q_k(x),\quad k\geqq 2,$$

the summation in (*) can be simplified to

(**) $$\mu(0,1)=(q_2-q_3+q_4-\cdots)-\sum_{1>x>C}[-q_2(x)+q_3(x)-q_4(x)+ +\cdots+\mu(0,x)].$$

For x above C and unequal to 1, consider the segment $[0,x]$. We prove that $C(x)=C\cap[0,x]$ is a cross-cut of the lattice $[0,x]$ such that $d(C(x))<d(C)$. Once this is done, it follows by the induction hypothesis that every term in brackets on the right of (**) vanishes, and the proof will be complete.

Conditions (a) and (b) in the definition of a cross-cut are trivially satisfied by $C(x)$, and condition (c) is verified as follows. Suppose Q is a maximal chain in $[0,x]$ which does not meet $C(x)$. Choose a maximal chain R in the segment $[x,1]$; then the chain $Q\cup R$ is maximal in L, and does not intersect C.

It remains to verify that $d(C(x))<d(C)$, and this is quite simple. There is a chain Q stretched between C and x whose length is $d(C(x))$. Then $d(C)$ exceeds the length of the chain $Q\cup R$, and since $x<1$, R has length at least 2, hence the length of $Q\cup R$ exceeds that of Q by at least one. The proof is therefore complete.

Theorem 3 gives a relation between the value $\mu(0,1)$ and the width of narrow cross-cuts or *bottlenecks* of a lattice. The proof of the following statement is immediate.

Corollary 1. (a) *If L has a cross-cut with one element, then* $\mu(0,1)=0$.

(b) *If L has a cross-cut with two elements, then the only two possible values of* $\mu(0,1)$ *are* 0 *and* 1.

(c) *If L has a cross-cut having three elements, then the only possible values of* $\mu(0,1)$ *are* 2, 1, 0 *and* -1.

In this connection, an interesting combinatorial problem is to determine all possible values of $\mu(0,1)$, given that L has a cross-cut with n elements.

Reduction of the main formula. In several applications of the cross-cut theorem, the computation of the number q_k of spanning sets may be long, and systematic procedures have to be devised. One such procedure is the following:

Proposition 1. *Let C be a cross-cut of a finite lattice L. For every integer $k \geqq 0$, and for every subset $A \subset C$, let $q(A)$ be the number of spanning sets containing A, and let $S_k = \sum_A q(A)$, where A ranges over all subsets of C having k elements. Set S_0 to be the number of elements of C. Then*

$$\mu(0,1) = S_0 - 2S_1 + 2^2 S_2 - 2^3 S_3 + \cdots.$$

Proof. For every subset $B \subset C$, set $p(B) = 1$ if B is a spanning set, and $p(B) = 0$ otherwise. Then

$$q(A) = \sum_{C \supseteq B \supseteq A} p(B).$$

Applying the Möbius inversion formula on the Boolean algebra of subsets of C, we get

$$p(A) = \sum_{B \supseteq A} q(B)\mu(A,B),$$

where μ is the Möbius function of the Boolean algebra. Summing over all subsets $A \subset C$ having exactly k elements,

$$q_k = \sum_{n(A)=k} p(A) = \sum_{n(A)=k} \quad \sum_{B \supseteq A} q(B)\mu(A,B).$$

Interchanging the order of summation on the right, recalling Proposition 5 of Section 3 and the fact that a set of $k+l$ elements possesses $\binom{k+l}{l}$ subsets of k elements, we obtain

$$q_k = S_k - \binom{k+1}{1} S_{k+1} + \binom{k+2}{2} S_{k+2} \cdots + (-1)^{n-k} \binom{n}{k} S_n.$$

A convenient way of recasting this expression in a form suitable for computation is the following. Let V be the vector space of all polynomials in the variable x, over the real field. The polynomials $1, x, x^2, \ldots,$ are linearly independent in V. Hence there exists a linear functional L in V such that

$$L(x^k) = S_k, \quad k = 0, 1, 2, \ldots.$$

Formula (*) can now be rewritten in the concise form

$$q_k = L\left(x^k - (k+1)x^{k+1} + \binom{k+2}{2} x^{k+2} - \cdots\right) = L\left(\frac{x^k}{(1+x)^{k+1}}\right).$$

Upon applying the cross-cut theorem, we find the expression (where q_0 and q_1 are also given by (*), but turn out to be 0)

$$\begin{aligned}\mu(0,1) &= L\left(\frac{1}{1+x} - \frac{x}{(1+x)^2} + \frac{x^2}{(1+x)^3} - \cdots\right) \\ &= L\left(\frac{1}{1+2x}\right) = L(1 - 2x + 4x^2 - 8x^3 + \cdots) \\ &= S_0 - 2S_1 + 4S_2 - \cdots, \quad \text{q.e.d.}\end{aligned}$$

The cross-cut theorem can be applied to study which alterations of the order relation of a lattice preserve the Euler characteristic. Every alteration which preserves meets and joins of the spanning subsets of some cross-cut will preserve the Euler characteristic. There is a great variety of such changes, and we shall not develop a systematic theory here. The following is a simple case.

Following BIRKHOFF and JÓNSSON and TARSKI we define the *ordinal sum* of lattices as follows. Given a lattice L and a function assigning to every element x of L a lattice $L(x)$, (all the $L(x)$ are distinct) the *ordinal sum* $P = \sum_L L(x)$ of the lattices $L(x)$ *over* the lattice L is the partially ordered set P consisting of the set $\bigcup_{x \in L} L(x)$, where $u \leqq v$ if $u \in L(x)$ and $v \in L(x)$ and $u \leqq v$ in $L(x)$, or if $u \in L(x)$ and $v \in L(y)$ and $x < y$. It is clear that P is a lattice if all the $L(x)$ are finite lattices.

Proposition 2. *If the finite lattice P is the ordinal sum of the lattices $L(x)$ over the non-trivial lattice L, and μ_p, μ_x and μ_L are the corresponding Möbius functions, then:*

If $L(0)$ is the one element lattice, then $\mu_p(0, 1) = \mu_L(0, 1)$.

Proof. The atoms of P are in one-to-one correspondence with the atoms of L and the spanning subsets are the same. Hence the result follows by applying the cross-cut theorem to the atoms.

In virtue of a theorem of JÓNSSON and TARSKI, every lattice P has a unique maximal decomposition into an ordinal sum over a "skeleton" L. This can be used in connection with the preceding Corollary to further simplify the computation of $\mu(0, n)$ as n ranges through P.

Homological interpretation. The alternating sums in the Cross-Cut Theorem suggest that the Euler characteristic of a lattice be interpreted as the Euler characteristic in a suitable homology theory. This is indeed the case. We now define* a *homology theory* $H(C)$ relative to an arbitrary cross-cut C of a finite lattice L. For the homological notions, we refer to Eilenberg-Steenrod.

Order the elements of C, say $a_1, a_2, \ldots, a_n$. For $k \geqq 0$, let a *k-simplex* σ be any subset of C of $k + 1$ elements which does *not* span. Let C_k be the free abelian group generated by the k-simplices. We let $C_{-1} = 0$; for a given simplex σ, let σ_i be the set obtained by omitting the $(i + 1)$-st element of σ, when the elements of σ are ordered according to the given ordering of C. The boundary of a k-simplex is defined as usual as $\partial_k \sigma = \sum_{i=0}^{k} (-1)^i \sigma_i$, and is extended by linearity to all of C_k, giving a linear mapping of C_k into C_{k-1}. The k-th homology group H_k is defined as the abelian group obtained by taking the quotient of the kernel of ∂_k by the image of ∂_{k+1}. The rank b_k of the abelian group H_k, that is, the number of independent generators of infinite cyclic subgroups of H_k, is the k-th *Betti number*.

Let α_k be the rank of C_k, that is, the number of k-simplices. The *Euler characteristic* of the homology $H(C)$ is defined in homology theory as

$$E(C) = \sum_{k=0}^{\infty} (-1)^k \alpha_k .$$

* This definition was obtained jointly with D. KAN, F. PETERSON and G. WHITEHEAD, whom I now wish to thank.

It follows from well-known results in homology theory that

$$E(C) = \sum_{k=0}^{\infty} (-1)^k b_k .$$

Let q_k be the number of spanning subsets with k elements as in Theorem 3. Then $q_{k+1} + \alpha_k$ is the total number of subsets of C having $k+1$ elements; if C has N elements, then $\alpha_k = \binom{N}{k+1} - q_{k+1}$. It follows from the Cross-Cut Theorem that

$$\begin{aligned} E(C) &= \sum_{k=0}^{\infty} (-1)^k \binom{N}{k+1} - \sum_{k=0}^{\infty} (-1)^k q_{k+1} \\ &= \sum_{k=0}^{\infty} (-1)^k \binom{N}{k+1} + \mu(0,1) . \end{aligned}$$

We have however

$$\sum_{k=0}^{\infty} (-1)^k \binom{N}{k+1} = -\sum_{i=1}^{\infty} (-1)^i \binom{N}{i} = 1 - \sum_{i=0}^{\infty} (-1)^i \binom{N}{i} = 1 - (1-1)^N = 1 ,$$

and hence

$$E(C) = 1 + \mu(0,1) = E;$$

in other words:

Proposition 3. *In a finite lattice, the Euler characteristic of the homology of any cross-cut C equals the Euler characteristic of the lattice.*

This result can sometimes be used to compute the Möbius functions of "large" lattices. In general, the numbers q_k are rather redundant, since any spanning subset of k elements gives rise to several spanning subsets with more than k elements. A method for eliminating redundant spanning sets is then called for. One such method consists precisely in the determination of the Betti numbers b_k.

We conjecture that the Betti numbers of $H(C)$ are themselves independent of the cross-cut C, and are also "invariants" of the lattice L, like the Euler characteristic $E(C)$. In the special case of lattices of height 4 satisfying the chain condition, this conjecture has been proved (in a different language) by DOWKER.

Example 1. *The Betti numbers of a Boolean algebra.* We take the cross-cut C of all atoms. If the height of the Boolean algebra is $n+1$, then every k-cycle, for $k < n-2$, bounds, so that $b_0 = 1$ and $b_k = 0$ for $0 < k < n-2$. On the other hand, there is only one cycle in dimension $n-2$. Hence $b_{n-2} = 1$ and we find $E = 1 + (-1)^{n-2}$, which agrees with Proposition 5 of Section 3.

A notion of Euler characteristic for *distributive* lattices has been recently introduced by HADWIGER and KLEE. For finite distributive lattices, KLEE's Euler characteristic is related to the one introduced in this work. We refer to KLEE's paper for details.

7. Geometric lattices

An ordered structure of very frequent occurrence in combinatorial theory is the one that has been variously called matroid (WHITNEY), matroid lattice (BIRKHOFF), closure relation with the exchange property (MACLANE), geometric lattice

(Birkhoff), abstract linear dependence relation (Bleicher and Preston). Roughly speaking, these structures arise in the study of combinatorial objects that are obtained by piecing together smaller objects with a particularly simple structure. The typical such case is a linear graph, which is obtained by piecing together edges. Several counting problems associated with such structures can often be attacked by Möbius inversion, and one finds that the Möbius functions involved have particularly simple properties.

We briefly summarize the needed facts out of the theory of such structures, referring to any of the works of the above authors for the proofs.

A finite lattice L is a *geometric lattice* when every element of L is the join of atoms, and whenever if a and b in L cover $a \wedge b$, then $a \vee b$ covers both a and b. Equivalently, a geometric lattice is characterized by the existence of a rank function satisfying $r(a \wedge b) + r(a \vee b) \leqq r(a) + r(b)$. Notice that this implies the chain condition. In particular if a is an atom, then $r(a \vee c) = r(c)$ or $r(c) + 1$. If M is a semimodular lattice, then the partially ordered subset of all elements which are joins of atoms is a geometric sublattice.

Geometric lattices are most often obtained from a closure relation on a finite set which satisfies the MacLane-Steinitz exchange property. The lattice L of closed sets in such a closure relation is a geometric lattice whenever every one-element set is closed. Conversely, every geometric lattice can be obtained in this way by defining one such closure relation on the set of its atoms.

The fundamental property of the Möbius function of geometric lattices is the following:

Theorem 4. *Let μ be the Möbius function of a finite geometric lattice L. Then:*

(a) *$\mu(x, y) \neq 0$ for any pair x, y in L, provided $x \leqq y$.*

(b) *If y covers z, then $\mu(x, y)$ and $\mu(x, z)$ have opposite signs.*

Proof. Any segment $[x, y]$ of a geometric lattice is also a geometric lattice. It will therefore suffice to assume that $x = 0$, $y = 1$ and that z is a dual atom of L.

We proceed by induction. The theorem is certainly true for lattices of height 2, where $\mu(0, 1) = -1$. Assume it is true for all lattices of height $n - 1$, and let L be a lattice of height n. By the Corollary to Proposition 4 of Section 5, with $b = 1$, and a an atom of L, we have

$$\mu(0, 1) = - \sum_{\substack{x \vee a = 1 \\ x \neq 1}} \mu(0, x) .$$

Now from the subadditive inequality

$$r(x \wedge a) + r(x \vee a) \leqq r(x) + r(a)$$

we infer that if $x \vee a = 1$, then $n \leqq \dim x + \dim a$, hence $\dim x \geqq n - 1$. The element x must therefore be a dual atom. It follows from the induction assumption and from the fact that L satisfies the chain condition, that all the $\mu(0, x)$ in the sum on the right have the same sign, and none of them is zero. Therefore, $\mu(0, 1)$ is not zero, and its sign is the opposite of that $\mu(0, x)$ for any dual atom x. This concludes the proof.

Corollary. *The coefficients of the characteristic polynomial of a geometric lattice alternate in sign.*

We next derive a combinatorial interpretation of the Euler characteristic of a geometric lattice, which generalizes a technique first used by WHITNEY in the study of linear graphs.

A subset $\{a, b, \ldots, c\}$ of a geometric lattice L is *independent* when

$$r(a \vee b \vee \cdots \vee c) = r(a) + r(b) + \cdots + r(c).$$

Let C_k be the cross-cut of L of all elements of rank $k > 0$. A maximal independent subset $\{a, b, \ldots, c\} \subset C_k$ is a *basis* of C_k. All bases of C_k have the same number of elements, namely, $n - k$ if the lattice has height n. A subset $A \subset C_k$ is a *circuit* (WHITNEY) when it is not independent but every proper subset is independent. A set is independent if and only if it contains no circuits.

Order the elements of L of rank k in a linear order, say $a_1, a_2, \ldots, a_l$. This ordering induces a lexicographic ordering of the circuits of C_k.

If the subset $\{a_{i_1}, a_{i_2}, \ldots, a_{i_j}\}$ $(i_1 < i_2 < \cdots < i_j)$ is a circuit, the subset $a_{i_1}, a_{i_2}, \ldots, a_{i_{j-1}}$ will be called a *broken circuit*.

Proposition 1. *Let L be a geometric lattice of height $n + 1$, and let C_k be the cross-cut of all elements of rank k. Then $\mu(0, 1) = (-1)^n m_k$, where m_k is the number of subsets of C_k whose meet is 0, containing $n - k + 1$ elements each, and not containing all the arcs of any broken circuit.*

Again, the assertion implies that $m_1 = m_2 = m_3 = \cdots$.

Proof. Let the lexicographically ordered broken circuits be $P_1, P_2, \ldots, P_\sigma$, and let S_i be the family of all spanning subsets of C_k containing P_i but not $P_1, P_2, \ldots$, or P_{i-1}. In particular, $S_{\sigma+1}$ is the family of all those spanning subsets not containing all the arcs of any broken circuit. Let q_j^i be the number of spanning subsets of j elements and not belonging to S_i. We shall prove that for each $i \geqq 1$

$$(*) \qquad \mu(0, 1) = q_2^i - q_3^i + q_4^i \cdots.$$

First, set $i = 1$. The set S_1 contains all spanning subsets containing the broken circuit P_1. Let $\bar{P}_1$ be the cicuit obtained by completing the broken circuit P_1. — A spanning set contained in S_1 contains either $\bar{P}_1$ or else P_1 but not $\bar{P}_1$; call these two families of spanning subsets A and B, and let q_j^A and q_j^B be defined accordingly. Then $q_j = q_j^1 + q_j^A + q_j^B$, and

$$\mu(0, 1) = q_2 - q_3 + q_4 \cdots = q_2^1 - q_3^1 + \cdots + \\ + q_2^A + (q_2^B - q_3^A) - (q_3^B - q_4^A) + \cdots.$$

Now, $q_2^A = 0$, because no circuit can contain two elements; there is a one-to-one correspondence between the elements of A and those of B, obtained by completing the broken circuit P_1. Thus, all terms in parentheses cancel and the identity (*) holds for $i = 1$.

To prove (*) for $i > 1$, remark that the element c_i of C_k, which is dropped from a circuit to obtain the broken circuit P_i, does not occur in any of the previous circuits, because of the lexicographic ordering of the circuits. Hence the induction can be continued up to $i = \sigma + 1$.

Any set belonging to $S_{\sigma+1}$ does not contain any circuit. Hence, it is an independent set. Since it is a spanning set, it must contain $n - k + 1$ elements. Thus, all the integers $q_{\sigma+1}$ vanish except $q_{n-k+1}^{\sigma+1}$ and the statement follows from (*), q. e. d.

Corollary 1. *Let* $q(\lambda) = \lambda^n + m_1\lambda^{n-1} + m_2\lambda^{n-2} + \cdots + m_n$ *be the characteristic polynomial of a geometric lattice of height* $n + 1$. *Then* $(-1)^k m_k$ *is a positive integer for* $1 \leqq k \leqq n$, *equal to the number of independent subsets of* k *atoms not containing any broken circuit.*

The *proof* is immediate: take $k = 1$ in the preceding Proposition.

The homology of a geometric lattice is simpler than that of a general lattice:

Proposition 2. *In the homology relative to the cross-cut* C_k *of all elements of rank* $k = 1$, *the Betti numbers* $b_1, b_2, \ldots, b_{k-2}$ *vanish.*

The *proof* is not difficult.

Example 1. *Partitions of a set.*

Let S be a finite set of n elements. A *partition* π of S is a family of disjoint subsets $B_1, B_2, \ldots, B_k$, called *blocks*, whose union is S. There is a (well-known) natural ordering of partitions, which is defined as follows: $\pi \leqq \sigma$ whenever every block of π is contained in a block of partition σ. In particular, 0 is the partition having n blocks, and I is the partition having one block. In this ordering, the partially ordered set of partitions is a geometric lattice (cf. Birkhoff).

The Möbius function for the lattice of partitions was first determined by Schützenberger and independently by Roberto Frucht and the author. We give a new proof which uses a recursion. If π is a partition, the *class* of π is the (finite) sequence $(k_1, k_2, \ldots)$, where k_i is the number of blocks with i elements.

Lemma. *Let* L_n *be the lattice of partitions of a set with* n *elements. If* $\pi \in L_n$ *is of rank* k, *then the segment* $[\pi, 1]$ *is isomorphic to* L_{n-k}. *If* π *is of class* $(k_1, k_2, \ldots)$, *then the segment* $[0, \pi]$ *is isomorphic to the direct product of* k_1 *lattices isomorphic to* L_1, k_2 *lattices isomorphic to* L_2, etc.

The *proof* is immediate.

It follows from the Lemma that if $[x, y]$ is a segment of L_n, then it is isomorphic to a product of k_i lattices isomorphic to L_i, $i = 1, 2, \ldots$. We call the sequence $(k_1, k_2, \ldots)$ the *class* of the *segment* $[x, y]$.

Proposition 3. *Let* $\mu_n = \mu(0, 1)$ *for the lattice of partitions of a set with* n *elements. Then* $\mu_n = (-1)^{n-1}(n-1)!$.

Proof. By the Corollary to Proposition 4 of Section 5, $\sum_{x \wedge a = 0} \mu(x, 1) = 0$. Let a be the dual atom consisting of a block C_1 containing $n - 1$ points, and a second block C_2 containing one point. Which non-zero partitions x have the property that $x \wedge a = 0$? Let the blocks of such a partition x be $B_1, \ldots, B_k$. None of the blocks B_i can contain two distinct points of the block C_1, otherwise the two points would still belong to the same block in the intersection. Furthermore, only one of the B_i can contain the block C_2. Hence, all the B_i contain one point, except one, which contains C_2 and an extra point. We conclude that x must be an atom, and there are $n - 1$ such atoms. Hence, $\mu_n = \mu(0, 1) = -\sum_x \mu(x, 1)$, where x ranges over a set of $n - 1$ atoms. By the Lemma, the segment $[x, 1]$ is isomorphic to the lattice of partitions of a set with $n-1$ elements, hence $\mu_n = -(n-1)\mu_{n-1}$. Since $\mu_2 = -1$, the conclusion follows.

Corollary. *If the segment* $[x, y]$ *is of class* $(k_1, k_2, \ldots, k_n)$, *then*

$$\mu(x, y) = \mu_1^{k_1} \mu_2^{k_2} \ldots \mu_n^{k_n} = (-1)^{k_1+k_2+\cdots+k_n-n} (2!)^{k_3} (3!)^{k_4} \ldots ((n-1)!)^{k_n}.$$

The Möbius inversion formula on the partitions of a set has several combinatorial applications; see the author's expository paper on the subject.

8. Representations

There is, as is well known, a close analogy between combinatorial results relating to Boolean algebras and those relating to the lattice of subspaces of a vector space. This analogy is displayed for example in the theory of q-difference equations developed by F. H. JACKSON, and can be noticed in many number-theoretic investigations. In view of it, we are led to surmise that a result analogous to Proposition 1 of Section 5 exists, in which the Boolean algebra of subsets of R is replaced by a lattice of subspaces of a vector space over a finite field. Such a result does indeed exist; in order to establish it a preliminary definition is needed.

Let L be a finite lattice, and let V be a finite-dimensional vector space over a finite field with q elements. A *representation* of L over V is a monotonic map p of L into the lattice M of subspaces of V, having the following properties:

(1) $p(0) = 0$.

(2) $p(a \vee b) = p(a) \vee p(b)$.

(3) Each atom of L is mapped to a line of the vector space V, and the set of lines thus obtained spans the entire space V.

A representation is *faithful* when the mapping p is one-to-one. We shall see in Section 9 that a great many ordered structures arising in combinatorial problems admit faithful representations. Given a representation $p: L \to M$, one defines the *conjugate map* $q: M \to L$ as follows.

Let K be the set of atoms of M (namely, lines of V), and let A be the image under p of the set of atoms of L. For $s \in M$, let $K(s)$ be the set of atoms of M dominated by s, and let $B(s)$ be a minimal subset of A which spans (in the vector space sense) every element of $K(s)$. Let $A(s)$ be the subset of A which is spanned by $B(s)$. A simple vector-space argument, which is here omitted, shows that the set $A(s)$ is well defined, that is, that it does not depend upon the choice of $B(s)$, but only upon the choice of s.

Let $C(s)$ be the set of atoms of L which are mapped by p onto $A(s)$. Set $q(s) = \vee C(s)$ in the lattice L; this defines the map q. It is obviously a monotonic function.

Lemma. *Let* $p: L \to M$ *be a faithful representation and let* $q: M \to L$ *be the conjugate map. Assume that every element of* L *is a join of atoms. Then* $p(q(s)) \geqq s$ *and* $q(p(x)) \leqq x$.

Proof. By definition, $q(s) = \vee C(s)$, where $C(s)$ is the inverse image of $A(s)$ under p. By property (2) of a representation,

$$p(q(s)) = p(\vee C(s)) = \vee p(C(s)) = \vee A(s).$$

But this join of the set of lines $A(s)$ in the lattice M is the same as their span in the vector space V. Hence $\bigvee A(s) \geqq s$, and we conclude that $p(q(s)) \geqq s$.

To prove that $q(p(x)) \leqq x$, it suffices to show that $A(p(x)) = B$, where B is the set of atoms in A dominated by $p(x)$. Clearly $B \subset A(p(x))$, and it will suffice to establish the converse implication. By (2), and by the fact that x is a join of atoms, we have $p(x) = \bigvee B$. Therefore every line l dominated by $p(x)$ is spanned by a subset of B. If in addition $l \in A$, then $l \leqq \bigvee C$ for some subset $C \subset B$, hence $l \in B$. This shows $B \supset A(p(x))$, q. e. d.

Theorem 5. *Let L be a finite lattice, where every element is a join of atoms, let $p: L \to M$ be a faithful representation of L into the lattice M of subspaces of a vector space V over a finite field with q elements, and let $q: M \to L$ be the conjugate map. For every $k \geqq 2$, let m_k be the number of k-dimensional subspaces s of V such that $q(s) = I$. Then*

$$(*) \qquad \mu(0, 1) = q^{\binom{2}{2}} m_2 - q^{\binom{3}{2}} m_3 + q^{\binom{4}{2}} m_4 - \cdots,$$

where μ is the Möbius function of L.

Proof. Let $Q = L^*$, let $c: L \to Q$ and $c^*: Q \to L$ be the canonical isomorphisms between L and Q. Define $\pi: Q \to M$ as $\pi = pc^*$, and $\varrho: M \to Q$ as $\varrho = cq$. We verify that π and ϱ give a Galois connection between Q and M satisfying the hypothesis of Theorem 1. If $\pi(x) = 0$, then there is a $y \in L$ such that $y = c^*(x)$ and $p(y) = 0$. It follows from the definition of a representation that $y = 0$. Hence $x = c(y) = 1$. Furthermore, $\varrho(0) = c(q(0)) = 1$. It follows from the preceding Lemma that π and ϱ are a Galois connection. Applying Theorem 1 and the result of Example 2 of Section 5, formula (*) follows at once.

Remark. It is easy to see that every lattice having a faithful representation is a geometric lattice. The converse is however not true, as an example of T. Lazarson shows.

A reduction similar to that of Proposition 1 of Section 7 can be carried out with Theorem 5 and representations, and another combinatorial property of the Euler characteristic is obtained.

9. The coloring of graphs

By way of illustration of the preceding theory, we give some applications to the classic problem of coloring of graphs, and to the problem of constructing flows in networks with specified properties. Our results extend previous work of G. D. Birkhoff, D. C. Lewis, W. T. Tutte and H. Whitney.

A *linear graph* $G = (V, E)$ is a structure consisting of a finite set V, whose elements are called vertices, together with a family E of two-element subsets of V, called edges. Two vertices a and b are adjacent when the set (a, b) is an edge; the vertices a and b are called the endpoints of (a, b). Alternately, one calls the vertices *regions* and calls the graph a *map*, and we use the two terms interchangeably, considering them as two words for the same object. If S is a set of edges, the *vertex set* $V(S)$ consists of all vertices which are incident to some edge in S.

A set of edges S is *connected* when in any partition $S = A \cup B$ into disjoint non-empty sets A and B, the vertex sets $V(A)$ and $V(B)$ are not disjoint. Every set of edges is the union of disjoint connected *blocks*.

The *bond closure* on a graph $G = (V, E)$ is a closure relation defined on the set E of edges as follows. If $S \subset E$, let $\bar{S}$ be the set of all edges both of whose endpoints belong to one and the same block of S. Every set consisting of a single edge is closed, and these are the only minimal non-empty closed sets.

Lemma 1. *The bond closure* $S \to \bar{S}$ *has the exchange property.*

Proof. Suppose e and f are edges, $S \subset E$, and $e \in \overline{S \cup f}$ but $e \notin \bar{S}$. Then every endpoint of e which is not in $V(S)$ is an endpoint of f; on the other hand, S and f have at least one point in common, otherwise $e \in \bar{S}$. Thus both e and f either connect the same two blocks of S, or else they have one endpoint in S and one common endpoint; hence $f \in \overline{S \cup e}$, q.e.d.

The lattice $L = L(G)$ of bond-closed subsets of E is called the *bond lattice* of the graph G. Suppose that E has n blocks and $p(\lambda)$ is the characteristic polynomial of L, then the polynomial $\lambda^n p(\lambda)$ is the *chromatic polynomial* of the graph G, first studied by G. D. BIRKHOFF. From Theorem 4 we infer at once the theorem of WHITNEY that the coefficients of the chromatic polynomial alternate in sign.

The chromatic polynomial has the following combinatorial interpretation. Let C be a set of n elements, called colors. A function $f: V \to C$ is a *proper coloring* of the graph, when no two adjacent vertices are assigned the same color. To every coloring f — not necessarily proper — there corresponds a subset of E, the *bond* of f, defined as the set of all edges whose endpoints are assigned the same color by f. The bond of f is a closed set of edges. For every closed set S, let $p(\lambda, S)$ be the number of colorings whose bond is S. Then we shall prove that $p(\lambda, S) = \lambda^n q(\lambda, S)$, where $q(\lambda, S)$ is the characteristic polynomial of the segment $[S, I]$ in the lattice L. Since every coloring has a bond $\sum_{T \geqq S} p(\lambda, T)$ equals the total number of colorings having some bond $T \geqq S$. But this number is evidently $\lambda^{k-r(s)}$, where k is the number of vertices of the graph and $r(S)$ is the rank of S in L. Applying the Möbius inversion formula on the bond-lattice, we get

$$(*) \qquad p(\lambda) = p(\lambda, 0) = \sum_{T \in L} \lambda^{k-r(T)} \mu(0, T).$$

But the number of colorings whose bond is the null set 0 is exactly the number of proper colorings.

WHITNEY's evaluation (cf. A logical expansion in Mathematics) of the chromatic polynomials of a graph in terms of the number of subgraphs of s edges and p connected components is an immediate consequence of the cross-cut theorem applied to the atoms of the bond-lattice of G. This result of WHITNEY's can now be sharpened in two directions: first, a cross-cut other than that of the atoms can be taken; secondly, the computation of the coefficients of the chromatic polynomial can be simplified by Proposition 1 of Section 8. The cross-cut of all elements of rank 2 is particularly suited for computation, and can be programmed. The interested reader may wish to explicitly translate the cross-cut theorem and the results of Section 8 into the geometric language of graphs.

Example 1. For a *complete graph* on n vertices, where every two-element subset is an edge, the bond-lattice is isomorphic to the lattice of partitions of a set with n elements. The chromatic polynomial is evidently $(\lambda)_n = \lambda(\lambda - 1) \ldots (\lambda - n + 1)$, and the coefficients $s(n, k)$ are the *Stirling numbers of the first kind.*

Thus, $\sum_{r(\pi)=k} \mu(0, \pi) = s(n, k)$. This gives a combinatorial interpretation to the Stirling numbers of the first kind.

For a map $\mathfrak{m}$ embedded in the plane, where regions and boundaries have their natural meaning and no region bounds with itself, one obtains an interesting geometric result by applying the cross-cut theorem to the dual atoms of the bond lattice $L(\mathfrak{m})$.

Let $\mathfrak{m}$ be a connected map in the plane; without loss of generality we can assume: (a) that all the regions of $\mathfrak{m}$, except one which is unbounded, lie inside a convex polygon, the outer boundary of $\mathfrak{m}$; (b) that all boundaries are segments of straight lines. The *dual graph* of $\mathfrak{m}$ is the linear graph made up of the boundaries of $\mathfrak{m}$. A *circuit* in a linear graph is defined as a simple closed curve contained in the graph. We give an expression of the polynomial $P(\lambda, \mathfrak{m})$ in terms of the circuits of the dual graph. The outer boundary is always a circuit.

A set of circuits of a map $\mathfrak{m}$ in the plane *spans*, when their union — in the set-theoretic sense — is the entire boundary of $\mathfrak{m}$.

Proposition 1. *For every integer $k \geqq 1$, let C_k be the number of spanning sets of k distinct circuits of a map $\mathfrak{m}$ in the plane. Then*

$$\mu_{\mathfrak{m}}(0, 1) = -C_1 + C_2 - C_3 + C_4 - \cdots$$

Proof. If the map has two regions, then $C_1 = 1$ and all other $C_c = 0$, so the result is trivial. Assume now that $\mathfrak{m}$ has at least 3 regions. Then $C_1 = 0$. All we have to prove is that the integers C_k are the integers q_k of Theorem 3, relative to the cross-cut of $L(\mathfrak{m})$ consisting of all the dual atoms.

By the Jordan curve theorem, every circuit divides the plane into two regions; this gives a one-to-one correspondence of the circuits with the dual atoms of $L(\mathfrak{m})$. Conversely, because we can assume that the map is of the special type described above, every dual atom in $L(\mathfrak{m})$ is a map with two connected regions, and so must have as a boundary a simple closed curve, q.e.d.

It has been shown by Richard Rado (p. 312) that the bond-lattice $L(G)$ of any linear graph G has a faithful representation. Accordingly, Theorem 5 can also be applied to obtain expression for $\mu(0,1)$. These expressions usually give sharper bounds than similar expressions based upon the cross-cut of atoms.

Farther-reaching techniques for the computation of the Möbius function of $L(G)$ are obtained by applying Theorem 1 to situations where P and Q are both bond-lattices of graphs. This we shall now do. A *monomorphism* of a graph G into a graph H is a one-to-one function f of the vertices of G *onto* the vertices of H, which induces a map $\bar{f}$ of the edges of G *into* the edges of H. Every monomorphism $f: G \to H$ induces a monotonic map $p: L(G) \to L(H)$, where $p(S)$ is defined as the closure of the image $\bar{f}(S)$ in H. It also induces a monotonic map $q: L(H) \to$ $\to L(G)$, where $q(T)$ is defined as the set of edges of G whose image is in T.

Lemma 2. $q(p(S)) = S$ *for S in $L(G)$ and* $p(q(T)) \leqq T$ *for T in $L(H)$.*

Proof. Intuitively, $p(S)$ is obtained by "adding edges" to S, and $q(p(S))$ simply removes the added edges. Thus, the first statement is graphically clear. The second one can be seen as follows. $q(T)$ is obtained from T by removing a number of edges. Taking $p(q(T))$, some of the edges may be replaced, but in general not all. Thus, $p(q(T)) \leqq T$.

Taking $M = L(H)^*$ and $c: L(H) \to M$ to be the canonical order-inverting map, we see that $\pi = cp$ and $\zeta = qc$ give a Galois connection between $L(G)$ and M. Now, $\pi(x) = 0$ is equivalent to $p(x) = 1$ for $x \in L(G)$. This can happen only if x has only one component, that is — since x is closed — only if $x = 1$ in $L(G)$. Thus $\pi(x) = 0$ if and only if $x = 1$. Secondly, $\varrho(0) = q(1) = 1$, evidently. We have verified all the hypotheses of Theorem 1, and we therefore obtain:

Proposition 2. *Let $f: G \to H$ be a monomorphism of a linear graph G into a linear graph H, and let μ_G and μ_H be the Möbius functions of the bond-lattices. Then*

$$\mu_G(0, 1) = \sum_{[a \in L(H);\, q(a)=0]} \mu_H(a, 1),$$

where q is the map of $L(H)$ into $L(G)$ naturally associated with f, as above.

Proposition 1 can be used to derive a great many of the reductions of G. D. BIRKHOFF and D. C. LEWIS, and provides a systematic way of investigating the changes of Möbius functions — and hence of the chromatic polynomial — when edges of a graph are removed. It has a simple geometric interpretation.

An interesting application is obtained by taking H to be the complete lattice on n elements. We then obtain a formula for μ which completes the statements of Theorems 3 and 5. Let G be a linear graph on n vertices. Let C be the family of two-element subsets of G which are not edges of G. Let F be the family of all subsets of C which are closed sets in the bond-lattice of the complete graph on n vertices built on the vertices of G. Then,

Corollary.
$$\mu_G(0, 1) = \sum_{a \in F} \mu(a, 1),$$

where μ is the Möbius function of the lattice of partitions (cf. Example 5) of a set of n elements.

Stronger results can be obtained by considering "epimorphisms" rather than "monomorphisms" of graphs, relating μ_G to the Möbius function obtained from G by "coalescing" points. In this way, one makes contact with G. A. DIRAC's theory of critical graphs. We leave the development of this topic to a later work.

10. Flows in networks

A network $N = (V, E)$ is a finite set V of vertices, together with a set of ordered pairs of vertices, called edges.

We shall adopt for networks the same language as for linear graphs.

A *circuit* is a sequence of edges S such that every vertex in $V(S)$ belongs to exactly two edges of S. Every edge has a positive and a negative endpoint. Given a function Φ from E to the integers from 0 to $\lambda - 1$, let for each vertex v, $\overline{\Phi}(v)$ be defined as

$$\overline{\Phi}(v) = \sum_e \eta(e, v)\, \Phi(e),$$

where the sum ranges over all edges incident to v, and the function $\eta(e, v)$ takes the value $+1$ or -1 according as the positive or negative end of the edge e abuts at the vertex v, and the value zero otherwise. The function Φ is a *flow* (mod. λ) when $\overline{\Phi}(v) \equiv 0$ (mod. λ) for every vertex v. The value $\Phi(e)$ for an edge e is called the *capacity* of the *flow* through e. The mod. λ restriction is inessential, but will be kept throughout.

A *proper flow* is one in which no edge is assigned zero capacity. TUTTE was the first to point out the importance of the problem of counting proper flows (cf. A contribution to the theory of chromatic polynomials) in combinatorial theory.

We shall reduce the solution of the problem to a Möbius inversion on a lattice associated with the network. This will give an expression for the number of proper flows as a polynomial in λ, whose coefficients are the values of a Möbius function.

Every flow through N is a proper flow of a suitable subnetwork of N, obtained by removing those edges which are assigned capacity 0. However, the converse of this assertion is not true: given a subnetwork S of N, it may not be possible to find a flow which is proper on the complement of N. This happens because every flow which assigns capacity zero to each edge of S may assign capacity zero to some further edges. We are therefore led to define a closure relation on the set of all subgraphs as follows: $\bar{S}$ shall be the set of all edges which necessarily are assigned capacity zero, in any flow of N which assigns capacity zero to every edge of S. In other words, if $e \notin \bar{S}$, then there is a flow in N which assigns capacity $\neq 0$ to the edge e, but which assigns capacity zero to all the edges of $\bar{S}$. It is immediately verified that $S \to \bar{S}$ is a closure relation. We call it the *circuit closure* of S. The circuit closure *has the exchange property*: if $e \in \overline{S \cup p}$ but $e \notin \bar{S}$, then $p \in \overline{S \cup e}$. Before verifying it, we first derive a geometric characterization of the circuit closure. A set S is circuit closed ($S = \bar{S}$) if and only if through every edge e not in S there passes a circuit which is disjoint from S. For if S is closed and $e \notin S$, then there is a flow through e and disjoint from S. But this can happen only if there is a circuit through e.

If there is a circuit through the edge p disjoint from $\overline{S \cup e}$, and a circuit through e disjoint from $\bar{S}$ and containing p, then there is — as has been observed by WHITNEY — also a circuit through e not containing $\bar{S} \cup p$. This implies that e is not in the closure of $\bar{S} \cup p$, and verifies the exchange property.

The lattice $C(N)$ of closed subsets of edges of the network N is the *circuit lattice of* N. An atom in this lattice is not necessarily a single edge.

Proposition 1. *The number of proper flows,* (mod. λ) *on a network N with v vertices, e edges and p connected components is a polynomial $p(\lambda)$ of degree $e - v + p$. This polynomial is the characteristic polynomial of the circuit lattice of N. The coefficients alternate in sign.*

Proof. The last statement is an immediate consequence of Theorem 4 of Section 8.

The total number of flows on N (not necessarily proper) is determined as follows. Assume for simplicity that N is connected. Remove a set D of $v - 1$ edges from N, one adjacent to each but one of the vertices.

Every flow on N can be obtained by first assigning to each of the edges not in D an arbitrary capacity, between 0 and $\lambda - 1$, and then filling in capacities for the edges in D to match the requirement of zero capacity through each vertex. There are λ^{e-v+1} ways of doing this, and this is therefore the total number of flows mod. λ. If the network is in p connected components, the same argument gives λ^{e-v+p}. Now, every flow on G is a proper flow on a unique closed subset $\bar{S}$, obtained by removing all edges having capacity zero.

Hence

$$\lambda^{e-v+p} = \sum_{\bar{S} \in \underline{C}(G)} p(\bar{S}, \lambda),$$

where $p(\bar{S}, \lambda)$ is the characteristic polynomial of the closed subgraph $\bar{S}$. Setting $n(s) = e(s) - v(s) + p(s)$, the number of edges, vertices and components of s, and applying the inversion formula, we get

$$p(G, \lambda) = \sum_{S \in \underline{C}(G)} \lambda^{n(s)} \mu(S, G), \quad \text{q.e.d.}$$

In the course of the proof we have also shown that $n(s)$ is the *rank* of S in the circuit lattice of G. The rank of the null subgraph is one.

The four-color problem is equivalent to the statement that every planar network without an isthmus has a proper flow mod 5. (An isthmus is an edge that disconnects a component of the network when removed.)

Most of the results of the preceding section extend to circuit lattices of a network, and give techniques for computation of the flow polynomials of networks. We shall not write down their translation into the geometric language of networks.

References

AUSLANDER, L., and H. M. TRENT: Incidence matrices and linear graphs. J. Math. Mech. **8**, 827 — 835 (1959).

BELL, E. T.: Algebraic Arithmetic. New York: Amer. Math. Soc. (1927),

— Exponential polynomials. Ann. of Math., II. Ser. **35**, 258—277 (1934).

BERGE, C.: Théorie des graphes et ses applications. Paris: Dounod 1958.

BIRKHOFF, GARRETT: Lattice Theory, third preliminary edition. Harvard University, 1963.

— Lattice Theory, revised edition. American Mathematical Society, 1948.

BIRKHOFF, G. D.: A determinant formula for the number of ways of coloring a map. Ann. of Math., II. Ser. **14**, 42—46 (1913).

—, and D. C. LEWIS: Chromatic polynomials. Trans. Amer. math. Soc. **60**, 355—451 (1946).

BLEICHER, M. N., and G. B. PRESTON: Abstract linear dependence relations. Publ. Math., Debrecen **8**, 55—63 (1961).

BOUGAYEV, N. V.: Theory of numerical derivatives. Moscow, 1870—1873, pp. 1—222.

BRUIJN, N. G. DE: Generalization of Polya's fundamental theorem in enumerative combinatorial analysis. Indagationes math. **21**, 59—69 (1959).

CHUNG, K.-L., and L. T. C. HSU: A combinatorial formula with its application to the theory of probability of arbitrary events. Ann. math. Statistics **16**, 91—95 (1945).

DEDEKIND, R.: Gesammelte Mathematische Werke, volls. I—II—III. Hamburg: Deutsche Math. Verein. (1930).

DELSARTE, S.: Fonctions de Möbius sur les groupes abéliens finis. Ann. of Math., II. Ser. **49**, 600—609 (1948).

DILWORTH, R. P.: Proof of a conjecture on finite modular lattices. Ann. of Math., II. Ser. **60**, 359—364 (1954).

DIRAC, G. A.: On the four-color conjecture. Proc. London math. Society, III. Ser. **13**, 193 to 218 (1963).

DOWKER, C. H.: Homology groups of relations. Ann. of Math., II. Ser. **56**, 84—95 (1952).

DUBREIL-JACOTIN, M.-L., L. LESIEUR et R. CROISOT: Leçons sur la théorie des treilles des structures algebriques ordonnées et des treilles géometriques. Paris: Gauthier-Villars 1953.
EILENBERG, S., and N. STEENROD: Foundations of algebraic topology. Princeton: University Press 1952.
FARY, I.: On straight-line representation of planar graphs. Acta Sci. math. Szeged **11**. 229—233 (1948).
FELLER, W.: An introduction to probability theory and its applications, second edition. New York: Wiley 1960.
FRANKLIN, P.: The four-color problem. Amer. J. Math. **44**, 225—236 (1922).
FRÉCHET, M.: Les probabilités associées à un système d'évenements compatibles et dépendants. Actualitées scientifiques et industrielles, nos. 859 et 942. Paris: Hermann 1940 et 1943.
FRONTERA MARQUÉS, B.: Una función numérica en los retículos finitos que se anula para los retículos reducibles. Actas de la 2a, Reunión de matemáticos españoles. Zaragoza 103—111 1962.
FRUCHT, R., and G.-C. ROTA: La función de Möbius para el retículo di particiones de un conjunto finito. To appear in Scientia (Chile).
GOLDBERG, K., M. S. GREEN and R. E. NETTLETON: Dense subgraphs and connectivity. Canadian J. Math. **11** (1959).
GOLOMB, S. W.: A mathematical theory of discrete classification. Fourth Symposium in Information Theory, London, 1961.
GREEN, M. S., and R. E. NETTLETON: Möbius function on the lattice of dense subgraphs. J. Res. nat. Bur. Standards **64**B, 41—47 (1962).
— — Expression in terms of modular distribution functions for the entropy density in an infinite system. J. Chemical Physisc **29**, 1365—1370 (1958).
HADWIGER, H.: Eulers Charakteristik und kombinatorische Geometrie. J. reine angew. Math. **194**, 101—110 (1955).
HALL, PHILIP: A contribution to the theory of groups of prime power order. Proc. London math. Soc., II. Ser. **36**, 39—95 (1932).
— The Eulerian functions of a group. Quart. J. Math. Oxford Ser. 134—151, 1936.
HARARY, F.: Unsolved problems in the enumeration of graphs. Publ. math. Inst. Hungar. Acad. Sci. **5**, 63—95 (1960).
HARDY, G. H.: Ramanujan. Cambridge: University Press 1940.
—, and E. M. WRIGHT: An introduction to the theory of numbers. Oxford: University Press 1954.
HARTMANIS, J.: Lattice theory of generalized partitions. Canadian J. Math. **11**, 97—106 (1959).
HILLE, E.: The inversion problems of Möbius. Duke math. J. **3**, 549—568 (1937).
HSU, L. T. C.: Abstract theory of inversion of iterated summation. Duke math. J. **14**, 465 to 473 (1947).
— On Romanov's device of orthogonalization. Sci. Rep. Nat. Tsing Hua Univ. **5**, 1—12 (1948).
— Note on an abstract inversion principle. Proc. Edinburgh math. Soc. (2) **9**, 71—73 (1954).
JACKSON, F. H.: Series connected with the enumeration of partitions. Proc. London math. Soc., II. Ser. **1**, 63—88 (1904).
— The q-form of Taylor's theorem. Messenger of Mathematics **38**, 57—61 (1909).
JÓNSSON, B.: Lattice-theoretic approach to projective and affine geometry. Symposium on the Axiomatic Method. Amsterdam, North-Holland Publishing Company, 1959, 188—205.
—, and A. TARSKI: Direct decomposition of finite algebraic systems. Notre Dame Mathematical lectures, no. 5. Indiana: Notre Dame 1947.
KAC, M., and J. C. WARD: A combinatorial solution of the two-dimensional Ising model. Phys. Review 88, 1332—1337 (1952).
KAPLANSKI, I., and J. RIORDAN: The problème des ménages. Scripta math. **12**, 113—124 (1946).
KLEE, V.: The Euler characteristic in combinatorial geometry. Amer. math. Monthly **70**, 119—127 (1963).

LAZARSON, T.: The representation problem for independence functions. J. London math. Soc. **33**, 21—25 (1958).
MACLANE, S.: A lattice formulation of transcendence degrees and p-bases. Duke math. J. **4**, 455—468 (1938).
MACMILLAN, B.: Absolutely monotone functions. Ann. of Math., II. Ser. **60**, 467—501 (1954).
MÖBIUS, A. F.: Über eine besondere Art von Umkehrung der Reihen. J. reine angew. Math. **9**, 105—123 (1832).
ORE, O.: Theory of graphs. Providence: American Mathematical Society 1962.
POLYA, G.: Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. Acta math. **68**, 145—253 (1937).
RADO, R.: Note on independence functions. Proc. London math. Soc., III. Ser. **7**, 300—320 (1957).
READ, R. C.: The enumeration of locally restricted graphs, I. J. London math. Soc. **34**, 417 to 436 (1959).
REDFIELD, J. H.: The theory of group-reduced distributions. Amer. J. Math. **49**, 433—455 (1927).
REVUZ, ANDRÉ: Fonctions croissantes et mesures sur les espaces topologiques ordonnés. Ann. Inst. Fourier **6** 187—268 (1955).
RIORDAN, J.: An introduction to combinatorial analysis. New York: Wiley 1958.
ROMANOV, N. P.: On a special orthonormal system and its connection with the theory of primes. Math. Sbornik, N. S. **16**, 353—364 (1945).
ROTA, G.-C.: Combinatorial theory and Möbius functions. To appear in Amer. math. Monthly.
— The number of partitions of a set. To appear in Amer. math. Monthly.
RYSER, H. J.: Combinatorial Mathematics. Buffalo: Mathematical Association of America 1963.
SCHÜTZENBERGER, M. P.: Contribution aux applications statistiques de la théorie de l'information. Publ. Inst. Stat. Univ. Paris, **3**, 5—117 (1954).
TARSKI, A.: Ordinal algebras. Amsterdam: North-Holland Publishing Company 1956.
TOUCHARD, J.: Sur un problème de permutations. C. r. Acad. Sci., Paris, **198**, 631—633 (1934).
TUTTE, W. T.: A contribution to the theory of chromatic polynomials. Canadian J. Math. **6**, 80—91 (1953).
— A class of Abelian group. Canadian J. Math. 8, 13—28 (1956).
— A homotopy theorem for matroids, I. and II. Trans. Amer. math. Soc. 88, 144—140 (1958).
— Matroids and graphs. Trans. Amer. math. Soc. **90**, 527—552 (1959).
WARD, M.: The algebra of lattice functions. Duke math. J. **5**, 357—371 (1939).
WEISNER, L.: Abstract theory of inversion of finite series. Trans. Amer. math. Soc. **38**, 474—484 (1935).
— Some properties of prime-power groups. Trans. Amer. math. Soc. **38**, 485—492 (1935).
WHITNEY, H.: A logical expansion in mathematics. Bull. Amer. math. Soc. **38**, 572—579 (1932).
— Characteristic functions and the algebra of logic. Ann. of Math., II. Ser. **34**, 405—414 (1933).
— The abstract properties of linear dependence. Amer. J. Math. **57**, 507—533 (1935).
WIELANDT, H.: Beziehungen zwischen den Fixpunktzahlen von Automorphismengruppen einer endlichen Gruppe. Math. Z. **73**, 146—158 (1960).
WINTNER, A.: Eratosthenian Averages. Baltimore (privately printed) 1943.

Department of Mathematics
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

(Received September 2, 1963)

On the Foundations of Combinatorial Theory II. Combinatorial Geometries

By Henry H. Crapo and Gian-Carlo Rota

It is a pleasure to dedicate this work to the founders of the theory:

Garrett Birkhoff
Robert Dilworth
Saunders MacLane
Richard Rado
W. T. Tutte
Hassler Whitney

Faceste come quei che va di notte,
Che porta il lume dietro e sé non giova,
Ma dopo sé fa le persone dotte
Purg. xxii, 67

1. Introduction

The purpose of the present series of papers is to give a systematic and thorough exposition of the foundations of the theory of combinatorial geometry.

It has become clear in the last ten years that the concept of combinatorial geometry, and its applied-mathematical counterpart, that of matroid or pregeometry, may well play in the current development of combinatorics a catalyzing role not unlike that of point-set topology in the development of functional analysis. Unfortunately, all expositions of relevant parts of the theory that are presently found in the literature take one of several one-sided approaches which obscure the unifying role and the broader interest of the theory. Thus, Hassler Whitney's pioneering paper of 1935, in which the main lines of the theory were set down, although still an introduction well worth reading, is definitely motivated by his desire to generalize the notion of a dual graph; the author abandoned the field shortly thereafter. Rado's work, important to the development of matching theory and the study of independence in infinite sets, proceeded in unfortunate isolation. The lattice-theoretic studies of the thirties, which led to the parallel idea of a geometric lattice, were confined to axiomatics (Birkhoff and the French School, M. L. Dubreil, Le Sieur, etc.) and algebraic dependence (MacLane). It remained for Dilworth in the forties to introduce the first significant combinatorial examples. Tutte's coordinatization papers of the fifties, which set the theory on its present footing, rely primarily upon graph-theoretic arguments, and ignore the geometric motivation made accessible by his predecessors.

The more recent work, some of it spectacular, by Barlotti, Berge, Bleicher, Bose, Brualdi, Camion, Dlab, M. L. Dubreil-Jacotin, Edmonds, Fulkerson, Gale, Higgs, Lehman, Marczewski, Minty, Mirsky, Perfect, Sachs, Schützenberger, Segre, Urbanik, Welsh, Wille, and by the present authors, serves only to underline the urgent need for a systematic unification of the theory.

The present work consists of the following main parts: (1) axiomatics of combinatorial geometry; (2) description of a large and disparate variety of examples; (3) a discussion of maps between geometries; (4) a brief presentation of the coordinatization theory; and (5) a sketch of our two main lines of future work, namely, the critical problem and matching theory.

The subject of combinatorial geometry can be broadly understood as the attempt to develop, and therewith generalize to a natural setting, what in nineteenth-century language would be called the theory of arithmetic invariants of finite sets of points in projective space. The term "arithmetic invariant" was used to distinguish (and quickly to dismiss) integer-valued invariants under the projective group, from the algebraic invariants (such as the well-known "brackets"), whose theory is now well understood. Remarkably, classical invariant theory failed to produce even a rudimentary theory of arithmetic invariants for point sets. Relevant examples of such invariants abound; the simplest is the rank of the given point set. Other examples can be gleaned from the canonical forms of the associated matrix and the rank of its minors. The most interesting arithmetic invariants arise from combinatorial considerations. Typical invariants include the numbers W_k of k-dimensional subspaces spanned by subsets of the given point set. (We call these numbers the *Whitney numbers* of the second kind.) Consider also the maximum size D of a family F of subspaces spanned by the points, subject to the restriction that no element of F should contain another (the *Dilworth number* of the point set.)

In fact, one can state in one short sentence those arithmetic invariants that fall under the scope of present methods. They are those which depend only upon the *lattice of subspaces* of the point set (such, evidently, are the rank, and the Whitney and Dilworth numbers just defined). The order-theoretic characterization of lattices arising in this way was achieved by Birkhoff in 1935, in immediate response to Whitney's pioneering effort. Such lattices, called *geometric lattices*, are defined as point lattices which have the Birkhoff covering property (see definition and discussion below). The projective-geometric and linear-algebraic qualities of such point sets are expressed by the MacLane-Steinitz *exchange axiom*. This axiom, as is well known, is all that is needed to derive the standard results about independence and bases in linear algebra.

We define a *combinatorial geometry* as a closure relation (see definition below) defined on subsets of a set S, and enjoying the exchange property.

(Of course, it is not in general true that the closure of the union of two sets is the union of their closures, as would be the case for a topological closure. We also add a finiteness condition, to simplify the exposition of the theory.) For simplicity, we also assume that every point in a geometry is a closed set. Without this additional assumption, the resulting structure is often described by the ineffably cacophonic term *matroid*, which we tend to avoid in favor of the term *pregeometry*. With each pregeometry (=matroid) there is canonically associated a geometry, the points of which are closures of points of the pregeometry. Therefore, when confronted with a matroid, the first safety measure to take is to run to the associated geometry whenever possible.

Clearly, a point set in any projective or affine space gives a combinatorial geometry, when closure is defined in terms of ordinary linear span. We hasten to add that not every combinatorial geometry is obtained from a point set in projective or affine space. A host of other combinatorial situations lead to significantly different examples, several of which are described in the text. By way of orientation, we now outline a few of these examples.

Instead of projective geometries one may invent more general geometries, such as circle geometries, or geometries which are obtained by taking various classes of algebraic varieties in the role of linear subspaces. The systematic study of these geometries has been recently undertaken by Wille (see Section 3.6). The geometric lattices associated with such generalizations of projective geometry turn out to be part distributive and part modular (They are not, however, modular.) An algebraic example is obtained by considering the lattice of relatively closed subfields of a field. This is a geometric lattice, as was first discovered by MacLane.

A remarkable example from combinatorial sources is the transversal geometry of Edmonds, Fulkerson, Mirsky and Perfect, which is obtained as follows. Given a relation $R \subseteq A \times B$ with finite sets A and B, one defines a subset I of A to be independent if the relation R, restricted to $I \times B$, dominates a one-one function from I into B, or *matching*. These are the independent sets of a pregeometry on the set A. Much of classical matching theory, from Philip Hall's marriage theorem to the deeper results of Tutte and others on the factorization of linear graphs, finds a pleasing systematization in the context of transversal geometries. When matching theory is formulated in geometric language, further extensions of the theory are suggested. We shall give the simplest result in this direction, the generalization of the marriage theorem to geometric lattices. This is a variant of a result of Rado.

The homology sequence on a simplicial complex provides a sequence of related geometries (see Chapter 6). This theory has been fully investigated only for one-dimensional complexes or graphs (Whitney, Tutte). The resulting interaction of topological and geometrical considerations promises to open up new connections with combinatorial topology.

Again along combinatorial lines, some remarkable examples of geometries are obtained by the following "measure-theoretic" process. The idea goes back to Dilworth. Let μ be an integer-valued function, on the subsets of a set S, with the following properties:

$$\mu(A) \geq \mu(B) \quad \text{if } A \text{ contains } B$$
$$\mu(A \cap B) + \mu(A \cup B) \leq \mu(A) + \mu(B)$$

Note that it is not assumed that μ is normalized by having any specified value on the empty set, or on one-element sets. We define a subset $I \subseteq S$ to be *independent* if $\mu(J) \geq \nu(J)$ for every nonempty subset J of I, where $\nu(J)$ denotes the size of the set J. The resulting pregeometries, whose theory is barely scratched, open a fertile line of investigation connecting with optimization theory, and with the little-investigated theory of non-additive set functions. We present the theory of such set functions in Chapter 7, including some remarkable and little-known results of Dilworth.

Geometric lattices often arise as lattices of substructures of combinatorial structures. The simplest such are the families of coverings of a set by blocks whose overlap structure is somehow prescribed. The simplest is the lattice of partitions, but there are many others. For example, there is the lattice of contractions of a linear graph, and a related structure, the lattice of cocontractions (see Chapter 6).

In Chapter 11 we give an exposition of one of the outstanding achievements in the theory. We prefer the precise term *orthogonality* to the much-used term *duality* previously used in this connection. The notion of orthogonality is due to Whitney, who was led to it by observing that the classical construction of the dual graph of a planar graph could be generalized to arbitrary graphs by use of the notion of a geometry.

We come now to a discussion of the main parts of the text concerning maps, representation, and the critical problem. In the process, we shall give some indication of what we feel are the prospects for, and purposes of, this theory.

In Chapter 9 we introduce various notions of maps among geometries. Combinatorial geometry can be viewed as a generalization of the classical notion of a vector space; correspondingly we introduce notions of maps which extend the notion of a linear transformation. In the special case of simplicial geometries, the maps we introduce reduce to the classical notion of a simplicial map. Any deeper results will require the full-scale introduction of homological methods, probably the most promising line of future investigation in this area.

The most thoroughly developed chapter of combinatorial geometry is the *representation* (or *coordinatization*) problem, which may be stated as follows. Given a geometry $G(S)$ on a set S, find a module M over an integral domain R, a set S_1 of submodules of M generated by single elements, and a one-to-one map of S onto S_1 which induces an isomorphism of the geometry $G(S)$ onto the geometry defined on the set S_1 in terms of ordinary linear dependence (see Section 3.2). For example, when M is a vector space over a field, the coordinatization problem is simply the problem of representing a geometry as the geometry defined by a finite point set in projective space.

In Chapter 15 we offer a simplified proof of a theorem of W. T. Tutte which reduces the problem of coordinatization to a combinatorial question of matching functions to copoints. The representation theory raises the following problem, which is largely unanswered: For which sets of points in projective space is the associated geometric lattice a complete set of invariants under the projective group? The main results are stated in the following numbered paragraphs; several of these results are not treated in the present work.

A complete representation theory is known for projective geometries of dimension 3 or more. (In stating results with this restriction, we avoid the problems associated with finite projective planes.)

(1) For geometries defined by point sets in projective geometries over the two-element field $GF(2)$, the lattice of subspaces is a complete set of invariants under projective transformations.

(2) A geometric lattice is isomorphic to the lattice of subspaces of a finite point set in a projective geometry over $GF(2)$ if and only if every coline is covered by at most three copoints, or equivalently, if and only if no interval of rank 2 in the lattice contains more than five elements (Whitney, Rado, Tutte; see Chapter 15 on binary lattices).

This theorem appears in Chapter 15, as an immediate consequence of the coordinatization theorem. Such lattices we call *binary* lattices.

(3) For any field other than $GF(2)$, the lattice of subspaces of a point set is not a complete set of invariants.

We come now to the deepest result in this area. Denote by F the *Fano geometry* (or Fano lattice), namely, the geometry and the lattice of subspaces defined by the Fano plane of seven points and seven lines. We denote by F^* the orthogonal geometry (see Chapter 11).

(4) A geometry is representable by a set of points in projective space over *every* field if and only if its geometric lattice is binary and does not have any interval containing a subgeometry isomorphic either to F or to F^*.

Such geometries and lattices are called *unimodular*. They have been variously studied by A. Hoffman, Camion, I. Heller, and Tutte. Their existence was first remarked by Poincaré in connection with homology theory, but was ignored in the subsequent development of topology.

Roughly speaking, the coordinatization of unimodular geometries is achieved by vectors whose entries are $+1$, -1, or 0. Furthermore, the matrix of such vectors enjoys the property that every minor equals $+1$, -1, or 0. Such matrices are called *totally unimodular* and arise in a variety of contexts, for example in integer programing.

When extra requirements are imposed upon the point set by which a geometry is to be coordinatized, the theory is further enriched. The simplest such requirement is that the geometry be representable as the set of *all* points in a projective space. One then obtains that classical theorem on modular geometric lattices, relating synthetic to analytic geometry, which occupied several generations of geometers from von Staudt to von Neumann.

It seems incredible that the development of coordinatization theory should have been entirely bypassed by classical projective geometers. One can only surmise that the possible cause was a pernicious insistence upon the tenets of the Erlanger program. In the light of classical geometry, combinatorial geometry may be considered as a revival of projective geometry in its most synthetic form.

Another fruitful restriction of the coordinatization problem involves the requirement that a geometric lattice be isomorphic to the lattice of contractions of a graph. It is important to see that this is actually a coordinatization problem. To see this, orient the edges of a linear graph, and let V be the vector space over a field F, say, of all 1-coboundaries of the graph. Recall that a 1-coboundary is a function f defined on edges e, whose values in F are of the form

$$f(e) = g(e_h) - g(e_t)$$

where g is a function of the vertices and e_h is the head, e_t the tail, of the edge e. Every edge e of the graph acts as a linear functional on V, by the evaluation

$$f \to f(e).$$

This gives an embedding of the set of edges of the graph into a subset S of the dual space V^*. The geometric lattice defined by the set S by ordinary linear dependence is isomorphic to the lattice of contractions of the graph.

Remarkably enough, the characterization of those geometric lattices which are isomorphic to lattices of contractions of the graph depends upon the two Kuratowski graphs, namely, the complete 5-graph, and the graph of "three girls to the three wells" (the complete bipartite graph). Let K_5 and $K_{3,3}$ be the lattices of contractions of these graphs, and let K_5^* and $K_{3,3}^*$ be the lattices of the orthogonal geometries. Then

(5) A geometric lattice is isomorphic to the lattice of contractions of a linear graph if and only if it is a unimodular lattice, and does not have any interval with a subgeometry isomorphic to K_5^* or $K_{3,3}^*$.

(6) (Kuratowski-Tutte) A geometric lattice is isomorphic to the lattice of contractions of a planar linear graph if and only if it is unimodular and does not have any interval with a subgeometry isomorphic to K_5, $K_{3,3}$, K_5^*, or $K_{3,3}^*$.

The concept of a *simplicial geometry* (see Chapter 6) opens a broad horizon of possibilities for coordinatization and embedding problems. It is an open problem to characterize the k-simplicial geometries, for values of k other than 1 or 2. This was accomplished for $k = 2$, partition lattices, by Ore, Sasaki, Fujiwara, and Sachs, but in ways which ill extend to the infinite case. It is an open problem, also, to characterize the subgeometries of these k-simplicial geometries, as Tutte has done for linear graphs, $k = 2$, by the exclusion of certain subgeometries from all intervals. In its complete generality, this problem reads: When is a geometry isomorphic to the geometry of cycles or cocycles of a simplicial complex? Finally, observing that the minimal dependent subsets of 3-simplicial geometries may be regarded as triangulations of closed surfaces, we may inquire as to which minimal dependent subsets correspond to triangulations of the sphere, the real projective plane, the torus, and surfaces of higher genus.

It is a pleasing consequence of the theorems we have stated that the classical Kuratowski theorem of planar graphs, which has yet to find a suitable topological setting, is placed in the company of several results of the same nature, when viewed within the context of combinatorial geometry. The common structure of the theorems we have stated is the following: "A geometric lattice can be coordinatized by a point set of type X if and only if no subgeometry of any of its intervals is isomorphic to the 'small' geometric lattices $A, B, \cdots, E$." Several other results of this same nature need to be worked out.

There is, however, another far more promising line of investigation which connects with representation theory, and which leads to the classical problem of finding the chromatic number of a graph, namely, the minimum number of colors sufficient to color the vertices of a linear graph, subject to the requirement that no two adjacent vertices be assigned the same color. It is shown by an argument using Möbius inversion that the chromatic number of the graph depends only upon the lattice of contractions of the linear graph. Thus, the classical conjecture of Hadwiger can be stated as follows: "The chromatic number of a linear graph is $\leq n$ if and only if the lattice of contractions of the linear graph does not contain any interval with a subgeometry isomorphic to the lattice of contractions of the complete $(n + 1)$-graph." In this form, the Hadwiger conjecture bears a striking resemblance to the results about coordinatization, and arouses the suspicion that techniques for proving coordinatization results may yield some understanding of the mystery of the coloring problem.

When the coloring problem is stated in geometric terms, the question naturally arises as to whether the same question can be formulated in the context of the present theory. It turns out that this is indeed the case. We now describe this generalization in two stages.

Consider first a geometry $G(S)$ defined on a point set S in a projective space P over $GF(q)$. Say that the set $\{H_1, \cdots, H_k\}$ of hyperplanes *distinguishes* the point set S when for every point t belonging to S there corresponds at least one hyperplane H_i, such that t does not belong to H_i. (That is to say, the supports of the linear functions defining the hyperplanes $H_1, \cdots, H_k$ cover the point set.) The least integer k for which a set of k distinguishing hyperplanes exists is here called the *critical exponent* of the geometry $G(S)$. The critical exponent also equals the maximum dimension of a subspace of P which does not meet S.

The critical exponent depends only upon the geometry. In fact, it can be shown (see Rota, "Foundations I") that the critical exponent k is the smallest positive integer for which the polynomial

$$\sum_{x \in L} \mu(0, x) q^{(n - r(x))k} = p_L(q^k)$$

takes positive value. Here μ is the Möbius function on the lattice L of the geometry, and n is the rank of the geometry.

By taking the point set defined by the edges of a linear graph, considered as linear functionals on the vector space of coboundaries, as described above, it is easily checked that the critical exponent is k if and only if the graph can be colored in q^k colors. It is especially interesting to consider the case $q = 2$.

The polynomial

$$p_L(\lambda) = \sum_{x \in L} \mu(0, x)\lambda^{n-r(x)}$$

is called the *characteristic polynomial* of the geometric lattice (see Rota, "Foundations, I"). An analogue of the critical problem for arbitrary geometric lattices is thus seen to be the problem of locating the integer zeros of the characteristic polynomial.

A large variety of combinatorial problems (for example, coding problems and the Segre independence problems, v. below) can be stated as critical problems. Investigations carried out so far indicate that many attacks used in classical studies of the coloring problem actually carry over to the critical problem generally. We are led to conjecture that the critical problem provides at long last a setting for the study of the coloring problem, a setting which possesses the much-looked-for level of natural generality.

It was an unfortunate historical accident that the map-coloring problem was the first instance of the critical problem to be studied. The systematic study of the critical problem generally should instead begin with simpler instances. Only thus can we ever hope to discover and develop suitable techniques which may eventually lead to a general understanding. The belief that the study of geometries will eventually lead to such an understanding is the motivation of the theory expounded in the present work.

The present work is a continuation of the program initiated by Rota in the paper referred to in the Bibliography as "On the Foundations of Combinatorial Theory, I." Whereas most of the results in that paper were new, the present series is at least partially expository. A brief history of the subject will be given at the end of the present series, together with credits.

Professor Crapo's work was funded by the National Research Council of Canada (grant A–2994). Professor Rota's work was supported by O.N.R. contract 00014–67–A–0204–0016

2. Geometries, matroids, and geometric lattices

Motivation. From classical examples, such as affine and projective spaces, let us abstract an axiomatic description of a combinatorial geometry.

Classically, the objects studied are points, lines, planes, etc. All these objects (or "flats") possess a well-defined rank (dimension). Points have rank 1, lines have rank 2, etc. Any set A of points in a geometry should thus possess a rank, $r(A)$, equal to the rank of the smallest flat (subspace) which contains A.

Our objective will be to abstract these and other familiar properties of classical projective geometry by placing certain conditions on the function $A \to \bar{A}$, which associates, with each subset A of the points, that smallest flat which contains A.

Consider, for instance, the geometry of Euclidean n-dimensional space E^n. For any set A of points in E^n, $\bar{A}$ is the smallest subspace (point, line, plane, etc., not necessarily containing the origin) containing the point set A. Within any subset $A \subseteq E^n$ we can always find a subset A_f with at most $n + 1$ elements, such that A_f and A span the same subspace.

Also, given any k-dimensional subspace A, the $(k + 1)$-dimensional subspaces B_α which contain A partition the points not in A. That is to say, if a point a is not in the subspace A, then $\overline{A \cup \{a\}}$ is a $(k + 1)$-dimensional subspace B. If b is any other point in B but not in A, then b equally well generates the subspace B over A. Thus $a \in \overline{A \cup \{b\}}$. These properties, of *finite basis* and *exchange*, lead us to the definition of a combinatorial geometry.

One critical feature of projective geometries will be found lacking in combinatorial geometries. In projective geometries the rank (dimension function) satisfies the identity

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

For combinatorial geometries, this becomes an *inequality*

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

which is often strict. After this relaxation to an inequality, we discover that an extraordinary variety of combinatorial structures are combinatorial geometries. While some of these structures are subsets of projective geometries, others arise from totally unrelated situations. This is our main motivation for the following axioms.

Basic Definitions. A *closure relation* on a set S is a function $A \to \bar{A}$ defined for all subsets $A \subseteq S$, satisfying

$$\bar{A} \subseteq S \tag{2.0}$$

$$A \subseteq \bar{A} \tag{2.1}$$

$$A \subseteq \bar{B} \text{ implies } \bar{A} \subseteq \bar{B} \tag{2.2}$$

for all subsets A, B of S. It follows that a closure relation is order-preserving,

$$A \subseteq B \text{ implies } \bar{A} \subseteq \bar{B}, \text{ and } \overline{A \cap B} = \bar{A} \cap \bar{B}$$

and is idempotent:

$$\bar{A} = \bar{\bar{A}}.$$

A set endowed with a closure relation is a *closure space*. A subset $A \subseteq S$ of a closure space S is *closed* if and only if $A = \bar{A}$.

We shall concern ourselves with closure relations with the *exchange property*:

For any elements $a, b \in S$, and for any subset $A \subseteq S$,

$$\text{if } a \in \overline{A \cup b} \text{ and } a \notin \bar{A} \text{ then } b \in \overline{A \cup a}. \tag{2.3}$$

(The notation $\overline{A \cup b}$ is an abbreviation for the correct, but needlessly pedantic, $\overline{A \cup \{b\}}$.)

A closure relation on a set S has *finite basis* if and only if

$$\text{any subset } A \subseteq S \text{ has a finite subset } A_f \subseteq A \text{ such that } \overline{A_f} = \overline{A}. \tag{2.4}$$

A *pregeometry* or *matroid* $G(S)$ is any closure space consisting of a set S and a closure relation with finite basis and the exchange property.

A pregeometry $G(S)$ is a *geometry* if and only if for the null set $\varnothing$

$$\overline{\varnothing} = \varnothing, \quad \text{and} \quad \bar{a} = a \quad \text{for all elements} \quad a \in S. \tag{2.5}$$

Let us associate a geometry with each pregeometry. In doing so, we make use of the partition of the complement of the closed set $\overline{\varnothing}$, as provided by the exchange property.

For any pregeometry $G(S)$ on a set S, the relation $(\sim)$ defined on S by

$$a \sim b \quad \text{if and only if} \quad \bar{a} = \bar{b}$$

is an equivalence relation. The elements of S which are not in the closure $\overline{\varnothing}$ of the empty set are partitioned by the equivalence relation into classes which may be regarded as the elements of a set S_0. The closure relation is well defined on these equivalence classes, and therefore canonically determines a closure relation on the set S_0. The set S_0, furnished with this closure relation, is a geometry, the geometry *canonically associated with the pregeometry* $G(S)$.

Whenever possible, we shall try to express all results in terms of geometries rather than matroids. In matroids, the possibility that two points may be dependent upon each other results in considerable nuisance, which is avoided by passing to the associated geometry. All major results of the theory are about geometries rather than matroids, and the only reason for considering matroids at all is that they actually arise in applications and in certain constructions.

Whenever possible, we shall try to picture finite geometries as geometries of points in affine space E^n of some dimension n. To draw such a geometry, we mark in only the nontrivial lines and planes. For instance, in the plane geometry on six points shown in Figure 2.1, the set $\underline{bd}$ is a line, but it is not drawn in. There are seven lines in all.

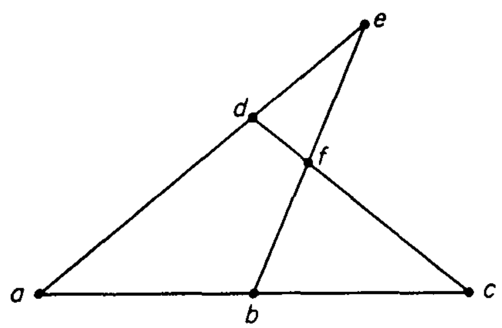


FIGURE 2.1

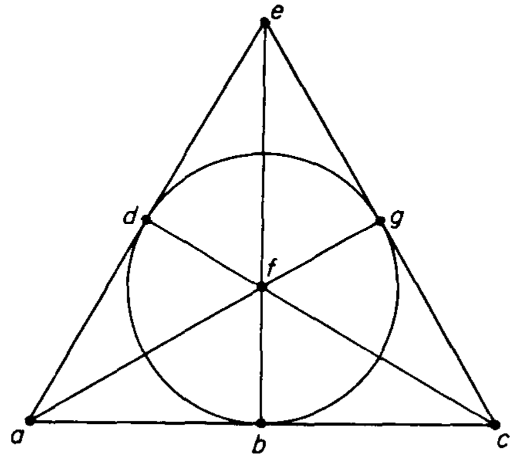


FIGURE 2.2

In certain marginal cases, where a geometry cannot be realized in affine space, a drawing can still be helpful. Consider the projective seven-point plane, for example. The seventh line $\underline{bdg}$ is indicated by a curve in the drawing.

In any closure space, the intersection of any collection of closed subsets is also closed. The closed subsets (or *flats*) of a geometry $G(S)$ form a *complete lattice* $L(S)$. (The reader will not be required to know any notions of lattice theory beyond the basic definitions.) The order relation on $L(S)$ is induced by that of containment, relating subsets of S. The binary operations $\vee$ and $\wedge$ in the lattice $L(S)$ are as follows: if A and B are flats of $G(S)$, then $A \vee B$ is the smallest flat containing both A and B,

$$A \vee B = \overline{A \cup B},$$

and $A \wedge B$ is the largest flat contained in both A and B,

$$A \wedge B = A \cap B.$$

The term *open set* refers as usual to the complement of a closed set. Clearly, the union of any collection of open sets is also open.

An example of a geometry on five points, together with its lattice of flats, is given in Figure 2.3. We shall generally use the lower-case letters

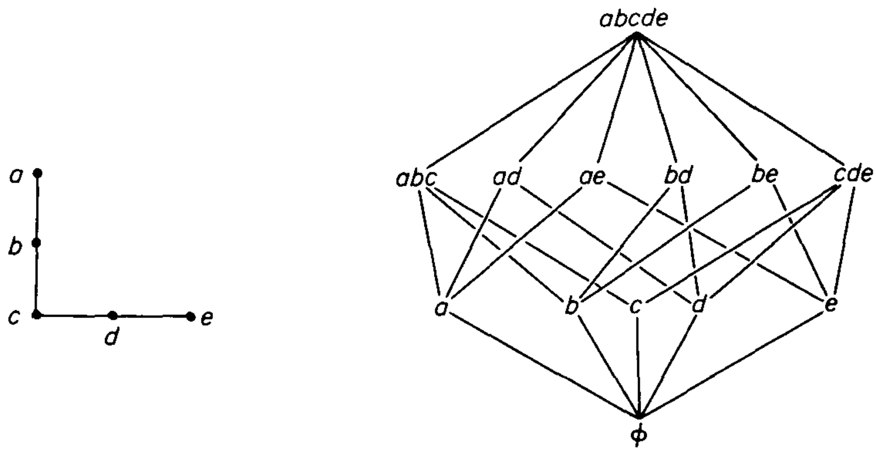


FIGURE 2.3

$z, y, x, \cdots$ to indicate flats of a geometry $G(S)$, when these flats are regarded sometimes as subsets of S and sometimes as elements of the lattice $L(S)$. For those flats which are points, we shall use the letters $a, b, \cdots$.

We now derive the characteristic properties of lattices $L(S)$ of geometries $G(S)$. Two terms are needed for this description. A *chain* in a lattice L is any linearly ordered subset of L. An element y *covers* an element x in a lattice L if and only if $x < y$, but $x < z < y$ for no element z in L.

We write $y \downarrow x$ if $y = x$ or y covers x (Read: y covers or equals x).

Lattices of geometries have two characteristic properties. They have no infinite chains, and the flats covering any flat x partition the points not in x. From these properties it will follow that every flat x has a well-defined finite rank, $r(x)$, equal to the common length of all maximal chains from 0 to x, where $0 = \bar{\varnothing}$ is the smallest flat.

A flat of rank 1 is a *point*, one of rank 2 is a *line*, etc. The entire set S is closed, and has some finite rank, say n. This is the unit element 1 of the lattice. A flat of rank $n - 1$ is a *copoint*, one of rank $n - 2$ is a *coline*, etc.

PROPOSITION 2.1. A flat y covers a flat x in the lattice L of a geometry $G(S)$ if and only if there is a point $\underline{a}$ not in x such that $x \vee a = y$.

Proof: If a flat y covers a flat x, consider the flats as subsets of the set S, and choose an element $a \in y - x$. Then $\underline{a}$ is a flat, and $x < x \vee a \leq y$, so $x \vee a = y$. Conversely, if x is a flat in L and $\underline{a}$ is a point not in x, let z be a flat such that $x < z \leq x \vee a$. Choose an element b in the difference set $z - x$. Then $b \notin \bar{x} = x$ and $b \in \overline{x \cup a} = x \vee a$, so the exchange property implies $a \in \overline{x \cup b} \leq \bar{z} = z$. But $x \leq z$ and $a \leq z$ imply $x \vee a \leq z$, that is: $x \vee a = z$, and $x \vee a$ covers x. ∎

PROPOSITION 2.2. In the lattice $L(G)$ of any geometry $G(S)$, $y \downarrow x$ implies $(y \vee z) \downarrow (x \vee z)$, for all flats x, y, z.

Proof: By Proposition 2.1, we may write $y = x \vee a$ for some point a. Then $y \vee z = (x \vee a) \vee z = (x \vee z) \vee a$, which must cover or equal $x \vee z$, again by Proposition 2.1. ∎

PROPOSITION 2.3. In the lattice of flats of a geometry, any chain is finite.

Proof: If there is an infinite chain of flats, there is either an infinite increasing subsequence or an infinite decreasing subsequence. From Axiom 2.4 we conclude that for any increasing sequence $\{A_i\}$ of closed subsets of S, the union $\bigcup A_i$ contains a finite set A such that $\bar{A} = \overline{\bigcup A_i}$. But for some index i, $A \subseteq A_i$, so $\overline{\bigcup A_i} = \bar{A} \subseteq \bar{A}_i = A_i$, and the sequence must terminate in the closed set A_i.

It remains to show that there can be no infinite decreasing sequence of flats. If $\{A_i\}$ is an infinite decreasing sequence of closed subsets of S, select an arbitrary infinite sequence $\{a_i\}$ of elements of the difference sets, $a_i \in A_{i-1} - A_i$, $i = 1, 2, \cdots$. By the ascending chain condition, $A_1 = \overline{\{a_1, a_2, \cdots, a_r\}}$ for some r. It follows that $a_{r+1} \in \overline{\{a_1, a_2, \cdots, a_r\}}$. Let i be the least suffix such that $a_{r+1} \notin \overline{\{a_i, a_{i+1}, \cdots, a_r\}}$, then $a_{r+1} \in \overline{\{a_{i-1}, \cdots, a_r\}}$. Hence, by the Exchange Property, $a_{i-1} \in \overline{\{a_i, a_{i+1}, \cdots, a_{r+1}\}} \subseteq A_i$, which is a contradiction. ∎

A lattice having no infinite chains is said to be *semimodular* whenever it has the *Birkhoff covering property*:

$$\text{if } x \text{ covers } x \wedge y, \text{ then } x \vee y \text{ covers } y.$$

This property is easily seen to be equivalent, in any lattice having no infinite chains, to the following property, as originally stated by Birkhoff:

$$\text{if } x \text{ and } y \text{ cover } x \wedge y, \text{ then } x \vee y \text{ covers } x \text{ and } y.$$

We leave to the reader the verification of the equivalence of these two properties, and we shall use them interchangeably throughout. We shall also make free use of the fact that every semimodular lattice satisfies the conclusion of Proposition 2.2.

We know, by Propositions 2.2 and 2.3, that the lattice of a geometry is semimodular, with no infinite chains. Furthermore, every flat of a geometry equals the lattice supremum of its points. A lattice with this latter property is said to be a *point lattice* (that is, every element is expressible as a supremum of elements covering the smallest element, denoted as usual by 0).

A lattice is *geometric* if and only if it is a semimodular point lattice with no infinite chains. We shall see that this description characterizes lattices of flats of geometries.

PROPOSITION 2.4. If L is a geometric lattice, then the equation

$$\bar{A} = \{a \in S;\, a \leq \sup A\}, \tag{2.6}$$

defined for all subsets A of the set S of all points of L, defines a geometry $G(S)$ on S. The lattice of flats of $G(S)$ is isomorphic to L.

Proof: If $a \in A$, for any set A of points, $A \subseteq S$, then $a \leq \sup A$. Thus $A \subseteq \bar{A}$. If $A \subseteq \bar{B}$, for two subsets $A, B \subseteq S$, then $a \in A$ implies $a \leq \sup B$, so $\sup A \leq \sup B$, and $\bar{A} \subseteq \bar{B}$. Thus Equation 2.6 defines a closure relation.

It is clear that a set of points is closed if and only if it is the set of all points beneath some lattice element. Since L is a point lattice, there is a one-to-one order-preserving correspondence between closed subsets of S and elements of L. If the closure relation is found to be a geometry, then it will follow that its lattice is isomorphic to L.

For any set A of points, $A \subseteq S$, let $x = \sup A$ in L. Well-order the set A, and let $C_0 = \varnothing$. Having constructed C_{i-1}, let a_i be the first element in $A - C_{i-1}$, under the well-ordering, and let $C_i = \overline{C_{i-1} \cup a_i}$. The sequence $\{C_i\}$ must be finite, because the lattice has no infinite chain, so $\{a_i\}$ is a finite subset of A with closure equal to that of A. Thus the closure relation has finite basis.

If $a \notin \bar{A}$, but $a \in \overline{A \cup b}$, for points a, b and some subset $A \subseteq S$, let $x = \sup A$, so $a \nleq x$, $a \leq x \vee b$. But $x \vee b$ covers x, because the lattice is semimodular. Therefore $x < x \vee a \leq x \vee b$ implies $x \vee a = x \vee b$. Consequently, $b \leq x \vee a$, and $b \in \overline{A \cup a}$, proving the exchange property. ∎

Rank, complementation, and modular pairs.

We define the *rank* $\lambda(x)$ of a flat x as the common length of all maximal chains from 0 to x in the lattice $L(S)$. Once we have verified that this concept of rank is well defined, we are justified in our use of geometric language. It becomes reasonable to speak of lines (rank 2), planes (rank 3), copoints (rank $n - 1$), and colines (rank $n - 2$), where $\lambda(1) = n$ is the rank of the entire geometry. (Recall that in a lattice L, 1 stands for the maximal element).

PROPOSITION 2.5. Given a semimodular lattice L having no infinite chains, and elements x, y in L such that $x \leq y$, then all maximal chains from x to y have the same length.

Proof: Let $x = s_0 < \cdots < s_n = y$ and $x = t_0 < \cdots < t_m = y$ be two maximal chains from x to y. If $n = 0$ or $n = 1$, then the chains coincide. Assume the truth of the theorem for all pairs x', y' between which there exists a maximal chain of length less than n. By the Birkhoff covering property, $s_1 \vee t_1$ covers both s_1 and t_1. Select a maximal chain $s_1 \vee t_1 = u_2 < u_3 < \cdots < u_p = y$ from $s_1 \vee t_1$ to y. Comparing the two paths from s_1 to y, we have $p = n$ by the induction hypothesis. Thus there is a maximal chain from t_1 to y having length $n - 1$, and $m = p = n$, again by the induction hypothesis. ∎

For each subset A of a geometry $G(S)$, let $\sigma(A)$ equal A, considered as an element of the lattice $L(G)$. Thus σ is a function (the canonical map, see Chapter 9) from the Boolean algebra $\mathscr{B}(S)$ of all subsets of S into $L(G)$, and we may say that the *geometric rank* $r(A)$ of a subset A is the rank $\lambda(\sigma(A))$. These ranks, the rank λ of lattice elements and the geometric rank r of subsets of S, satisfy a certain characteristic linear inequality.

PROPOSITION 2.6. In any semimodular lattice L having no infinite chains, (in particular in every geometric lattice) the rank function λ satisfies the inequality

$$\lambda(x \wedge y) + \lambda(x \vee y) \leq \lambda(x) + \lambda(y), \tag{2.7}$$

for all elements x, y in L. If $G(S)$ is a geometry, the geometric rank r satisfies the analogous inequality

$$r(A \cap B) + r(A \cup B) \leq r(A) + r(B) \tag{2.8}$$

for all subsets A, B of S.

Proof: Assuming for a moment that the inequality (2.7) holds for geometric lattices, let $\sigma(A) = x$ and $\sigma(B) = y$ for two subsets A, B for a geometry $G(S)$. Then $\overline{A \cap B} \subseteq \bar{A} \cap \bar{B}$, the latter being a closed set, the image of which in $L(G)$ is equal to $x \wedge y$. Thus $r(A \cap B) \leq \lambda(x \wedge y)$. Also $\overline{A \cup B} = \overline{\bar{A} \cup \bar{B}}$, the latter being a closed set with image $x \vee y$ in $L(G)$. Thus $r(A \cup B) = \lambda(x \vee y)$ and $r(A \cap B) + r(A \cup B) \leq \lambda(x \wedge y) + \lambda(x \vee y) \leq \lambda(x) + \lambda(y) = r(A) + r(B)$, so the geometric rank satisfies Inequality 2.8.

Assume now that L is a semimodular lattice with no infinite chains, and that x, y are elements of L. Choose a maximal chain $x \wedge y = x_0 < x_1 < \cdots < x_n = x$ from $x \wedge y$ to x. Let $y_i = x_i \vee y$. Then $y_i \downarrow y_{i-1}$ for $i = 1, \cdots, n$, by the conclusion of Proposition 2.2 which, as remarked, holds in any semimodular lattice. Thus except for possible repetition of some elements, $y_0 \le y_1 \le \cdots \le y_n$ is a maximal chain from y to $x \vee y$, and $\lambda(x) - \lambda(x \wedge y) \ge \lambda(x \vee y) - \lambda(y)$. ∎

Given a closed set A in a geometry $G(S)$, we may inquire whether there exists a closed set B, having no points in common with A, and such that $\overline{A \cup B} = S$. Such flats always exist, and are called "*complements*" of A. In the lattice $L(G)$, a flat y is a *complement* of a flat x if and only if $x \wedge y = 0$, and $x \vee y = 1$. Actually, we shall see that complements exist even within an *interval* $[s, t] = \{z; s \le z \le t\}$ of a geometric lattice:

PROPOSITION 2.7. If L is a geometric lattice, and $[s, t]$ is an interval in L, then for any element x in the interval, there exists an element y such that $x \wedge y = s$ and $x \vee y = t$.

Proof: See Figure 2.4. Given any lattice element y_i in the interval $[s, t]$ such that $x \wedge y_i = s$ and $x \vee y_i < t$, we may select a point b in the difference set $t - (x \vee y_i)$, that is, a point b such that $b \le t$ but $b \not\le x \vee y_i$, and set $y_{i+1} = y_i \vee b$. Then $x \vee y_{i+1} \le t$ and covers $x \vee y_i$. If $x \wedge y_{i+1} \ne s$, we may select a point $c \le x \wedge y_{i+1}$, $c \not\le s$. Then $c \le x$, so $c \not\le y_i$, and $y_i < c \vee y_i \le c \vee y_{i+1} = y_{i+1}$, that is: $c \vee y_i = y_{i+1}$. But then $x \vee y_i = (c \vee x) \vee y_i = x \vee (c \vee y_i) = x \vee y_{i+1}$, a contradiction. Consequently, y_{i+1} also satisfies $x \wedge y_{i+1} = s$, and the construction may be repeated until $x \vee y_j = t$, after finitely many steps, and y_j is a complement of x in the interval $[s, t]$. ∎

A flat y such that $x \wedge y = s$, $x \vee y = t$, is called a *relative complement* of x in the interval $[s, t]$. Clearly, any element y is a relative complement

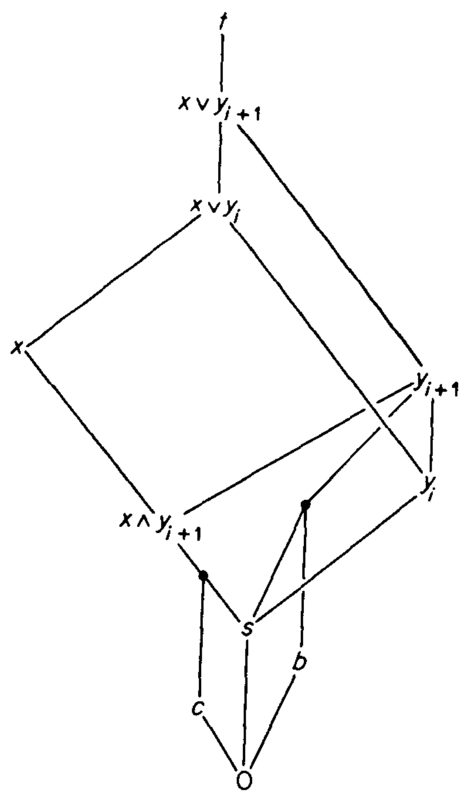


FIGURE 2.4

of x in one interval, the interval $[x \wedge y, x \vee y]$. Because we are dealing with lattices without infinite chains, any flat x also has at least one *minimal* relative complement, in any interval containing x. Minimal relative complements are the analogues of "orthogonal complements" in geometric constructions, and are thus of prime importance. They are characterized as follows. A lattice is *relatively complemented* whenever a relative complement exists of any element, in any interval containing it.

PROPOSITION 2.8. Consider the following statements concerning a pair x, y of elements of a lattice L:

(a) For all elements $z \in L$ such that $z \leq x$, $(x \wedge y) \vee z = x \wedge (y \vee z)$.
(b) x is a minimal relative complement of y in the interval $[x \wedge y, x \vee y]$.
(c) $\lambda(x \wedge y) + \lambda(x \vee y) = \lambda(x) + \lambda(y)$.

Statements (a) and (b) are equivalent in any relatively complemented lattice. Statements (a) and (c) are equivalent in any semimodular lattice with no infinite chains. Thus all three statements are equivalent in a geometric lattice.

Proof: We prove first that (b) implies (a) in any relatively complemented lattice. Given elements x, y, z in any lattice, with $z \leq x$, we have $y \leq y \vee z$, so $x \wedge y \leq x \wedge (y \vee z)$, and $z \leq y \vee z$, so $z \leq x \wedge (y \vee z)$. Thus the inequality, $(x \wedge y) \vee z \leq x \wedge (y \vee z)$, holds in any lattice.

Assume now that the lattice L is relatively complemented. Let the element $t < x$ be a relative complement of the element $x \wedge (y \vee z)$ in the interval $[(x \wedge y) \vee z, x]$, assuming that $(x \wedge y) \vee z < x \wedge (y \vee z)$. Such an element t (see Figure 2.5) satisfies $t \vee y \geq y \vee z \geq x \wedge (y \vee z)$, but $t \vee y \geq x \wedge (y \vee z)$ and $t \vee y \geq t$ together imply that $t \vee y \geq x$, because t is a relative complement of $x \wedge (y \vee z)$ in the interval $[(x \wedge y) \vee z, x]$. Thus $t \vee y \geq x \vee y$, and equality $t \vee y = x \vee y$ follows from the obvious inequality $t \vee y \leq x \vee y$. On the other hand, $x \wedge y \leq t \leq x$ implies $t \wedge y = x \wedge y$. Since $t < x$ and t is a relative complement of y in the interval $[x \wedge y, x \vee y]$, it follows that $t = x$, because x is a *minimal* relative complement. Hence condition (b) implies condition (a). The converse is clear, if z is taken to be a relative complement of y in the interval $[x \wedge y, x \vee y]$, with $z < x$.

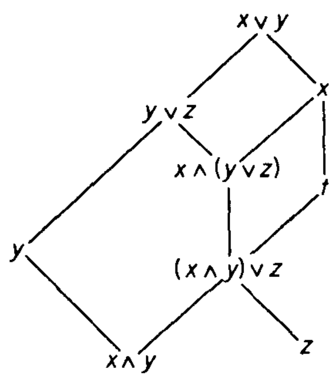


FIGURE 2.5

Assume next that the lattice L is semimodular, with no infinite chains. Choose a maximal chain $x \wedge y = x_0 < x_1 < \cdots < x_n = x$ from $x \wedge y$ to x. Then $\{y_i = x_i \vee y; i = 0, \cdots, n\}$ is, except for possible repetition of elements, a maximal chain from y to $x \vee y$. $\lambda(x \vee y) - \lambda(y) < \lambda(x) - \lambda(x \wedge y)$ if and only if there exists an index i such that $x_{i-1} \vee y = x_i \vee y$. If such an index exists, then $(x \wedge y) \vee x_{i-1} = x_{i-1} < x_i \leq x \wedge (y \vee x_i) = x \wedge (y \vee x_{i-1})$, and statement (a) is not satisfied for $z = x_{i-1}$. Conversely, if statement (a) is not satisfied for some element $z \leq x$, the elements $s = (x \wedge y) \vee z$ and $t = x \wedge (y \vee z)$ stand in the relation $s < t$. Choose a maximal chain $x \wedge y = x_0 < \cdots < x_i = s < \cdots < x_j = t < \cdots < x_n = x$. Then $x_i \vee y = x_j \vee y$, so $\lambda(x \vee y) - \lambda(y) < \lambda(x) - \lambda(x \wedge y)$, and statement (c) is not satisfied. ❙

If two flats x, y in a geometry satisfy the conditions of Proposition 2.8, we say they are a *modular pair*, and write $(x, y)M$. Two flats x, y are *skew* if $(x, y)M$ and $x \wedge y = 0$. If a flat x bears the relation $(x, y)M$ to every flat y of the geometry, then the flat x is a *modular* flat. The points, the zero 0, and the unit 1 are modular, in any geometric lattice. If all pairs of flats are modular, the geometry is *modular*.

PROPOSITION 2.9. For any two elements x, y in a geometric lattice, there is a natural order-preserving one-one function from the interval $[x \wedge y, x]$ into the interval $[y, x \vee y]$ if and only if x and y are a modular pair, given by the function $z \mapsto z \vee y$.

Proof: If x and y are not a modular pair, then $\lambda(x \vee y) - \lambda(y) < \lambda(x) - \lambda(x \wedge y)$, and no maximal chain in $[x \wedge y, x]$ can be sent, by an order-preserving one-one function, into the shorter interval $[y, x \vee y]$. Assume now that $(x, y)M$. Consider the functions

$$[x \wedge y, x] \underset{g}{\overset{f}{\rightleftarrows}} [y, x \vee y]$$

defined by $f(z) = y \vee z$ and $g(w) = x \wedge w$. Both functions are order-preserving. The composite $g(f)$ is the identity function on the interval $[x \wedge y, x]$, because $(x, y)M$ and

$$g(f(z)) = x \wedge (y \vee z) = (x \wedge y) \vee z) = z$$

for all elements z in the interval $[x \wedge y, x]$. Thus the function f is the required one-one function. ❙

As a consequence of the existence of relative complements in geometries, every flat is expressible as a set intersection of maximal proper flats:

PROPOSITION 2.10. In any geometric lattice L, every element $x \in L$ is expressible as an infimum of copoints of L.

Proof: Let x be an element of L, and let $y = \inf C$, where C is the set of copoints containing x. Assume $x < y$. Let z be a relative complement of y in the interval $[x, 1]$. Then $z \neq 1$, and hence there exists a copoint c containing z. Now, $y \vee z = 1$ implies $y \vee c = 1$, so $y \not\leq c$. But $x \leq c$, contradicting the definition of y. ❙

3. Some classical examples

The most important classes of geometries, historically speaking, are the following: chain groups, function spaces, algebraic extensions of fields, coverings, Wille incidence geometries, graphs and simplicial geometries. As a preliminary to the discussion of these examples, let us look at a few small but otherwise perfectly general geometries. In the next Section, we shall discuss some examples arising in applied mathematics.

3.1. SMALL GEOMETRIES

Given a five-element set, how many different geometries may be constructed upon it? It will turn out that all such geometries can be represented by sets of points in an affine space of dimension no greater than 4 when the closure of a set is the ordinary linear span, so we shall draw them as such. First, there is the unique geometry of rank 2: five points on a line.

FIGURE 3.1

A rank 3 geometry on five points may have at most one line of four points, and at most two lines of three points. There are four possibilities:

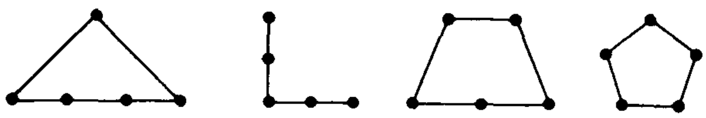


FIGURE 3.2

A rank 4 geometry on five points has at most one nontrivial flat, either a three-point line, a four-point plane, or five points in general position on a flat of rank 4.

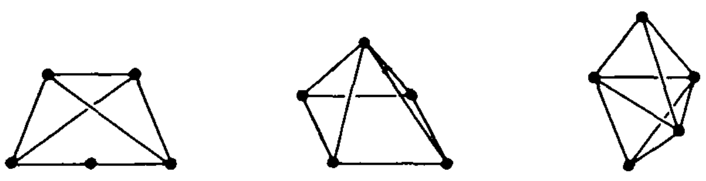


FIGURE 3.3

If the geometry on five points has rank 5, the points must be in general position in a space of rank 5.

A similar count of the number $g_{n,k}$ of essentially different geometries of rank k on an n-element set, $n = 1, \cdots, 8$, yields the following tabulation. Let g_n be the total $g_{n,1} + \cdots + g_{n,n}$. Then the recursion

$$g_{n+1} = (g_n)^{3/2}$$

seems approximately correct, on the basis of this data alone. This would suggest that there are some thirty thousand essentially different geometries on a nine-element set.

TABLE 1. TABULATION OF $g_{n,k}$

	8 pts.	7 pts.	6 pts.	5 pts.	4 pts.	3 pts.	2 pts.	1 pt.
rank 1	·	·	·	·	·	·	·	1
2	1	1	1	1	1	1	1	
3	68	23	9	4	2	1		
4	617	49	11	3	1			
5	217	22	4	1				
6	40	5	1					
7	6	1						
8	1							
total	950	101	26	9	4	2	1	1

3.2. CHAIN GROUPS

Let M be a module over a commutative integral domain R. The set, upon which we shall define a closure relation, will be an arbitrary finite subset S of the module M.

For any subset $A \subseteq S$, let $\bar{A}$ be the set

$\bar{A} = \{x \in S$; some nonzero R-multiple of x is expressible as a finite linear combination of elements of A, with coefficients in $R\}$.

PROPOSITION 3.1. The subset S, furnished with the relation $A \to \bar{A}$, is a pregeometry.

Proof: For any subset $A \subseteq S$, $A \subseteq \bar{A}$. Assume $A \subseteq \bar{B}$ and $x \in \bar{A}$. Then for some nonzero $r \in R$, and coefficients $r_i \in R$, $rx = r_1m_1 + \cdots + r_km_k$, with $m_i \in A$. Since the elements m_i are in $\bar{B}$, there exist nonzero elements $s_1, \cdots, s_k$ in R such that s_im_i is a finite linear combination of elements of B. Since R is an integral domain, the product $s_1 \cdots s_nr$ is nonzero. Thus $s_1 \cdots s_nrx$ is a nonzero multiple of x, which is expressible as a finite linear combination of elements of B, because R is commutative. So $x \in \bar{B}$, $\bar{A} \subseteq \bar{B}$, and the function $A \to \bar{A}$ is a closure.

Say that $y \in \overline{A \cup x}$, $y \notin \bar{A}$. Then for some $r \neq 0$,

$$ry = sx + \sum_{i=1}^{n} r_iz_i, \qquad z_i \in A.$$

But $y \notin \bar{A}$ implies $s \neq 0$, so

$$sx = ry + \sum_{i=1}^{n} (-r_i)z_i, \qquad z_i \in A$$

and $x \in \overline{A \cup y}$. ∎

This pregeometry is called the *chain- group* pregeometry $C(S)$. It is by far the most important example of a geometry. The reader is urged to peruse the connection between the exchange property and the "elimination" of a variable in linear algebra that is displayed by the preceding example. This connection embodies the "yoga" of the theory of combinatorial geometry.

The traditional example of a chain-group geometry is that on a set S of vectors in a vector space over a field. Then the dependence of a vector x on a subset $A \subseteq S$ reduces to linear dependence, because any relation $rx = r_1a_1 + \cdots + r_na_n$ may be divided throughout by the nonzero field element r.

An especially simple example of chain-group geometry occurs when an Abelian group G is regarded as a module over the integers. The closed subsets of G are simply the subgroups H of G for which the factor group G/H has no elements of finite order.

We shall discuss below several examples of chain groups which have arisen in combinatorial problems.

The *rank* of a subset of a chain group is the analogue of the *dimension* of a subspace of a projective space. However, the most important examples of chain groups in the present context are those where $M = P_n$ is a projective space over a fixed field F, so that we set $R = F$, and let S be a subset of P_n. In this case the lattice of flats of the chain group $C(S)$ is isomorphic to the lattice consisting of those subspaces of P_n which are spanned by subsets of S. In these examples we see combinatorial geometry as the study of those properties of points in projective space which depend only on join and intersection. This is the typical example of combinatorial geometry, and provides the motivation and a visualization for many notions.

We urge the reader to constantly picture, insofar as possible, all notions and results of combinatorial geometry in terms of sets of points in projective space.

3.3. FUNCTION SPACES

A *subtractive algebra* is a set X, together with a binary operation $(x, y) \rightarrow x - y$ (read "x minus y") and a constant 0 (read "zero") satisfying the following axioms:

$$(1)\ x - x = 0 \quad \text{for all} \quad x \in X$$
$$(2)\ \text{if } x - 0 = 0, \quad \text{then} \quad x = 0.$$

Every loop (and, in particular, every group) gives rise to a subtractive algebra, where 0 is the loop identity and $x - y = xy^{-1}$.

Now let S be a finite set, and X a subtractive algebra. A *proper function space* on S shall be a set V of functions from S to X with the following properties:

(1) if $f, g \in V$ and $r = f - g$, then $r \in V$.

(2) if, for some subset $A \subseteq S$, element $s \in S$, and function $g_0 \in V$, $g_0(p) = 0$ for all $p \in A$ and $g_0(s) \neq 0$, then, for any element x in X such that $x = f(s)$ for some $f \in V$, there exists a function $g \in V$ such that $g(p) = 0$ for all $p \in A$, and $g(s) = x$.

The typical example of a proper function space is a vector space of functions from a set S to a field F, where subtraction in F is the operation in the subtractive algebra. Proper function spaces furnish the most general context in which the following hull-kernel construction gives rise to a geometry.

Let V be a proper function space on a set S over a subtractive algebra X. For any subset $A \subseteq S$, the *hull* of A is a subset of V,

$$h(A) = \{f \in V; f(s) = 0, \forall s \in A\}.$$

For any subset $B \subseteq V$, the *kernel* of B is a subset of S,

$$k(B) = \{s \in S; f(s) = 0, \forall f \in B\}.$$

Since the functions h and k are order-inverting between the Boolean algebras $\mathscr{B}(S)$ and $\mathscr{B}(V)$ of all subsets of S and V,

$$\mathscr{B}(S) \underset{k}{\overset{h}{\rightleftarrows}} \mathscr{B}(V)$$

and since their composites $k(h)$ and $h(k)$ are increasing on $\mathscr{B}(S)$ and $\mathscr{B}(V)$, the functions h and k form a *Galois connection* between the Boolean algebras.

It is an easy consequence of the general theory of Galois connections (see for example Rota, Foundations 1) that the composite functions are closure operators. The images of functions h and k are precisely the closed subsets of V and S, respectively. The functions h and k take unions into intersections. Finally, the lattices of closed subsets of S and of V are anti-isomorphic (use the restriction of h). The reader may easily prove these facts for himself, beginning with a proof of the equivalence

$$A \subseteq k(B) \quad \text{if and only if} \quad B \subseteq h(A).$$

PROPOSITION 3.2. Let V be a proper function space on a set S. The set S, with closure $\bar{A} = k(h(A))$, is a pregeometry or matroid.

Proof: We prove that the closure has the exchange property. Assume $s \notin k(h(A))$ for some subset $A \subseteq S$, but $s \in k(h(A \cup t))$ for some element $t \in S$. Find a function $f \in V$ vanishing on A (that is, $f(p) = 0$ for $p \in A$) but not on s. It does not vanish on t. If $t \notin k(h(A \cup s))$, then there is a function g vanishing on A and on s, but not on t. By Property (2) of proper function spaces, we may assume without loss of generality that $f(t) = g(t)$. Let $r = f - g$. Then $r \in V$, and by Axioms (1) and (2) for the subtractive algebra X,

$$\begin{aligned} r(p) &= f(p) - g(p) = 0 - 0 = 0 \quad \text{for all } p \in A, \\ r(s) &= f(s) - g(s) = f(s) - 0 \neq 0, \\ r(t) &= f(t) - g(t) = 0. \end{aligned}$$

This contradicts the assumption that $s \in k(h(A \cup t))$, and completes the proof. ∎

We say that the proper function space V *distinguishes the points* of S if, for every ordered pair s, t of distinct elements of S, there is a function $f \in V$ such that $f(s) = 0$ but $f(t) \neq 0$. Whenever V distinguishes the points of S, the pregeometry obtained above is a geometry. In any event, let us denote the geometry arising from this hull-kernel construction by the symbol $F(S)$, the function-space geometry on the set S.

Let M be a module over a commutative integral domain R. (For instance M may be an Abelian group. Then R is the integers.) If $C(S)$ is the chain-group pregeometry on a finite subset S of M, the set S also has a natural function-space pregeometry $F(S)$, obtained by taking as the proper function space V the set of restrictions to S of all elements of the dual module M^* (the set of R-linear maps of M into R). It is easy to check that V is in fact a proper function space on S, and that the pregeometries $F(S)$ and $C(S)$ are the same. Thus chain-group pregeometries are in reality merely a special case of function space pregeometries, in which the subtractive algebra is necessarily a commutative integral domain.

Conversely, given an R-function space pregeometry $F(S)$ on a finite set S, derived from a proper function space V of functions from S into a commutative integral domain R, V is an R-module and has a dual $V^* = \mathrm{Hom}_R(V, R)$. Each element of S acts as a linear functional on V, so it can be regarded as a member of V^*. In case $F(S)$ is a geometry, each $s \in S$ corresponds to a distinct functional so S becomes a subset of V^*, and as V^* is an R-module, S acquires therefrom a chain-group structure $C(S)$, which is the same as $F(S)$.

We can now state that *any integral-domain function-space geometry* (not so for a pregeometry) *is isomorphic to a chain-group geometry, and vice versa.* Although strictly speaking the isomorphism makes one of the two examples superfluous, it is nevertheless useful to be able to use both the language of chain groups and that of proper function spaces.

As an instance of the usefulness of having both modes of expression available, we will prove that *any chain-group geometry is a vector-space geometry.*

The precise meaning of the statement is this: if S is a finite subset of an R-module M and $C(S)$ is the resulting chain-group geometry on S, then S can be embedded in a vector space Y over some field F so as to yield a chain-group geometry $C'(S) = C(S)$. Specifically we take F to be the field of fractions of R. The first step in the proof is to use the fact that $C(S)$ is isomorphic to an R-function space geometry $V(S)$. V, as a submodule of Hom (S, R), is also a subset of the F-vector space Hom (S, F), in which it generates a vector subspace W. W is a proper function space on S over F and gives a new geometry $W(S)$. It is easily seen that $W(S)$ and $V(S)$ are the same. (Let $f \in W$. By construction

$$f = \sum_1^n \frac{a_i}{b_i} f_i, \qquad a_i, b_i \in R, f_i \in V, n \geq 0.$$

Let $b = b_1 \cdots b_n$. Then $bf \in V$ and $\ker f = \ker bf$. This shows the flats of $W(S)$ are all flats of $V(S)$. The converse is immediate since $V \subseteq W$.) Hence we obtain an F-function space geometry $W(S) = C(S)$. By the assumption that $C(S)$ is a geometry, $W(S)$ is isomorphic to a chain-group geometry $C'(S)$ over F. ∎

Although this isomorphism sometimes makes one of the two previous examples superfluous, it is nevertheless useful in many contexts to be able to use both the language of chain groups and the language of proper function spaces. However, the reduction of a proper function space to a chain group, by an extension of the preceding construction beyond integral domains, is not always possible.

3.4. ALGEBRAIC EXTENSIONS OF FIELDS

The transcendence degree of an extension of a field is defined to be the cardinal number of a maximal set of algebraically independent transcendentals in the extension. The transcendence degree of a field K over a subfield F can be ascertained from the collection of fields M which are extensions of F, and which are relatively algebraically closed in K (that is, every element of K algebraic over M is in M).

Recall that an element b of a field K depends algebraically upon a subfield M if and only if b is a solution of some polynomial equation $p(x) = 0$ with coefficients in the subfield M.

PROPOSITION 3.3. Let F be an algebraically closed field, let $x_1, \cdots, x_n$ be independent transcendentals, and let $F(x_1, \cdots, x_n)$ be the associated extension field (the field of rational functions in n indeterminates, over F). Then the relatively algebraically closed subfields of $F(x_1, \cdots, x_n)$, containing F, form a geometric lattice L of rank n.

Proof: The points of the lattice L may be represented as those relatively algebraically closed subfields $F(y)$, where y is a transcendental, that is, a single rational function in the indeterminates $x_1, \cdots, x_n$. For any set T of such transcendentals, a transcendental x depends algebraically on T when x is a solution of a polynomial equation in one variable, with coefficients in $F(T)$. Then the enlargement of each set T of transcendentals, to the set of all transcendentals dependent algebraically on T, is a closure operator. A set of transcendentals is closed if and only if it is the set of transcendentals in some relatively algebraically closed field, contained between F and $F(x_1, \cdots, x_n)$.

If a transcendental x depends algebraically upon a set T of transcendentals, it depends also upon a finite set T_f of transcendentals occurring in the coefficients of some polynomial equation. Thus the closure relation has the finite basis Property 2.4.

If a transcendental x depends algebraically upon a set $T \cup y$ of transcendentals, but not upon the set T, then x is the solution of some polynomial equation, in one variable ξ, with coefficients in the field $F(T \cup y)$. Multiply this polynomial by an appropriate element of $F(T \cup y)$, substitute a variable η for y, so that it becomes a polynomial equation in two variables ξ and η over the field $F(T)$, with solution $\xi = x$, $\eta = y$. The variable η must occur nontrivially in this polynomial, for otherwise x would depend algebraically on T. Substituting $\xi = x$, we see y depends algebraically on the set $T \cup x$, and the closure relation has the exchange property. ∎

3.5. COVERINGS

A *covering* π of a set X is a family of subsets of X, with union X. A covering is a *partition of type* $\underline{n}$ (or an $\underline{n}$*-partition*) of the set X if

every member of π has at least n elements (3.1)

every n-element subset of X is contained in a *unique* member of π. (3.2)

The members of π are the *blocks* of the covering. For $n = 1$, an n-partition is an ordinary partition. For $n = 2$, the blocks may be called lines, and our requirement is that two points determine a unique line.

PROPOSITION 3.4. For any n-partition of a set X, $n \geq 2$, the following subsets:

> the set X,
> the blocks,
> and all subsets with fewer than n elements,

are all the flats of a geometry.

Proof: We are given the closed subsets of X. The closure relation with these closed sets is defined as follows. For any subset $A \subseteq S$ with fewer than n elements, $\bar{A} = A$. If a subset A has at least n elements, then it is either contained in a unique block B, so that $\bar{A} = B$, or else it is contained in no block, and $\bar{A} = X$. It is clear that $A \subseteq \bar{A}$ for all subsets $A \subseteq X$. Assume $A \subseteq \bar{B}$. Then $\bar{B}$ either has fewer than n elements, or is a block, or is equal to X. In the first instance, A has fewer than n elements, and is closed. In the second instance, A is either closed, or has closure $\bar{B}$. In any event, $\bar{A} \subseteq \bar{B}$, and the relation $A \to \bar{A}$ is a closure relation.

Any subset $A \subseteq X$ has a subset with at most $n + 1$ elements and the same closure. Thus the closure has finite basis, and it remains to prove the exchange property. Assume $a \in \overline{A \cup b}$ and $a \notin \bar{A}$, for some subset $A \subseteq X$ and two elements $a, b \in X$. Since the sets A and $\overline{A \cup b}$ differ by a single element, there are three possibilities for $\bar{A}$ and $\overline{A \cup b}$. Either $\overline{A \cup b} = X$ and $\bar{A}$ is a block, or $\overline{A \cup b}$ is a block and $A = \bar{A}$ has $n - 1$ elements, or both $A = \bar{A}$ and $A \cup b = \overline{A \cup b}$ have fewer than n elements. In the first instance, $a \notin \bar{A}$, so $\overline{A \cup a} = X$. In the second instance, $A \cup a$ has n elements, and is a subset of the block $\overline{A \cup b}$. In the third instance, $a = b$. In any case, $\overline{A \cup a} = \overline{A \cup b}$, and $b \in \overline{A \cup a}$. ∎

From each geometric lattice $L(S)$ of rank $n + 1$, a geometric lattice $TL(S)$ of rank n may be formed by identifying all the copoints of the lattice $L(S)$ with the element 1. If this *truncation* operation is performed on the lattice of a partition of type n, the resulting lattice is composed of all subsets of a set X which have fewer than n elements, together with the entire set X. Such geometries have a particularly simple structure: they are also obtainable by truncating the Boolean algebra $\mathscr{B}(S)$ of all subsets of the set S to rank n.

The converse of this observation also holds. Using the truncation operation, we obtain two equivalent characterizations of lattices of partitions of type n.

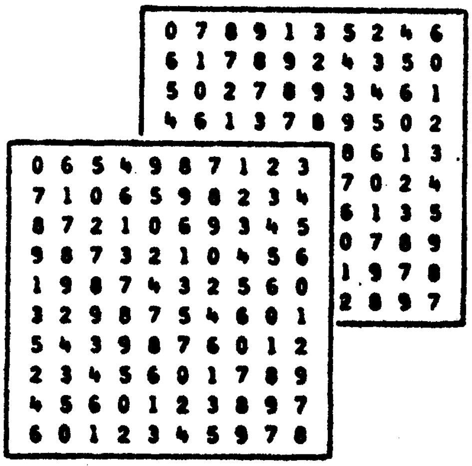
PROPOSITION 3.5. A geometric lattice $L(S)$ of rank $n + 1$ is isomorphic to the lattice of a partition of type n if and only if its truncation $TL(S)$ is isomorphic to the rank n truncation of the Boolean algebra $\mathscr{B}(S)$, if and only if every lattice interval $[0, x]$ to a coline x is distributive (and is therefore a Boolean algebra: the lattice of the free geometry on $n - 1$ points).

Proof: The preceding discussion establishes the necessity of these conditions. If the intervals $[0, x]$ to colines x are all distributive, we shall see in Chapter 4, Proposition 4.11, (or else, the reader can easily verify for himself) that all $n - 1$ element subsets of S are independent and closed. Thus the truncation $TL(S)$ coincides with the truncation of the Boolean algebra $\mathscr{B}(S)$ to rank n.

Since the lattice $L(S)$ is geometric, each n-element set $A = \{a_1, \cdots, a_n\}$ must be in a unique copoint. This copoint is obtained by adjoining the element a_n to the closed set $\{a_1, \cdots, a_{n-1}\}$, and by forming its closure. By the covering property, $\bar{A}$ must be a copoint. Each copoint has at least n elements, so the lattice $L(S)$ is the lattice of a partition of type n. ❙

University of Waterloo
Massachusetts Institute of Technology
(Received February 24, 1970)

(to be continued)



COMBINATORIAL
MATHEMATICS
YEAR

February 1969 - June 1970

ON THE FOUNDATIONS OF COMBINATORIAL THEORY III.

Theory of Binomial Enumeration

by

Ronald Mullin
University of Waterloo
Waterloo,
Ontario

Gian-Carlo Rota
Massachusetts Institute
of Technology
Cambridge, Massachusetts

Department of Statistics
University of North Carolina at Chapel Hill
Institute of Statistice Mimeo Series No. 600.24
April 1970

This is the final version of lecture notes for a series of lectures given by the second author at the Summer Study Institute on Combinatorial Mathematics and Its Applications in the Natural Sciences, held on June 2-27, 1969, in Chapel Hill, and sponsored by the U.S. Air Force Office of Scientific Research under Grant No. AFOSR-68-1406 and the U.S. Army Research Office, Durham, under Grant No. DA-ARO-D-31-124-G910.

ON THE FOUNDATIONS OF COMBINATORIAL THEORY

III. Theory of Binomial Enumeration

by

Ronald Mullin

and

Gian-Carlo Rota

1. Introduction.

The present work is born from the interplay of two seemingly disparate branches of combinatorial theory. The first is the classical calculus of finite differences, which has been in the past more often related to numerical analysis than to problems of enumeration. In the calculus of finite differences, there occur several sequences of polynomials which are used in interpolation, numerical quadrature, and several other connections. Typical of such sequences of polynomials are the lower factorials

$$(1) \qquad p_n(x) = (x)_n = x(x-1)\ldots(x-n+1), \quad m = 0,1,2,$$

and the upper factorials

$$(2) \qquad p_n(x) = x^{(n)} = x(x+1)\ldots(x+n-1), \quad n = 0,1,2, \quad .$$

Less well known, but equally significant polynomial sequences are the Abel polynomials, studied by Abel, Hurwitz and others:

$$(3) \qquad p_n(x) = x(x-an)^{n-1}, \quad n = 0,1,2,\ldots$$

and the exponential polynomials, studied by Touchard and others,

$$\varphi_n(x) = \sum_{k\geq 0} S(n,k)x^k, \tag{4}$$

where $S(n,k)$ denote the familiar Stirling numbers of the second kind. Another significant sequence is the Laguerre polynomials

$$L_n(x) = \sum_{k\geq 0} \frac{n!}{k!} \binom{n-1}{k-1}(-x)^k, \tag{5}$$

which have an extensive literature. These sequences of polynomials, as well as a large number of other sequences that have arisen in classical analysis and combinatorics, share a common property: that of being of binomial type. We say that a sequence of polynomials $p_n(x)$, where $p_n(x)$ is of exactly of degree of n, is of binomial type when it satisfies the sequences of identities

$$p_n(x+y) = \sum_{k\geq 0} \binom{n}{k} p_k(x)p_{n-k}(y), \quad n = 0,1,2,\ldots \quad . \tag{6}$$

It will be shown in the course of this study (and it is verified without difficulty using the results below) that each one of the sequences of polynomials mentioned above is of binomial type.

This work is a study of certain analytic or (more suggestively) algebraic-combinatorial properties of sequences of polynomials of binomial type. The main problem we aim at is the following: given two sequences $p_n(x)$ and $q_n(x)$, both of binomial type, there clearly exist coefficients c_{nk}, the so-called connection constants,

$$(7) \qquad p_n(x) = \sum_{k>0} c_{nk} q_k(x)$$

which express one sequence of polynomials in terms of the other. Our problem is to determine as efficiently as possible the coefficients c_{nk} in terms of minimal data on the polynomials $p_n(x)$ and $q_n(x)$. A few classical instances of this problem are given below.

In trying to solve this problem we were led to develop a systematic theory of polynomial sequences of binominal type. The main novelties we introduce in this theory are, first, a systematic use of operator methods as against less efficient generating function methods, which were used almost exclusively in the past, and secondly a solution of the connection problem stated above, which eluded past workers in the field, and which we believe to be remarkably simple.

Patches and bits of the theory developed in this work can be found in the literature of the last 50 years, starting with the work of Pincherle and Amaldi in 1900, following through the Danish and Italian schools of calculus of finite differences, culminating with the work of the great Danish actuarialist Steffensen. The statement (though not, alas, the proof) of Theorem 4 below is due to him. A few other results, such as the Expansion Theorem, where at least intuited by Pincherle and his school. But, we believe that our notion of umbral operator (a term introduced by Sylvester and extensively used by the invariant theorists and by E.T. Bell, though never correctly defined), together with our solution of the connection constants problem that it yields, gives a new direction to the calculus of finite differences, even for workers interested in purely analytic matters.

It turns out that there is a second and entirely different point of view from which the theory of polynomials of binomial types can be looked at. Each of the polynomial sequences listed above can be interpreted as counting the number of ways of placing "balls" into "boxes", subject to various restrictions. This ties in with the classical theory of distributions and occupancy, which can be alternatively considered as making words out of an alphabet, subject to various restrictions on the successions of letters. More precisely, we are given a set S with n elements and a set X with x elements, and we consider functions from the set S to the set X subject to various restrictions. The restrictions are such that they do not limit the range of the functions but only the domain. Thus, for example, the lower factorial powers (1) count the number of one-to-one functions from a set of n elements to a set of x elements. Similarly, the upper factorials

(8) $$x^{(n)} = x(x+1)(x+2)\dots(x+n-1)$$

count the different ways of placing the balls S into the boxes x when a linear ordering is to be chosen of the balls within each box.

In the same vein, the Abel polynomials

(9) $$x(x-an)^{n-1},\ n = 0,1,\dots,\ an<x,$$

can be considered in combinatorial terms. Indeed, consider a circle of circumference x, and a set of n arcs each of length a and each having the same radius of curvature as the circle. If we drop the arcs randomly on the circumference of the circle then the probability that no two arcs overlap is easily seen to be

$$(10) \qquad \frac{x(x-na)^{n-1}}{x^n} .$$

Thus the Abel polynomials "count" the ways (i.e., the measure, since this case is continuous) in which the arcs may be placed without overlapping.

Whenever we count a set of functions from a set of S to a set X, subject to restrictions on the domain, then, letting $p_n(x)$ be the number of such functions, we see immediately that $p_n(x)$ is a polynomial and that the sequence $p_n(x)$ must be of binomial type. Thus, sequences of polynomials of binomial type arise naturally as the unifying concept in the theory of distribution and occupancy.

Accordingly, the present study will be divided into two parts. In the first (the present) part we concentrate on the analytic properties of polynomial sequences of binomial type; the relationship to problems of distributions and occupancy is discussed only in Sections 2 and 10, and is meant only as an introduction to the second part. It turns out that every sequence of binomial type with positive integral coefficients can be associated to a counting problem of a certain class of "reluctant" functions, as defined in the next Section. In the second part of this work we shall intrepret the analytic results derived here in purely combinatorial, that is, set-theoretic terms.

Perhaps the most satisfying results of this investigation are, first, the unexpected relations of sequences of binomial type with problems of enumeration of rooted labeled trees, (Section 2), and secondly, the solution of the problem of the connection constants, which has deep combinatorial implications.

In several special cases, classical analysis has already answered the problem of the connection constants. For example, we have

$$x^n = \sum_{k\geq 0} S(k,n)(x)_k \tag{11}$$

$$(x)_n = \sum_{k\geq 0} s(k,n)x^k \tag{12}$$

$$x^{(n)} = \sum_{k\geq 0} |s(k,n)|x^k \tag{13}$$

where $s(k,n)$ and $S(k,n)$ are the Stirling number of the first and second kind. Another example is

$$x^{(n)} = \sum_{k\geq 0} \frac{n!}{k!} \binom{n-1}{k-1}(x)_k \tag{14}$$

$$(15) \qquad L_n(x) = \sum_{k \geq 0} (-1)^k \frac{n!}{k!} \binom{n-1}{k-1} x^k$$

where $L_n(x)$ are the Laguerre polynomials.

We hope that this introuction has given an idea of the scope of the present investigation. In the next Section we briefly outline some combinatorial connections, thereafter to dismiss them in favor of the analytic theory, until Section 10.

We thank M. Shalor and S. Smoliar for help in the preparation of this manuscript, and the Statistics Department of the University of North Carolina, as well as the U.S. Army Research Center at the University of Wisconsin, and Florida Atlantic University for giving us the opportunity to present these results to an actively contributing audience.

2. Reluctant Functions.

Given a function $f: S \to X$, where from now on S will be a finite set with n elements and X will be a finite set of x elements, we can associate with it "functorially" two objects: the range of f, namely, the sub-set of elements of X which are images of some elements of S under the function f; and the coimage of f, which consists of the partition of the set X defined by the following equivalence relation: an element a of X is equivalent to an element b of S if and only if $f(a) = f(b)$. Thus, the coimage of f is a partition of the set S.

We are now going to rather drastically generalize these concepts.

We define a reluctant function from S to X as follows. It is a function f from S to the disjoint union $S \cup X$, subject to the following restriction. For every element $s \in S$, the element $f(f(s))$ is defined if and only if $f(s) \in S$; similarly, $f(f(f(s)))$ is defined if and only if $f(f(s)) \in S$, etc. Our requirements is that only a finite number of terms of the sequence s, $f(s)$, $f(f(s))$, $f(f(f(s)))$,... be well-defined. A more suggestive, if less precise, way of stating the same condition is the following: for every element $s \in S$, the "orbit" $s, f(s), f(f(s)), f(f(f(s))), \ldots$ of s under iteration of the function f "eventually" ends up in X, where it stops. Thus, one might say that f "reluctantly" maps S into X.

The range of a reluctant function f will consist of those elements of X which are images of some element of S, just like in the case of an ordinary function. On the other hand, we need to generalize the notion of coimage of an ordinary function, as defined above, to the newly introduced concept of a reluctant function. Whereas the coimage of an ordinary function is simply a partition of the set S, the coimage a reluctant function is going to be more than a partition of S. In fact, for every element x of X which is in the range of the reluctant function f, the inverse image of the element x is defined as the set of all elements s of S such that the sequence of its successors $f(s), f(f(s)), \ldots$ eventually ends up in X. The inverse images of distinct elements of X are disjoint subsets of S. Thus, to every reluctant function there is associated a partition of the set S, just like in the case of an ordinary function. However, within each block of such a partition there is a natural structure of a forest of rooted trees describing the "history" of the elements of that block before they end up in X. Thus, we are led to define the coimage of a reluctant function to be a partition of ths set S, together with a structure of a rooted forest (i.e., set of rooted trees) defined on each block of the partition. Each rooted forest covering one block of the coimage is the "inverse image" — in the generalized sense just described — of an element x of X.

Note that each block of the coimage can be further partitioned into the connected components, that is, the trees, of the rooted forest. The resulting partition is a refinement of the coimage and has the additional property that each block has the structure of a rooted tree. We call this finer partition π of S, together with the structure of rooted tree (See Harary or Moon for definitions) on each block of π, the pre-image of the reluctant function f (recall that a rooted tree is a partially ordered set). Thus, the coimage of f is obtanied by "piecing togehter" all those blocks of the pre-image of f which are "eventually mapped" to the same element x in X.

Clearly, the pre-image of any reluctant function is a rooted labeled forest on the set S, following classical terminology. Given any rooted labeled forest L on S, with k blocks, that is, consisting of k rooted trees, there are evidently x^k reluctant functions whose pre-image is the forest L.

By way of example, let us consider the set of all reluctant functions from S to X (notice that our use of the word "from" and "to" is not strictly correct, but is nevertheless suggestive so we shall keep using it). Let c_{nk} be the number of rooted labeled forests with k blocks on the set S. Then the number of reluctant functions from S to X is evidently given by the polynomial

$$(1) \qquad \sum_{k\geq 0} c_{nk}x^k = A_n(x).$$

It is easy to see, by a simple combinatorial argument which imitates the standard set-theoretic proof of the binomial theorem, that the sequence of polynomials $A_n(x)$ is of binomial type. It is less obvious, and it will trivially follow from the present theory (see Section 10) that the polynomials $A_n(x)$ are given by the expression

$$(2) \qquad A_n(x) = x(x+n)^{n-1},$$

that is, that they are a special case of the Abel polynomials, corresponding to $a = -1$. This gives immediately the classical result of Cayley counting the number of rooted trees, since rooted trees correspond to reluctant functions having as pre-image a partition with one block, and so are the coefficients c_{n1} in (2), which equal n^{n-2}.

We define a binomial class B of reluctant functions as follows. To every set S and set X we assign a set $F(S,X)$ of reluctant functions from S to X. The assignment is "functorial" — or, in combinatorial language, "unlabeled". This means that isomorphisms of the sets S with S_0 and X with X_0 induce a natural isomorphism of the sets $F(S,X)$ with $F(S_0,X_0)$. Thus, if the polynomials $p_n(x)$ denotes the size of the sets $F(S,X)$, the function $p_n(x)$ depends only on the size n the set S and the size x of the set X.

We come now to the crucial condition. In set-theoretic terms, the condition states that there is a natural isomorphism

$$(4) \qquad F(S,X\oplus Y) = \sum_{A\subseteq S} F(A,X)\otimes F(S-A,Y).$$

Here, $\oplus$ and Σ denote disjoint sum of sets, $\otimes$ denotes product of sets, and $=$ stands for natural isomorphism.

The variable A ranges over all subsets of the set S. We set (for good reasons) $F(\emptyset, X) = 1$ for all non-empty sets X.

Taking the sizes of both sides of (*) we obtain the equation

$$p_n(x+y) = \sum_{k\geq 0} \binom{n}{k} p_k(x) p_{n-k}(y),$$

which expresses the fact that the polynomials $p_n(x)$ are a sequence of binomial type.

Roughly speaking, condition (*) states that by "piecing togehter" two reluctant functions in the family B, we again obtain a reluctant function in the family. It is a generalized set-theoretic version of the binomial theorem.

Two important ways of defining binomial classes B of reluctant functions are the following. Let T be a family of rooted trees (it is immaterial whether they are labeled or unlabeled). The family $B(T)$ will consist of all reluctant functions whose pre-images are labeled forests on S each of whose components is isomorphic to a tree in the family T. Clearly $B(T)$ is a binomial class of reluctant functions. In the example considered above, the family T consisted of all rooted trees.

Thus, we see that the enumeration of labeled forests is closely connected with the theory of polynomials of binomial type. The family T can be specifed in innumerable ways, which will be considered in the second part of the present work. For the moment, we shall give some illustrations that show that the classical polynomials listed in Section 1 can be interpreted as enumerating binomial classes of reluctant functions. We have already seen above that the Abel polynomials can be interpreted as enumerating the binomial class of all reluctant functions, as least for $a = -1$. A somewhat more elaborate argument would show that all the other Abel polynomials, for a a negative integer, enumerate other binomial classes of reluctant functions

Perhaps the simplest example is given by the sequence x^n. This enumerates the binomial class B(T), where T consists of a single tree, with one root.

Another interesting example is the sequence of Laguerre polynomials $L_n(-x)$. These enumerate the binomial class B(T), where T is the set of all linearly ordered rooted trees. We leave the easy verification of this fact to the reader.

A fourth example comes for the inverses of the Abel polynomials, considered in Section 10, namely, functional digraphs, enumerated by the polynomials

$$p_n(x) = \sum_{k \geq 0} \binom{n}{k} k^{n-k} x^k,$$

which do not appear at first sight to be of binomial type. We prove that they are, by showing that they enumerate a binomial class $B(T)$. Simply take T to be the family of all rooted trees, all of whose branches have length at most two!

Given a binomial class $B(T)$ of reluctant functions, we can consider the subclass of those functions having the property that _their coimage coincides with their pre-image_. We denote this subclass by $B_m(T)$, and call it the _monomorphic class_ associated with $B(T)$; it generalizes the notion of a one-to-one function.

The monomorphic class associated with x^n consists precisely of all one-to-one functions, enumerated by the lower factorials $(x)_n$. The monomorphic class associated with the Laguerre polynomials turns out to be enumerated by the upper factorials $x^{(n)}$ (as follows from the combinatorial interpretation of $x^{(n)}$ given above).

We state without proof (but the proof is easy) an important result about monomorphic classes. If the sequence

$$p_n(x) = \sum_{k=0}^{n} a_{nk} x^k$$

enumerates the binomial class $B(T)$, then the sequence of polynomials

$$q_n(x) = \sum_{k=0}^{n} a_{nk} (x)_k$$

enumerates the monomorphic class $B_m(T)$. This fact makes formula (14) of the preceeding Section immediately obvious, and a similar interpretation can be given to (11).

The substitution of $(x)_k$ for x^k is an instance of umbral substitution, studied generally in Section 7. It will be seen in the second part of this work that the general umbral substituions of one basic sequence into another have combinatorial interpretations in terms of "piecing together" trees and other set-theoretic operations.

These examples such suffice to orient the reader to the combinatorial aspect of the theory we are about to develop. The notion of reluctant function does not exhaust the interpretation of sequences of polynomials of binomial type. For example it does not interpret combinatorially those sequences of polynomials of binomial type which have negative or non-integral coefficients. Nevertheless, we shall see in the second part of this work that all sequences of polynomial type with non-negative coefficients can be set-theoretically (or probabilistically) interpreted by a generalization of the notion of reluctant function, whereas those with negative coefficients can be interpreted by sieving methods (Möbius inversions, etc.). There is also an obvious connection with the theory of compound Poisson processes.

Apologizing for this sketchy introduction, we proceed to begin the analytic theory.

3. Fundamentals.

Throughout this paper, we shall be concerned with the algebra (over a field of characteristic zero) of all polynomials in one variable, to be denoted $\underline{P}$.

By a polynomial sequence we shall denote a sequence of polynomials $p_i(x)$, $i = 0,1,2,\ldots$ where $p_i(x)$ is exactly of degree i, for all i.

A polynomial sequence is said to be of binomial type if it satisfies the infinite sequence of identities

$$p_n(x+y) = \sum_{k\geq 0} \binom{n}{k} p_k(x) p_{n-k}(y), \quad n = 0,1,2,\ldots$$

All the polynomial sequences mentioned above are of binomial type. For some sequences, such as x^n, this is a trivial observation, but for others, such as the Abel and Touchard polynomials, the verification that they are of binomial type will be a consequence (a rather simple one, to be sure) of our theory.

Our study will revolve primarily around the study of linear operators on $\underline{P}$ considered as a vector space. Henceforth, all operators we consider will be tacitly assumed to be linear. We denote the action of an operator T or the polynomial $p(x)$ by $Tp(x)$; this notation is not strictly correct; a correct version is $(Tp)(x)$. However, this notational license results in greater readability. By way of orientation, we list some of the operators of frequent occurrence in the theory of binomial enumeration. The most important are the <u>shift operators</u>. A shift operator, written E^a, is an operator which translates the argument of a polynomial by a, where a is an element of the field, that is, $E^a p(x) = p(x+a)$.

An operator T which commutes with all shift operators is called a <u>shift-invariant operator</u>, i.e.,

$$TE^a = E^a T .$$

The following are important examples of shift-invariant operators:

(i) Identity operator $I: x^n \to x^n$.

(ii) Differentiation operator $D: x^n \to nx^{n-1}$.

(iii) Difference operator $\Delta = E-I: (x)_n \to n(x)_{n-1}$, where we write E in place of E^1, where 1 is the identity of the field.

(iv) The Abel operator $DE^a = E^a D: x(x-na)^{n-1} \to nx(x-(n-1)a)^{n-1}$.

(v) Bernoulli operator $J: p(x) \to \int_x^{x+1} p(t)dt$.

(vi) Backward difference operator $\nabla = I-E^{-1}$: $x^{(n)} \to nx^{(n-1}$

(vii) Laguerre operator L: $p(x) \to -\int_0^\infty e^{-t}p'(x+t)dt$.

(viii) Hermite operator H: $p(x) \to \sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} e^{-t^2/2}p(x+t)dt$.

(ix) Central difference operator
$\delta = E^{1/2}-E^{-1/2}$: $p(x) \to p(x+1/2)-p(x-1/2)$.

(x) Euler (mean) operator $M = (1/2)(I+E)$: $p(x) \to (1/2)(p(x)+p(x+1))$.

We define a <u>delta operator</u>, usually denoted by the letter Q, as a shift-invariant operator for which Qx is a non-zero constant.

The derivative, difference, backward difference, central difference, Laguerre, and Abel operators are delta operators.

Delta operators possess many of the properties of the derivative operator, as we proceed to show.

<u>Lemma 1</u>: If Q is a delta operator, then $Qa = 0$ for every constant a.

<u>Proof</u>: Since Q is shift invariant, then

$$QE^a x = E^a Qx.$$

By the linearity of Q,

$$QE^{a}x = Q(x+a) = Qx+Qa = c+Qa,$$

since Qx is equal to some non-zero constant c by definition. But also

$$E^{a}Qx = E^{a}c = c$$

and so $c+Qa = c$. Hence $Qa = 0$, Q.E.D.

Lemma 2: If $p(x)$ is a polynomial of degree n and Q is a delta operator, then $Qp(x)$ is a polynomial of degree $n-1$.

Proof: It is sufficient to prove the conclusion for the special case $p(x) = x^{n}$, that is, to show that the polynomial $r(x) = Qx^{n}$ is of degree $n-1$ (exactly). From the binomial theorem and the linearity of Q we have

$$Q(x+a)^{n} = \sum_{k\geq 0} \binom{n}{k} a^{k} Qx^{n-k}.$$

Also by the shift invariance of Q

$$Q(x+a)^{n} = QE^{a}x^{n} = E^{a}Qx^{n} = r(x+a)$$

so that

$$r(x+a) = \sum_{k\geq 0} \binom{n}{k} a^k Qx^{n-k}.$$

Putting $x = 0$, we have r expressed as a polynomial in a:

$$r(a) = \sum_{k\geq 0} \binom{n}{k} a^k [Qx^{n-k}]_{x=0}.$$

The coefficient of a^n is

$$[Qx^{n-n}]_{x=0} = [Q1]_{x=0} = 0$$

by Lemma 1. Further, the coefficient of a^{n-1} is

$$\binom{n}{n-1} [Qx^{n-n+1}]_{x=0} = n[Qx]_{x=0} = nc\neq 0.$$

Hence r is of degree $n-1$, Q.E.D.

Let Q be a delta operator. A polynomial sequence $p_n(x)$ is called the sequence of <u>basic polynomials</u> for Q if:

(1) $p_o(x) = 1$

(2) $p_n(0) = 0$ whenever $n>0$

(3) $Qp_n(x) = np_{n-1}(x)$

Using Lemma 2, it is easily shown by induction that <u>every delta operator has a unique sequence of basic polynomials associated with it</u>. For example, the basic polynomials for the deriative operator are x^n.

We shall now see that severval properties of the polynomial sequence x^n can be generalized to an arbitrary sequence of basic polynomials. The first property we noticed about x^n was that it was of binomial type. This turns out to be true for every sequence of basic polynomials, and is one of our basic results.

<u>Theorem 1.</u>

(a) If $p_n(x)$ is a basic sequence for some delta operator Q, then it is a sequence of polynomials of binomial type.

(b) If $p_n(x)$ is a sequence of polynomials of binomial type, then it is a basic sequence for some delta operator.

<u>Proof:</u>

(a) Iterating property (3) of basic polynomials, we see that

$$Q^k p_n(x) = (n)_k p_{n-k}(x)$$

and hence that for $k = n$,

$$[Q^n p_n(x)]_{x=0} = n!$$

while

$$[Q^k p_n(x)]_{x=0} = 0, \; k<n.$$

Thus, we may express $p_n(x)$ in the following form:

$$p_n(x) = \sum_{k\geq 0} \frac{p_k(x)}{k!} [Q^k p_n(x)]_{x=0}.$$

Since any polynomial $p(x)$ is a linear combination of the basic polynomials $p_n(x)$, this expression also holds for all polynomials $p(x)$, i.e.,

$$p(x) = \sum_{k\geq 0} \frac{p_k(x)}{k!} [Q^k p(x)]_{x=0}.$$

Now suppose $p(x)$ is the polynomial $p_n(x+y)$. Then

$$p_n(x+y) = \sum_{k\geq 0} \frac{p_k(x)}{k!} [Q^k p_n(x+y)]_{x=0}.$$

But

$$\begin{aligned}
[Q^k p_n(x+y)]_{x=0} &= [Q^k E^y p_n(x)]_{x=0} \\
&= [E^y Q^k p_n(x)]_{x=0} \\
&= [E^y (n)_k p_{n-k}(x)]_{x=0} \\
&= (n)_k p_{n-k}(y)
\end{aligned}$$

and so

$$p_n(x+y) = \sum_{k\geq 0} \binom{n}{k} p_k(x) p_{n-k}(y)$$

which means that $p_n(x)$ is of binomial type.

(b) Conversely, suppose $p_n(x)$ is a sequence of binomial type. Putting $y = 0$ in the binomial identity, we have

$$\begin{aligned} p_n(x) &= \sum_{k\geq 0} \binom{n}{k} p_k(x) p_{n-k}(0) \\ &= p_n(x)p_o(0)+np_{n-1}(x)p_1(0)+\dots \quad . \end{aligned}$$

Since each $p_i(x)$ is exactly of degree i, it follows that $p_o(0) = 1$ (and hence $p_o(x) = 1$) and $p_i(0) = 0$ for all other i. Thus properties (1) and (2) of basic sequences are satisfied.

We now find a delta operator for which such a sequence $p_n(x)$ is the sequence of basic polynomials. Let Q be the operator defined by the property that $Qp_o(x) = 0$ and $Qp_n(x) = np_{n-1}(x)$ for $n\geq 1$. Clearly Qx must be a non-zero constant. Hence all that remains to be shown is that Q is shift-invariant.

As before we may trivially rewrite the generalized binomial theorem in terms of Q:

$$p_n(x+y) = \sum_{k\geq 0} \frac{p_k(x)}{k!} Q^k p_n(y)$$

and, by linearity, this may be extended to all polynomials:

$$p(x+y) = \sum_{k\geq 0} \frac{p_k(x)}{k!} Q^k p(y).$$

Now replace p by Qp and interchange x and y on the right to get

$$(Qp)(x+y) = \sum_{k\geq 0} \frac{p_k(y)}{k!} Q^{k+1} p(x).$$

But

$$(Qp)(x+y) = E^y(Qp)(x) = E^y Qp(x)$$

and

$$\sum_{k\geq 0} \frac{p_k(y)}{k!} Q^{k+1} p(x) = Q[\sum_{k\geq 0} \frac{p_k(y)}{k!} Q^k p(x)]$$

$$= Q(p(x+y))$$

$$= QE^y p(x).$$

Thus we have

$$E^y Qp(x) = QE^y p(x),$$

for all polynomials $p(x)$, i.e., Q is shift-invariant, Q.E.D.

4. Expansions.

We shall study next the various ways of expressing a shift-invariant operator in terms of a delta operator and its powers. The difficulties caused by convergence questions are minimal, and we shall get around them in the easiest possible way.

Consider a sequence of shift-invariant operators T_n on $\underset{\sim}{P}$. We say that the sequence converges to T, written $T_n \to T$, if for every polynomial $p(x)$ the sequence of polynomials $T_n p(x)$ converges pointwise to the polynomial $Tp(x)$. The convergence of an infinite series of operators is to be understood accordingly.

The following theorem generalizes the Taylor expansion theorem to arbitrary delta operators and basic polynomials.

Theorem 2. (First Expansion Theorem). Let T be a shift-invariant operator, and let Q be a delta operator with basic set $p_n(x)$. Then

$$T = \sum_{k \geq 0} \frac{a_k}{k!} Q^k$$

where

$$a_k = [Tp_k(x)]_{x=0}.$$

<u>Proof:</u> Since the polynomials $p_n(x)$ are of binomial type then, as usual, we rewrite the binomial formula as

$$p_n(x+y) = \sum_{k\geq 0} \frac{p_k(y)}{k!} Q^k p_n(x).$$

Now we may regard this as a polynomial in the variable y and apply T to both sides to get:

$$Tp_n(x+y) = \sum_{k\geq 0} \frac{Tp_k(y)}{k!} Q^k p_n(x).$$

Again, by linearity, this expression can be extended to all polynomials p. After doing this and setting y equal to zero we get

$$Tp(x) = \sum_{k\geq 0} \frac{[Tp_k(y)]_{y=0}}{k!} Q^k p(x) \qquad \text{Q.E.D.}$$

Obviously, the best-known example of this Theorem is when $T = I$ and $Q = D$; then $p_n(x) = x^n$, and we have Taylor's expansion. A second example is Newton's expansion, which has three forms. If $Q = \Delta$, then $p_n(x) = (x)_k$ and the coefficients are $a_k = [T(x)_k]_{x=0}$. If $Q = \nabla$ then $p_n(x) = x^{(n)}$ and $a_k = [Tx^{(k)}]_{x=0}$. The basic polynomials for $Q = \delta = E^{1/3}-E^{-1/2}$ will be determined later.

The following remark will be used occasionally:

Lemma: If Q is a delta operator, and $p(x)$, $q(x)$ any polynomials, then

$$[p(Q)q(x)]_{x=0} = [q(Q)p(x)]_{x=0}.$$

Proof: By linearity, we need only consider the cases when $q(x) = p_k(x)$ and $p(Q) = Q^n$, where $p_k(x)$ are the basic polynomials of Q. But it is easy to see that the relation holds in this case. Q.E.D.

As a further example of the use of the expansion theorem, we derive the classical Newton-Cotes formulas of numerical integration. We wish to find an expansion, in terms of Δ, of the Bernoulli operator J_r defined by:

$$J_r p(x) = \int_x^{x+r} p(t)dt.$$

Noting that J_r is a shift-invariant, we have the identities

$$J_r = \frac{(I+\Delta)^r - 1}{\Delta} \cdot \frac{\Delta}{D}$$
$$= \frac{(I+\Delta)^r - I}{\Delta} \cdot J_1$$

which reduces the problem to finding an expansion of J_1 in terms of Δ. Using the First Expansion Theorem, this is fairly simple:

$$J = \sum_{k\geq 0} \frac{a_k}{k!} \Delta^k$$

where

$$a_k = [J(x)_k]_{x=0} = \int_0^1 (x)_k dx$$

where we note that the a_k are the <u>Bernoulli numbers</u> of the <u>second kind</u>. J_2 evaluated in this way gives <u>Simpson's rule</u>:

$$\int_x^{x+2h} p(t)dt = 2h(1+\Delta+\frac{1}{6}\Delta^2+\frac{1}{180}\Delta^4+\frac{1}{180}\Delta^8+\dots)p(x).$$

A final example is the classical Euler's transformation

$$\sum_{k\geq 0}(-1)^k f(k) = 1/2 \sum_{n\geq 0} \frac{(-1)^n}{2^n} \Delta^n f(0)$$

Which follows from the identities:

$$\sum_{k\geq 0}(-1)^k E^k = \frac{I}{I+E}$$

$$= \frac{I}{2I+\Delta}$$

$$= 1/2 \frac{I}{I+1/2\Delta}$$

$$= 1/2 \sum_{n\geq 0} \frac{(-1)^n}{2^n} \Delta^n.$$

Of course, in this case we are disregarding convergence questions.

We now turn our attention to the Abel polynomials. The delta operator in this case is E^aD. Thus, the Abel polynomials are basic polynomials and hence are of binomial type. Therefore, by Thm. 1 we have proved Abel's identity:

$$(x+y)(x+y-na)^{n-1} = \sum_{k\geq 0} \binom{n}{k} x(x-ka)^{k-1} y(y-(n-r)a)^{n-k-1},$$

not easily proved by direct methods. We can use the Expansion Theorem to get an Abel expansion of e^x. Indeed, we do get the following beautiful expansion

$$e^x = \sum_{k\geq 0} \frac{x(x-ka)^{k-1}}{k!} e^{ka},$$

convergent for $a<0$.

Theorem 3. Let Q be a delta operator, and let F be the ring of formal power series in the variable t, over the same field. Then there exists an isomorphism from F onto the ring Σ of shift-invariant operators, which carries

$$f(t) = \sum_{k\geq 0} \frac{a_k}{k!} \text{ into } \sum_{k\geq 0} \frac{a_k}{k!} Q^k.$$

Proof: The mapping is already linear and by the Expansion Theorem, it is onto. Therefore, all we have to verify is that the map preserves products. Let T be the shift-invariant operator corresponding to the formal power series $f(t)$ and let S be the shift-invariant operator corresponding to

$$g(t) = \sum_{k\geq 0} \frac{b_k}{k!} t^k.$$

We must verify that

$$[TSp_n(x)]_{x=0} = \sum_{k\geq 0} \binom{n}{k} a_k b_{n-k}$$

where $p_n(x)$ are the basic polynomials of Q. Now

$$[TSp_r(x)]_{x=0} = [(\sum_{k\geq 0} \frac{a_k}{k!} Q^k \sum_{n\geq 0} \frac{b_n}{n!} Q^n) p_r(x)]_{x=0}$$

$$= [\sum_{k\geq 0} \sum_{n\geq 0} \frac{a_k b_n}{k!n!} Q^{k+n} p_r(x)]_{x=0}.$$

But $p_n(0) = 0$ for $n>0$ and $p_o(x) = 1$. Hence, it follows that the only non-zero terms of the double sum occur when $n = r-k$. Thus

$$[TSp_r(x)]_{x=0} = [\sum_{k\geq 0} \frac{a_k b_{r-k}}{k!(r-k)!} Q^r p_r(x)]_{x=0}$$

$$= [\sum_{k\geq 0} \frac{a_k b_{r-k}}{k!(r-k)!} r! p_0(x)]_{x=0}$$

$$= \sum_{k\geq 0} \binom{r}{k} a_k b_{r-k}, \qquad \text{Q.E.D.}$$

<u>Corollary 1</u>. A shift-invariant operator T is invertible if and only if $T1\neq 0$.

In the following, we shall write $P = p(Q)$, where P is a shift-invariant operator and $p(t)$ is a formal power series, to indicate that the operator P corresponds to the formal power series $p(t)$ under the isomorphism of Theorem 3. Note that $p(0) = 0$ and $p'(0)\neq 0$ whenever P is a shift-invariant delta operator. For such formal power series, a unique inverse formal power series $p^{-1}(t)$ exists.

<u>Corollary 2.</u> Let Q be a delta operator with basic polynomials $p_n(x)$, and let $q(D) = Q$. Let $q^{-1}(t)$ be the inverse formal power series. Then

$$\sum_{n\geq 0} \frac{p_n(x)}{n!} u^n = e^{xq^{-1}(u)}.$$

<u>Proof:</u> Expand E^a in terms of Q. The coefficient a_n are $p_n(a)$. Hence

$$\sum_{n\geq 0} \frac{p_n(a)}{n!} Q^n = E^a,$$

a formula which can be considered as a generalization of Taylor's formula, and which specializes (for example, for $Q = \Delta$ it gives Newton's expansion) to several classical expansions. Now use the Isomorphism Theorem, with D as the delta operators. We get

$$\sum_{n\geq 0} \frac{p_n(a)}{n!} q(t)^n = e^{at},$$

whence the conclusion, upon setting $u = q(t)$ and $a = x$, Q.E.D.

As an aside, we remark at this point a possibly useful connection between basic polynomials and orthogonal polynomials:

<u>Proposition</u>. Let $p_n(x)$, $n = 0,1,2,\ldots$ be a sequence of polynomials of binomial type. Then there exists a unique inner product $(p(x), q(x))$, on the vector space $\underline{P}$ of all polynomials $p(x)$, under which the sequence $p_n(x)$ is an orthogonal sequence and $(p_n(x), p_n(x)) = n!$. Under this inner product we have

$$[Q^n p(x)]_{x=0} = (p(x), p_n(x)/\sqrt{n!}$$

so that

$$p(x) = \sum_{n\geq 0}^{\infty} \frac{p_n(x)}{n!} [Q^n p(x)]_{x=0} = \sum_{n\geq 0} \frac{p_n(x)}{n!} (p(x), p_n(x)).$$

Proof: Let T be the (uniquely defined) operator mapping $p_n(x)$ to x^n, for all n. Define the inner product as follows:

$$(p(x), q(x)) = [(Tp)(Q)q(x)]_{x=0}.$$

An argument similar to the proof of the Lemma preceding Theorem 2 shows that this bilinear form is symmetric (set $p(x) = p_n(x)$ and $q(x) = p_k(x)$), and that $p_n(x)$ is orthogonal to $p_k(x)$ for $k \neq n$. Finally,

$$(p_n(x), p_n(x)) = [(Tp_n)(Q)p_n(x)]_{x=0} =$$

$$[Q^n p_n(x)]_{x=0} = n!,$$

which shows that the bilinear form is positive definite.

It is trivially verified that $[Q^n p(x)]_{x=0} = (p(x), p_n(x))/\sqrt{n!}$. Thus the Expansion Theorem, in the form

$$q(a) = [E^a q(x)]_{x=0} = \sum_{n\geq 0} \frac{q_n(a)}{n!} [Q^n q(x)]_{x=0}$$

is the same as the orthogonal expansion of $q(x)$ relative to the above inner product, Q.E.D.

We note that for the Laguerre polynomials, discussed below, the inner product just introduced reduces to the classical inner product making the Laguerre polynomials an orthogonal set.

Note that for the operators (i), (ii), (iii), (iv), (vi), (vii), (ix) described at the beginning of this Section the polynomials defined there are the basic sets, as will be shown in the course of this study.

5. Closed Forms.

We now introduce a class of linear operators of an altogether different kind. Let $p(x)$ be a polynomial in the parameter x. Multiplying each term of $p(x)$ by a factor x, i.e., replacing each occurrence of x^n by x^{n+1}, $n \geq 0$, — we obtain a new polynomial in x which we may denote $xp(x)$. The first x in this expression may be regarded as a linear operator since it represents a linear transformation of polynomial into polynomials. We call this the multiplication operator and we denote it by the parameter $\underline{x}$ underlined. Thus, $\underline{x}$: $p(x) \mapsto xp(x)$. Note that the operator $\underline{x}$ is not shift-invariant.

Before proceeding further, it should be noted that $E^a p(x) = p(x+a)$ is a polynomial in the formal parameter $x+a$. Since the multiplication operator is not shift-invariant, we have the operator identity:

$$E^a \underline{x} = (\underline{x+a}) E^a,$$

where $\underline{x+a}$: $p(x) \mapsto (x+a)p(x)$.

Proposition 1. If T is a shift-invariant operator, then

$$T' = T\underline{x} - \underline{x}T$$

is also a shift-invariant operator.

The proof is a straightforward verification. We call T' the Pincherle derivative of the operator T.

We saw in the previous Section, as a special case of the Expansion Theorem, that any shift-invariant operator T can be expressed as an expansion in the delta operator D, i.e., $T = \sum_{k\geq 0} \frac{a_k}{k!} D$ where $a_k = [Tx^k]_{x=0}$. Further, by the isomorphism theorem, (Theorem 3) the formal power series corresponding to T is $\sum_{k\geq 0} \frac{a_k}{k!} t^k = f(t)$. We call $f(t)$ the indicator of T.

Proposition 2. If T has indicator $f(t)$, then T' has $f'(t)$ as its indicator.

Proof: Straightforward verification of coefficients by Theorem 3.

We note in passing Pincherle's Formula:

$$Tx^n p(x) = \sum_{k\geq 0} \binom{n}{k} x^{n-k} T^{(k)} p(x).$$

Note that by the isomorphism theorem of the preceding Section, we also have

$$(TS)' = T'S+TS'.$$

<u>Proposition 3</u>. Q is a delta operator if and only if $Q = DP$ for some shift-invariant operator P, where P^{-1} exists.

<u>Proof:</u> If Q is a delta operator, then it can be written

$$Q = \sum_{k\geq 0} \frac{a_k}{k!} D^k$$

where

$$a_k = [Qx^k]_{x=0}.$$

But

$$a_0 = [Q1]_{x=0} = 0$$

$$a_1 = [Qx]_{x=0} \neq 0$$

by definition of a delta operator. Thus if we set

$$P = \sum_{k\geq 0} \frac{a_{k+1}}{(k+1)!} D^k$$

then the conclusion follows at once.

Conversely, suppose $Q = DP$ where P is shift-invariant and P^{-1} exists. Since D and P are shift-invariant, then Q must be also. Further, shift-invariant operators commute (by Theorem 3), so that

$$Qx = DPx = PDx = P1 \neq 0,$$

since $P1 \neq 0$ for an invertible shift-invariant operator. Hence Q is a delta operator. Q.E.D.

<u>Theorem 4</u> (<u>Closed forms for basic polynomials</u>). If $p_n(x)$ is a sequence of basic polynomials for the delta operator $Q = DP$, then

(1) $p_n(x) = Q'P^{-n-1}x^n$

(2) $p_n(x) = P^{-n}x^n - (P^{-n})x^{n-1}$

(3) $p_n(x) = xP^{-n}x^{n-1}$

(4) (Rodriguez-type formula) $p_n(x) = x(Q')^{-1}p_{n-1}(x)$.

<u>Proof:</u> We shall first show that (1) and (2) define the same polynomial sequence:

$$Q'P^{-n-1} = (DP)'P^{-n-1}$$

$$= (D'P+DP')P^{-n-1}.$$

Thus, if we can show that $q_n(0) = 0$ for $n>0$, we will complete the proof that $q_n(x)$ is the sequence of basic polynomials for Q, and it will follow that they will satisfy formulas (1), (2), and (3). Now, from the equivalence of equations (1), (2), (3) we see that

$$q_n(x) = xP^{-n}x^{n-1}$$

and hence $q_n(0) = 0$ for $n \geq 0$. Thus (1), (2), and (3) have been proven, and $q_n(x) = p_n(x)$.

To prove (4), we first invert formula (1), getting:

$$x^n = (Q')^{-1}P^{n+1}p_n(x).$$

Notice that Q' is invertible, as is easily verified. Inserting this into the right side of formula (3) we get:

$$p_n(x) = xP^{-n}(Q')^{-1}P^n p_{n-1}(x)$$

$$= x(Q')^{-1}p_{n-1}(x)$$

which is the Rodrigues-type formula, Q.E.D.

The following formulas, numbered (5) and (6), relate the basic polynomials of two different delta operators in an analogous way. Their proof is immediate.

<u>Corollary</u>. Let $R = DS$ and $Q = DP$ be delta operators with basic polynomials $r_n(x)$ and $p_n(x)$, respectively, where S^{-1} and P^{-1} exists. Then:

(5) $p_n(x) = Q'(R')^{-1}P^{-n-1}S^{n+1}r_n(x)$

(6) $p_n(x) = x(RQ^{-1})^n x^{-1} r_n(x)$.

<u>Example 1</u>. The Abel polynomials are the basic polynomials of the Abel operator E^aD. Indeed from formula (3):

$$p_n(x) = xE^{-an}x^{n-1}$$

$$= x(x-an)^{n-1}.$$

<u>Example 2</u>. The lower factorials $(x)_n$ are the sequence of basic polynomials for the lower difference operator $\Delta = E-I$. Since $\Delta' = E$, the Rodrigues formula (4) gives immediately

$$p_n(x) = xE^{-1}p_{n-1}(x)$$

which by iteration gives the lower factorial power $(x)_n$.

5-8

Note that the basic polynomials for the central difference operator δ can be obtained from (6) and Δ much as the Abel polynomial were obtained from (3).

6. The Automorphism Theorem.

Let $\mathcal{A}(\underline{P})$ be the algebra of all linear operators on the algebra of all polynomials $\underline{P}$. Let Σ be the subalgebra of shift-invariant operators on $\underline{P}$. We now prove our main result.

Theorem 5. Let T be an operator in $\mathcal{A}(\underline{P})$, not necessarily shift-invariant. Let P and Q be delta operators with basic polynomials $p_n(x)$ and $q_n(x)$, respectively. Assume that

$$Tp_n(x) = q_n(x), \quad \text{for all } n \geq 0,$$

then T^{-1} exists and

(a) the map $S \to TST^{-1}$ is an automorphism of the algebra Σ.

(b) T maps every sequence of basic polynomials into a sequence of basic polynomials.

(c) Let $P = p(D)$ and $Q = q(D)$, where $p(t)$ and $q(t)$ are formal power series. Let the delta operator R have formal power series expansion $r(D)$ and basic polynomials $r_n(x)$. Then

$$Tr_n(x) = s_n(x)$$

is a sequence of basic polynomials for the delta operator

$$S = r(p^{-1}(q(D))),$$

where p^{-1} is the inverse formal power series of $p(t)$, that is $p(p^{-1}(t)) = p^{-1}(p(t)) = t$.

<u>Proof</u>:

(a) We have the string of identities:

$$\begin{aligned} TPp_n(x) &= T(np_{n-1}(x)) \\ &= nTp_{n-1}(x) \\ &= nq_{n-1}(x) \\ &= Qq_n(x) \\ &= QTp_n(x) \end{aligned}$$

and since every polynomial is a linear combination of the basic polynomials, by linearity, we infer that $TPp(x) = QTp(x)$ for all polynomials $p(x)$, that is, $TP = QT$. It is clear that T is invertible, since it maps polynomials of degree n into polynomials of degree n, for all n. Hence

$$TPT^{-1} = Q$$

whence

$$TP^nT^{-1} = Q^n$$

for all $n \geq 0$. Let S be any shift-invariant operator and let the expansion of S in terms of P be

$$S = \sum_{n \geq 0} \frac{a_n}{n!} P^n.$$

Then

$$TST^{-1} = T\Big(\sum_{n \geq 0} \frac{a_n}{n!} P^n\Big)T^{-1} = \sum_{n \geq 0} \frac{a_n}{n!} Q^n \qquad \text{(I)}$$

and thus TST^{-1} is a shift-invariant operator. Furthermore the map $S \mapsto TST^{-1}$ is onto since any shift-invariant operator can be expanded in terms of Q. Thus, the map is an automorphism, as claimed.

Remark: We have also shown that T maps delta operators into delta operators, since for delta operators the constant co-efficient a_0 vanishes.

(b) Let $s_n(x) = Tr_n(x)$ and let $S = TRT^{-1}$. By the remark above, S is a delta operator since R is.

Also

$$Ss_n(x) = TRT^{-1}s_n(x)$$

$$= TRr_n(x)$$

$$= nTr_{n-1}(x)$$

$$= ns_{n-1}(x).$$

To complete the proof that $s_n(x)$ are the basic polynomials of S we need only show that $s_n(0) = 0$ for $n>0$. Now we can write

$$r_n(x) = \sum_{k\geq 1} a_k p_k(x)$$

since $a_0 = 0$ because R is a delta operator, and hence $r_n(0) = 0$. Hence

$$Tr_n(x) = \sum_{k\geq 1} a_k q_k(x) = s_n(x)$$

so that

$$s_n(0) = 0, \quad n>0,$$

as desired.

(c) Now Q and R can be written as power series in P, say $Q = f(P)$ and $R = q(P)$. In equation (I) above let

$$R = g(P) = \sum_{k \geq 0} \frac{a_n}{n!} P^n;$$

then

$$S = TRT^{-1} = g(Q)$$

and therefore

$$R = g(P) = g(p(D))$$

and

$$S = g(Q) = g(q(D)).$$

Finally we see that

$$r(D) = g(p(D))$$

$$g(D) = r(p^{-1}(D))$$

and

$$S = g(Q) = r(p^{-1}(q(D))), \qquad \text{Q.E.D.}$$

7. Umbral Notation.

In order to simplify the complex notation which has been appearing in many of the above formulas, we will make use and for the first time make rigorous the "umbral calculus" or "symbolic notation" first devised by Sylvester and later used informally by many authors. If $\{a_n(x)\}$ is a polynomial sequence then we simply note that there is a unique linear operator L on $\underline{P}$ such that $L(x^n) = a_n(x)$. We say that L is the umbral representation of the sequence $\{a_n(x)\}$. In particular, an operator T with the properties specified in the preceding Theorem will be called an umbral operator.

If $f(x)$ is a polynomial then we use the notation $f(\underline{a}(x))$ to denote the image of $f(x)$ under the operator L. For example, $\underline{a}(x)$ denotes $a_1(x)$, while $[\underline{a}(x)]^2$ denotes $a_2(x)$. Similarly, $[\underline{a}(x)+b][\underline{a}(x)+c]$ denotes $a_2(x)+(b+c)a_1(x)+bc$. This is in essence the umbral notation, which we signify by boldface lettering.

Loosely speaking, umbral notation is a simple technique using exponents to denote subscripts. For example, the defining property for a polynomial sequence to be of binomial type

$$p_n(x+y) = \sum_{k\geq 0} \binom{n}{k} p_k(x) p_{n-k}(y)$$

can be restated umbrally as

$$\underline{p}^n(x+y) = [\underline{p}(x)+\underline{p}(y)]^n.$$

Note that, in view of our definition in terms of the operator L, this identity has a well-defined meaning.

Theorem 6. If P and Q are delta operators with basic sequences $p_n(x)$ and $q_n(x)$, and expansions $p = p(D)$ and $Q = q(D)$, then the umbral composition

$$r_n(x) = p_n(\underline{q}(x))$$

is the sequence of basic polynomials for the delta operator

$$R = p(q(D)).$$

Proof: Let T be the umbral operator defined by

$$Tx^n = q_n(x).$$

By the Automorphism Theorem of the preceeding Section, it follows that T takes any basic sequence into another basic sequence. Now if

$$p_n(x) = \sum_{i=0}^{n} a_i x^i$$

then

$$Tp_n(x) = T(\sum_{i=0}^{n} a_i x^i)$$

$$= \sum_{k=0}^{n} a_i Tx^i$$

$$= \sum_{k=0}^{n} a_i q_i(x)$$

$$= p_n(\underline{q}(x)).$$

Thus $r_n(x)$ is a sequence of basic polynomials and by the Automorphism Theorem, it is the basic sequence for

$$R = TPT^{-1} = p(q(D)), \qquad \text{Q.E.D.}$$

<u>Corollary</u>: If $p_n(x)$ is a sequence of basic polynomials then there exists a basic sequence $q_n(x)$ such that

$$p_n(\underline{q}(x)) = x^n.$$

We say that $q_n(x)$ is the <u>inverse sequence</u> of $p_n(x)$.

<u>Theorem 7</u>. (<u>Summation Formula</u>). Suppose $p_n(x)$ and $q_n(x)$ are the basic sequences for the delta operators P and Q respectively. If $q_n(x)$ is inverse to $p_n(x)$, then

$$p_n(x) = \sum_{k\geq 0} \frac{x^k}{k!} [Q^k x^n]_{x=0}.$$

The proof is similar to the preceeding and is left to the reader.

We are now in a position to solve the problem stated in the Introduction: given basic sequences $p_n(x)$ and $q_n(x)$, with delta operators $P = p(D)$ and $Q = q(D)$, how are the coefficients c_{nk}

$$q_n(x) = \sum_{k\geq 0} c_{nk} p_k(x)$$

linking the $p_n(x)$ to the $q_n(x)$, the so-called <u>connection constants</u>, to be determined? The answer is dismayingly simple. Consider the polynomials

$$r_n(x) = \sum_{k\geq 0} c_{nk} x^k,$$

and consider the umbral operator T defined by

$$Tx^n = p_n(x).$$

Then clearly

$$q_n(x) = Tr_n(x) = r_n(\underline{p}(x)),$$

so that $r_n(x)$ are of binomial type and $R = r(D)$ being their delta operator, we find $q(t) = r(p(t))$, or $r(t) = q(p^{-1}(t))$. Theorem 4 then provides explicit expressions for the $r_n(x)$. One couldn't expect a simpler answer.

As an example, consider the connection constants between $q_n(x) = x^n$ and $p_n(x) = (x)_n$. Here $q(t) = t$ and $p(t) = e^t-1$. Thus, $r(t) = \log(1+t)$ and, as we shall see below, the polynomials $r_n(x)$ turn out to be the exponential polynomials $\varphi_n(x)$, discussed below.

As a second example, let $p_n(x) = (x)_n$ and $q_n(x) = x^{(n)}$. An easy computation shows that $r(t) = t/(t-1)$, whose basic polynomials are the Laguerre polynomials, also discussed below.

An instructive example the reader may work out for himself — thereby obtaining a number of classical and new identities, is to take $p_n(x) = x(x-na)^{n-1}$ and $q_n(x) = x(x-nb)^{n-1}$ for $a\neq b$. These examples could be multiplied ad infinitum, and a great number of combinatorial identities in the literature can be seen to fall into the simple pattern we have just outlined.

Remark. It can be shown that every automorphism of the algebra Σ is of the form $S \to TST^{-1}$ for some umbral operator T, but this fact will not be needed, so we omit the proof.

8. The Exponential Polynomials.

The exponential polynomials, studied by Touchard and other authors, are a good testing ground for the theory developed so far. We shall see that their basic properties and the identities they satisfy are almost trivial consequences of the theory.

Consider the sequence of lower factorials $(x)_n$, which as we have seen is the basic sequence for the delta operator $\Delta = e^D - I$. In this case the inverse sequence is the sequence of basic polynomials for the operator $Q = \log(I+D)$. We denote these polynomials by $\varphi_n(x)$; these are the <u>exponential polynomials</u>.
From the Corollary above we have umbrally

$$\underline{\varphi}(\underline{\varphi}-1)(\underline{\varphi}-2)\ldots(\underline{\varphi}-n+1) = x^n.$$

Further by the summation formula

$$\varphi_n(x) = \sum_{k\geq 0} \frac{x^k}{k!} [\Delta^k x^k]_{x=0}$$

$$= \sum_{k\geq 0} S(k,n)x^k,$$

where $S(n,k)$ denote the Stirling numbers of the second kind.

Now let us apply the Rodrigues formula to see what we get. Since $Q = \log(I+D)$, we have $Q' = (I+D)^{-1}$ and hence

$$\varphi_n(x) = x(Q')^{-1}\varphi_{n-1}(x)$$

$$= x(I+D)\varphi_{n-1}(x)$$

$$= x\varphi_{n-1}(x)+x\varphi'_{n-1}(x),$$

which is the recursion formula for the exponential polynomials.

The next property of these exponential polynomials which we shall prove is expressed umbrally as

$$\varphi_{n+1}(x) = x(\underline{\varphi}+1)^n.$$

Let T be the umbral operator that takes $(x)_n$ into x^n, so that $Tx^n = \varphi_n(x)$. Hence

$$Tx(x-1)_{n-1} = xx^{n-1}$$

or changing n to $n+1$

$$Tx(x-1)_n = xx^n = xT(x)_n$$

can be rewritten as

$$TxE^{-1}(x)_n = xT(x)_n .$$

We can extend this by linearity to all polynomials $p(x)$ so that

$$TxE^{-1}p(x) = xTp(x).$$

Replacing $p(x)$ by $p(x+1)$ we have

$$TxE^{-1}p(x+1) = xTp(x+1).$$

Hence $$Txp(x) = xTp(x+1).$$

Since $Tx^n = \varphi_n(x)$ then $Tp(x) = p(\underline{\varphi}(x))$ and it follows that

$$\underline{\varphi}(x)p(\underline{\varphi}(x)) = Txp(x)$$

$$= xTp(x+1)$$

$$= xp(\underline{\varphi}(x)+1).$$

If we let $p(x) = x^n$ then

$$\varphi_{n+1}(x) = [\varphi(x)]^{n+1} = \varphi(x)[\varphi(x)]^n$$

$$= x[\varphi(x)+1]^n$$

which is what we wanted to prove.

In a similar vein one can prove the remarkable Dobinský-type formula:

$$\varphi_n(x) = e^{-x} \sum_{k\geq 0} \frac{x^k k^n}{k!} ,$$

which we shall leave as an exercise to the reader.

9. Laguerre Polynomials.

As a further example of the above theory, we shall develop some properties of the Laguerre polynomials. The Laguerre operator L is defined by

$$Lp(x) = -\int_0^\infty e^{-x} p'(x+t)\,dt.$$

It is a delta operator and as such has a sequence of basic polynomials which we shall call $L_n(x)$. By straightforward calculation, we find that the expansion of L in terms of D has coefficients

$$[Lx^n]_{x=0} = n!, \quad n \geq 1$$

$$= 0, \quad n=0$$

and hence we find that

$$L = \frac{D}{D-I}.$$

Hence from formula (3) of Theorem 4 we have

$$(*) \qquad L_n(x) = x(D-I)^n x^{n-1}.$$

Since for all polynomial $p(x)$ we also have

$$e^x De^{-x} p(x) = e^x(e^{-x}p'(x) - e^{-x}p(x))$$

$$= (D-I)p(x)$$

then $e^x De^{-x} = D-I$ and hence

$$e^x D^n e^{-x} = (D-I)^n.$$

Therefore we obtain the classical Rodrigues formula,

$$L_n(x) = xe^x D^n e^{-x} x^{n-1}.$$

From formula (*) we find by binomial expansion that

$$L_n(x) = \sum_{k=1}^{n} \frac{n!}{k!} \binom{n-1}{k-1} (-x)^k$$

where the coefficients

$$\frac{n!}{k!} \binom{n-1}{k-1}$$

are known as the Lah numbers. Our notation for the polynomials L_n corresponds to the notation in Bateman for the polynomials $L_n^{(-1)}$, that is,

$$L_n^{(-1)}(x) = \frac{1}{n!} L_n(x).$$

We now come to the most important fact about the Laguerre polynomials. The indicator of L is

$$f(t) = \frac{t}{t-1}$$

and hence

$$f(f(t)) = \frac{\frac{t}{t-1}}{\frac{t}{t-1} - 1} = \frac{t}{t-t+1} = t$$

Thus, by the Automorphism Theorem we infer that the Laguerre polynomials are a self-inverse set. Thus, we have as an immediate consequence the beautiful identity

$$x^n = \sum_{k=1}^{n} \frac{n!}{k!} \binom{n-1}{k-1} (-1)^k L_k(x) = L_n(\underline{L}(x)).$$

Other identities concerning Laguerre polynomials stem from the fact that

$$L = \frac{D}{D-I} .$$

Since $L_n(x)$ are the basic polynomials of L we have

$$\frac{D}{D-I} L_n(x) = nL_{n-1}(x)$$

and hence the classical recursion formula

$$L_n'(x) = n(D-I)L_{n-1}(x).$$

In fact, if we expand $\frac{D}{D-I}$ into series form

$$\frac{D}{D-I} = -D \cdot \frac{I}{I-D} = -D(I+D+D^2+\ldots)$$

we can use this to get other known recursion formulas.

10. A Glimpse of Combinatorics.

Although we intend to leave most of the combinatorial applications of the preceeding theory to the second part of this work, we shall outline two typical results which we hope will orient the reader to applications to problems of enumeration, typical of the second part of this work.

Theorem 8. Let P be an invertible shift-invariant operator. Let $p_n(x)$ be a sequence of basic polynomials satisfying

$$[x^{-1}p_n(x)]_{x=0} = n[P^{-1}p_{n-1}(x)]_{x=0},$$

for all $n>0$. Then $p_n(x)$ is the sequence of basic polynomials for the delta operator $Q = DP$.

Proof: Define the operator Q by $Q1 = 0$ and

$$Qp_n(x) = np_{n-1}(x)$$

and extending by linearity. Note that Q is shift-invariant. In terms of Q, the preceding identity can be rewritten in the form

$$[x^{-1}p_n(x)]_{x=0} = [P^{-1}Qp_n(x)]_{x=0}.$$

By linearity, this extends to an identity for all polynomials $p(x)$ — an argument we have often used in this work. Thus, recalling that $[x^{-1}p(x)]_{x=0} = [Dp(x)]_{x=0}$ whenever $p(0) = 0$, we have

$$[Dp(x)]_{x=0} = [P^{-1}Qp(x)]_{x=0}$$

for all polynomials $p(x)$, including those for which $p(0) \neq 0$. Setting $p(x) = q(x+a)$ we obtain, using the shift-invariance of P and Q,

$$\begin{aligned} Dq(a) &= [P^{-1}QE^{a}q(x)]_{x=0} \\ &= [E^{a}P^{-1}Qq(x)]_{x=0} \\ &= P^{-1}Qq(a) \end{aligned}$$

for all constants a. But this means that $D = P^{-1}Q$, or $Q = DP$, Q.E.D.

<u>Corollary 1</u>. Given any sequence of constants c_{n1}, $n = 1,2,\ldots,$ there exists a unique sequence of basic polynomials $p_n(x)$ such that $[x^{-1}p_n(x)]_{x=0} = c_{n1}$, that is,

$$p_n(x) = \sum_{k\geq 1} c_{nk}x^k, \quad n = 1,2,\ldots\ .$$

Corollary 2. Let $g(x)$ be the indicator of Q in the above. Then $g = f^{-1}$, where $f(t) = \sum_{k\geq 0} c_{k,1} \frac{t^k}{k!}$.

Proof: From above

$$D = QP^{-1} = \sum_{k\geq 0} c_{k,1} \frac{Q^k}{k!} = f(Q)$$

and the result follows.

We now give some applications of the above theory.

Application 1. Let $t_{n,k}$ be the number of forests of rooted labeled trees (i.e., trees with a distinguished vertex) with n vertices and k components, then

$$A_n(x) = \sum_{k\geq 0} t_{n,k} x^k = x(x+n)^{n-1}.$$

Proof: Since $t_{n,1}$ is the number of rooted trees on n vertices, then $t_{n,1} = nA_{n-1}(1)$ since each such tree on n vertices may be obtained by mapping a forest on $n-1$ vertices onto a single new root vertex. The resulting root may be labeled in n ways, i.e., either by using a new symbol or by using one of the $n-1$ old symbols and replacing it by the new symbol. But this relation may be written

$$[x^{-1} A_n(x)]_{x=0} = n[EA_{n-1}(x)]_{x=0}$$

and hence the delta operator for A_n is DE^{-1} by Theorem 8. Thus the associated polynomials are the Abel polynomials $x(x+n)^{n-1}$.

Corollary (Cayley). The number of labeled trees on n vertices is n^{n-2}.

Proof: Since the number of rooted labeled trees is n^{n-1} the number of unrooted trees is n^{n-2} since each free tree can be labeled in n ways.

Application 2. Let S_n be a symmetric group on n symbols and let $c_{n,k}$ be the number of group elements which consist of precisely k cycles. If $C_n(x) = \sum_{k\geq 0} c_{n,k}x^k$ then $C_n(x) = x^{(n)}$.

Proof: We note that in this case $C_{n,1} = (n-1)!$ which is clearly the number of group elements consisting of just one cycle, and thus by Corollary 2 this is the required sequence.

Functional Digraphs. A digraph, D, (with loops permitted) on n symbols is a functional digraph if and only if it satisfies the following two postulates,

1) each component of D contains precisely one consistently directed circuit; and

2) each non-circuit edge is directed towards the circuit contained in its component.

An idempotent is a functional digraph all of whose components contain a distinguished vertex which meets every edge of that component.

Application 3. The polynomial $p_n(x) = \sum_{k\geq 0} \binom{n}{k} k^{n-k} x^k$ is of binomial type. Let $h_{n,k}$ be the number of idempotent on n symbols with precisely k components. Then $h_{n,k} = \binom{n}{k} k^{n-k}$ since the k distinguished vertices, V, may be chosen in $\binom{n}{k}$ ways and the remaining $n-k$ points may be directed into V in k^{n-k} ways. However, we may also view each idempotent as a structure generated by its components. It is interesting to note that the coefficients $h_{n,1} = n$ and the associated delta operator has indicator $f^{[-1]}(t)$ where $f(t) = te^t$. Thus these polynomials are the inverse sequence of the Abel polynomials. Several identities for them may be derived in much the same way as we related the exponential polynomials to the lower factorials in Section 6.

Anticipating some developments in the second part of this paper, we may state the following principle. In order to enumerate by a sequence c_{n1} a class of rooted trees, graded by the number of vertices, one forms the associated basic set, which will enumerate a class of reluctant functions, and then proceedsto apply Theorem 8 or a variant of it, which will reflect the "composition rule" of such class of trees. The connection constants between two polynomial sequences enumerating sequences of reluctant functions have a combinatorial interpretation in terms of "piecing together" one set of trees in terms of another. Thus our starting point in the second part of this work will be: given two families F_1 and F_2 of rooted labeled forests, in how many ways can be a member of F_2 be "pieced together" from members of F_1? The simplest case of this is Cayley's theorem above, where F_1 consists of a single edge and F_2 consists of all labeled rooted forests.

11. Bibliography

Rota, G.-C., On the Foundations of Combinatorial Theory, I, Theory of Möbius functions, Zeit. für. Wahr., 2 (1964), 340-368.

Crapo, H.H. and Rota, G.-C., On the Foundation of Combinatorial Theory, II, Combinatorial Geometries, M.I.T. Press (to appear).

Boas, R.P. and Buck, R.C., Poynomial Expansions of Entire Functions, Springer, New York, 1964.

Martin, W.T., On Expansions in Terms of a Certain General Class of Functions, American J. Math., 58 (1936), 407-420.

Pringsheim, A., Zur Geschichte des Taylorschen Lehrsatzes, Bibliotheca Math., (3) 1 (1900), 433-479.

Steffensen, J.F., The Poweroid, an Extension of the Mathematical Notion of Power, Acta Mathematica, 73 (1941), 333-366.

Touchard, J., Nombres Exponentiels et Nombres de Bernoulli, Canad, J. Math., 8 (1956), 305-320.

Riordan, J., Combinatorial Identities, Wiley, New York, 1968.

Al-Salaam, W.A., Operational Expressions for the Laguerre and other Polynomials, Duke Math J., 31 (1964), 127-142.

Carlitz, L. and Riordan, J., The Divided Central Differences of Zero, Canadian J. Math., 15 (1963), 94-100.

Hurwitz, A., Ueber Abel's Verallgemeinerung der binomischen Formel, Acta Math., 26 (1902), 199-203.

Lah, I., Eine neue Art von Zahlen, ihre Eigneschaften ... Mittelungsbl. Math. Stat., 7 (1955), 203-216.

Riordan, J., Inverse Relations and Combinatorial Identities, American Math. Monthly, 71 (1964), 485-498.

Rota, G.-C., The Number of Partitions of a Set, American Math. Monthly, 71 (1964), 498-504.

Moon, J.W., Counting Labelled Trees, A Survey of Methods and Results, A publication of the Department of Mathematics of the University of Alberta, 1969.

Pincherle, S. and Amaldi, Le Operazioni Distributive, Zanichelli, Bologna, 1900.

Pincherle, S., Opere Scalte, 2 vols., Cremonese, Rome, 1954.

Sylvester, J.J., Collected Mathematical Papers, Cambridge, 1904-1912, 4 vols.

Bell, E.T., Postualational Bases for the Umbral Calculus, American J. Math., 62 (1940), 717-724.

Harary, F., A Seminar in Graph Theory, Holt, New York, 1967.

Hall, Marshall, Combinatorial Theory, Blaisdell, Waltham 1967.

Harary, Frank, Graph Theory, Addison-Wesley, Reading, 1969.

Los Alamos Scientific Laboratory
Brookhaven National Laboratory
The Rand Corporation
University of Waterloo
University of North Carolina
University of Colorado
Florida Atlantic University
The Rockefeller University
Massachusetts Institute of Technology

Studies in Applied Mathematics. Vol. XLIX No. 3, September 1970.

On the Foundations of Combinatorial Theory IV Finite Vector Spaces and Eulerian Generating Functions

By Jay Goldman and Gian-Carlo Rota**

1. Introduction

The purpose of this paper is to carry out a small part of the program that was begun in Foundations I. Our main concern is the study of the combinatorial aspects of the lattice of subspaces of a vector space over a finite field, and its use in deriving various classical and new identities to be found in the literature under various guises and disguises. The central idea is to obtain as systematically as possible a set-theoretic interpretation in terms of enumeration of vector spaces and of linear transformation between vector spaces over finite fields, of various identities classically known as q-identities. These identities have almost universally been studied from different points of view, namely, from the point of view of the theory of partitions of numbers, and from the point of view of the theory of elliptic functions. The analogy between these identities and classical binomial identities has been remarked many times. In fact, it is the theme of the entire work of Jackson and of the small school of English formalists that he left. Unfortunately, Jackson's work is purely analytic, and does not reveal the set-theoretic basis for this analogy. We believe that our systematic attempt at such an interpretation reveals the structure of this analogy, which is the similarity between the lattices of subspaces of a finite vector space and the lattice of subsets of a finite set. The numerical analog of this similarity is the fact that as $q \to 1$, every finite identity on vector spaces tends to an identity on a Boolean algebra.

We begin with a brief study of the Gaussian coefficients, namely the number of subspaces of dimension k in a vector space of dimension n, displaying various analogs of binomial identities, which we prove by set-theoretic means. We then proceed to develop the method of Eulerian generating functions, which are derived as a subalgebra of the incidence algebra of the lattice of finite dimensional subspaces of an infinite-dimensional vector space (always over a finite field.) We call this subalgebra the reduced incidence algebra.

* This paper is the result of research supported by the Office of Naval Research Contract N00014-56A-0298-0017, NR-042-097.

Sections 4 and 5 are perhaps the most interesting. At the beginning of Section 4 we give a set-theoretic interpretation of formulas relating Eulerian generating functions to enumeration with the reduced incidence algebra. We then proceed to apply this principle of interpretation to various situations, obtaining enumerations of various quantities connected with vector spaces. Section 5 contains analogs of the binomial theorem for finite vector spaces and applications of the Möbius inversion formula over a finite vector space leading to various classical q-identities. We conclude with a speculative section relating to future work in this field.

We are greatly indebted to the previous work of Philip Hall, who was the first to develop the Möbius inversion formula in the context of p-groups, and to the numerous and profound papers of L. Carlitz, H. W. Gould, Sharma, Chak, Segre and several other authors, whose papers could not be listed in the bibliography because of their number.

2. The Gaussian coefficients

We begin with the q-analog of the binomial coefficients. Just as $\binom{n}{k}$ counts the number of elements of rank (size) k in the lattice of subsets of a set of n elements, we let $\binom{n}{k}_q$ be the number of subspaces of rank (= dimension) k in lattice $L(V_n)$ of subspaces of an n-dimensional vector space over the finite field $GF(q)$. The numbers of elements of rank k in this lattice are called the *Gaussian coefficients.*

It is easy to derive a formula for the Gaussian coefficients $\binom{n}{k}_q$. First enumerate all ordered bases of k-dimensional subspaces of V_n, as follows. Choose the first vector y_1 in any one of $q^n - 1$ ways (that is, excluding the zero vector). There are q vectors linearly dependent upon y_1, so the next vector y_2 can be chosen in $q^n - q$ ways. y_1 and y_2 span a two-dimensional subspace containing q^2 vectors, so we may choose y_3 in $q^n - q^2$ way, etc. Thus there are $(q^n - 1)(q^n - q)\ldots(q^n - q^{k-1})$ linearly ordered sets of k linearly independent vectors in V_n. But each k-dimensional subspace has, by the same argument, (let $n = k$) $\prod_{i=0}^{k-1}(q^k - q^i)$ ordered bases. Thus dividing out the overcount we obtain the well-known expression

$$\binom{n}{k}_q = \frac{(q^n - 1)(q^n - q)\ldots(q^n - q^{k-1})}{(q^k - 1)(q^k - q)\ldots(q^k - q^{k-1})} = \frac{(q^n - 1)(q^{n-1} - 1)\ldots(q^{n-k+1} - 1)}{(q^k - 1)(q^{k-1} - 1)\ldots(q - 1)}. \tag{1}$$

Note that as $q \to 1$, $\binom{n}{k}_q \to \binom{n}{k}$, and thus the Gaussian coefficient can be expected to share many of the properties of binomial coefficients. It is a heuristic principle that all identities between Gaussian coefficients yield as corollaries identities between binomial coefficients. Perhaps the simplest example is

$$\binom{n}{k}_q = \binom{n}{n-k}_q. \tag{2}$$

This can be seen immediately from (1), or by noting that the lattice $L(V_n)$ is *selfdual* and thus there are as many k spaces of V_n as $n - k$-*spaces*. Letting $q \to 1$ we get the familiar identity $\binom{n}{k} = \binom{n}{n-k}$.

Since (2) was derived for q any power of prime and n any positive integer, we can now think of it as an algebraic function of the variable q varying over a wider domain. Similarly we take (1) as a definition when q is a variable.

If we think of q as a real variable $0 < q < 1$ then as $n \to \infty$,

$$\binom{n}{k}_q \to \frac{1}{(1-q)\dots(1-q^k)}, \tag{3}$$

a fact which will be used later. Identity (3) remain true as q ranges over the q-adic integers.

We shall now use combinatorial arguments on the lattice $L(V_n)$ to derive identities for the Gaussian coefficients by counting sets of subspaces in different ways, in analogy with the counting of subsets in a Boolean algebra. We limit ourselves to a few examples, from which the reader will be able to glean the power and the methods.

PROPOSITION 1. (q-Pascal triangle)

$$\binom{n}{h}_q = \binom{n-1}{h-1}_q + q^h\binom{n-1}{h}_q. \tag{4}$$

Note: when $q \to 1$, (4) reduces to the usual Pascal triangle identity.

Proof: Choose a basis $x_1, \dots, x_n$, and let V_{n-1} be the space spanned by $x_1, \dots, x_{n-1}$. Now let V_h be an h-dimensional subspace of V_n. There are two possibilities for V_h.

Case 1. V_h includes the whole line spanned by x_n. If so, then $V_h \cap V_{n-1}$ is a subspace of dimension $h - 1$, and this intersection can be chosen in $\binom{n-1}{h-1}_q$ ways, accounting for the first term on the right of (4).

Case 2. V_h does not include the vector x_n. But then, the projection of V_h onto V_{n-1} along the line x_n is a subspace of dimension h, call it W_h, of V_{n-1}.

One then obtains V_h by choosing such a W_h, and then "lifting it up", that is, choosing a basis $y_1, \dots, y_h$ of W_h, and adding to each y_i a multiple of x_n. There are altogether q^h ways of performing the latter operation, and $\binom{n-1}{h}_q$ ways of performing the former. This accounts for the second term on the right of (4), and concludes the proof.

The next proposition yields both a q-identity and also enumerates a useful quantity.

PROPOSITION 2. $N_{k,1}$, the number of k-subspaces of V_n containing a fixed one-dimensional subspace, is given by

$$N_{k,1} = \binom{n-1}{n-k}_q = \frac{\binom{n}{k}_q\binom{k}{1}_q}{\binom{n}{1}_q}. \tag{5}$$

Proof: $N_{k,1} = \binom{n-1}{n-k}_q$ follows immediately from the self duality of the lattice, as already remarked, see (2). By flipping $L(V_n)$ upside down the number of k-spaces containing a given 1 space is equal to the number of $(n-k)$-spaces contained in a given $n-1$ space which by (1) is $\binom{n-1}{n-k}_q$.

To derive the right side of (5) we look at the bipartite graph whose distinct sets of vertices A and B are the k-spaces and 1-spaces respectively and we connect a k-space $V \in A$ to an l-space $V' \in B$ by an edge iff $V \supseteq V'$. Now we count the number of edges in this graph in two ways. First there are $\binom{n}{k}_q$ vertices in A and $\binom{k}{l}_q$ edges at each of these vertices. On the other hand there are $\binom{n}{l}_q$ vertices in B and $N_{k,l}$ edges at each vertex. Thus

$$\binom{n}{k}_q \binom{k}{l}_q = N_{k,l} \binom{n}{l}_q, \qquad \text{q.e.d.}$$

The next identity is a q-generalization of the binomial theorem, first proved by Cauchy by purely algebraic means. Other q-binomial theorems will be derived in Section 5.

PROPOSITION 3.

$$y^n = \sum_{k=0}^{n} \binom{n}{k}_q (y-1)(y-q)\dots(y-q^{k-1}) \tag{6}$$

Note: As $q \to 1$ we get $y^n = \sum_{k=0}^{n} \binom{n}{k} (y-1)^k$ or letting $z = y - 1$, $(z+1)^n = \sum_{k=0}^{n} \binom{n}{k} z^k$. Thus (6) is a q-analog of the binomial expansion of $(z+1)^m$.

Proof: (6) counts, in two ways, all linear transformations of V_n into a space Y with y vectors. Indeed: let $x_1, \dots, x_n$ be a basis for V_n. Then each of the x_i can map into any of the y vectors of Y and these determine the linear transformation. There are altogether y^n choices for the x_i. This accounts for the left side of the identity.

On the right hand side we enumerate linear transformations by the dimension of their null spaces. Given a subspace V_k of dimension k (and there are $\binom{n}{k}_q$ of these) let $z_1, \dots, z_{n-k}, z_{n-k+1}, \dots, z_n$ be a basis of V_n such that $z_{n-k+1}, \dots, z_n$ generates V_k. A linear transformation has V_k as its null space if and only if it maps $z_{n-k+1}, \dots, z_n$ into zero and the remaining $n-k$ vectors $z_1, \dots, z_{n-k}$ onto an independent set in Y. z_1 can be mapped anywhere into Y except the zero vector i.e. in $y-1$ ways. The vector z_2 can be mapped anywhere except to the line spanned by the image of z_1. Since such a line has q points we have $y-q$ possibilities for z_2. Proceeding in this by now familiar way we find there are

$$(y-1)(y-q)\dots(y-q^{n-k-1})$$

maps whose nullspace is a given k space V_k. Thus, there are $\binom{n}{k}_q (y-1)(y-q)\ldots(y-q^{n-k-1}) = \binom{n}{n-k}_q (y-1(y-q)\ldots(y-q^{n-k-1})$ linear transformations whose nullspace has dimension k. Summing over k we get the right side of (6).

Identity (6) can be interpreted as a q-analog of the classical binomial distribution, as follows. In the space Y, choose independently and at random a set of n vectors. What is the probability that they shall span a subspace of dimension k? By the preceding argument, this probability is

$$\binom{n}{k}_q \frac{(y-1)(y-q)\ldots(y-q^{k-1})}{y^n}.$$

As q tends to one, this tends to the classical binomial distribution

$$\binom{n}{k}\left(1-\frac{1}{y}\right)^k\left(\frac{1}{y}\right)^{n-k}.$$

3. Eulerian generating functions

Our chief tool in the study of $L(V_n)$ will be the Möbius function, which in this case gives the q-adic generalization of the principle of inclusion-exclusion. In this section we shall review some of the important points of Foundations I in the context of finite vector spaces and introduce some new concepts.

The Möbius inversion formula in $L(V_n)$ is as follows:

Let $N_=(V)$ and $N_\geq(V)$ be any two functions defined on $L(V_n)$, $V \in L(V_n)$, (with values in a commutative ring which we generally take to be the integers) satisfying the system of equations

$$N_\geq(V) = \sum_{W \geq V} N_=(W); \tag{1}$$

then there exists a function $\mu(V, W)$ defined on $L(V_n)$, independent of the functions such that

$$N_=(V) = \sum_{W \geq V} \mu(V, W) N_\geq(W). \tag{2}$$

The function μ is given by the formula

$$\mu(V, W) = \mu(0, W/V) = (-1)^k q^{\binom{k}{2}} \tag{3}$$

where $k = (\dim W/V) = (\dim W - \dim V)$.

Since $\mu(V, W)$ depends only on the difference of dimensions we set

$$\mu_k = (-1)^k q^{\binom{k}{2}} = (-1)^k q^{k(k-1)/2}. \tag{4}$$

Since $L(V_n)$ is self-dual an equivalent form of Möbius inversion is

(1a) $$N_\leq(V) = \sum_{W \leq V} N_=(W) \quad \text{implies}$$

(2a) $$N_=(V) = \sum_{W \leq V} N_\leq(W)\mu(W, V).$$

If both $N_=(V)$ and $N_\leq(V)$ depend only on the dimension of V, that is, $N_=(V) = a_k$ if dim $V = k$ and similarly $N_\geq(V) = b_k$, then collecting terms of the same dimension (1a)–(3) gives at once the beautiful numerical inversion formula

$$b_n = \sum_{k=0}^{n} \binom{n}{k}_q a_k; \qquad a_n = \sum_{k=0}^{n} (-1)^k q^{\binom{k}{2}} \binom{n}{k}_q b_k. \tag{5}$$

This is the q-analog (let $q \to 1$) of the classical inversion formula (see Riordan)

$$b_n = \sum_{k=0}^{n} \binom{n}{k} a_k; \qquad a_k = \sum_{k=0}^{n} (-1)^k \binom{n}{k} b_k \tag{6}$$

which arises from μ-inversion over the lattice of subsets of an n-element set.

But let us put the μ function in a more general setting. It is just one of many functions of two variables in $L(V_n)$ which form an interesting structure.

DEFINITION. The incidence algebra $I(V_n)$ of V_n is the set of all functions $F(V, W)$ of two variables defined on $L(V_n)$, which take values in a commutative ring R, (which we generally take to be the integers) and such that $F(V, W) = 0$ unless $V \leq W$, together with the following operations:

a) Addition: if $f, g \in I(V_n)$
let $h(V, W) = (f + g)(V, W) \equiv f(V, W) + g(V, W)$ be their sum,

b) if $c \in R$ and $f \in I(V_n)$
let $(cf)\ (V, W) \equiv c(f(V, W))$

c) if $f, g \in I(V_n)$ their *convolution* (or product) is given by $L(V, W) = f * g(V, W) \equiv \sum_{V \leq Z \leq W} f(V, Z)g(Z, W)$

It is easily verified that $I(V_n)$ is an algebra over R.

If one embeds the partial order of $L(V_n)$ in a linear order and lists the subspaces in the order $W_1, W_2, \ldots$, then a typical element f of $I(V_n)$ can be thought of as a matrix $A_{ij} = f(W_i, W_j)$ and $I(V_n)$ is isomorphic to an algebra of upper triangular matrices.

The zeta function, $\zeta(V, W) = 1$ if $V \leq W$, and 0 otherwise, belongs to $I(V_n)$ and its inverse is the Möbius function.

If one translates this to the isomorphism with upper triangular matrices then it is seen that Möbius inversion is a special matrix inversion. (See Foundation I for further details.)

It turns out in practice that one often doesn't have to study the full incidence algebra but a special subalgebra of particular combinatorial interest.

For this we recall that an *interval* or *segment* $[V, W]$ in $L(V_n)$ is given by

$$[V, W] = \{Z | V \leq Z \leq W\}.$$

DEFINITION. The reduced incidence algebra $R(V_n)$ of $L(V_n)$ is the subalgebra of $I(V_n)$ consisting of all function $f \in I(V_n)$ s.t. if $[V, W]$ is isomorphic to $[V', W']$ then $f(V, W) = f(V', W')$ i.e. those functions constant on isomorphism classes of intervals. Isomorphism is taken in the sense of partially ordered sets.

It is easily verified that $R(V_n)$ is a subalgebra.

We next determine the reduced incidence algebra of $L(V_\infty)$, the lattice of finite dimensional subspaces of a countably infinite dimensional vector space over $GF(q)$. All definitions and results of this section hold for $L(V_\infty)$ and it proves more convenient to work with this lattice which contains as sublattices $L(V_n)$ for all n.

Let the *height* of a segment $[V, W]$ be $(\dim W - \dim V) = \dim(W/V)$. Then any two segments are isomorphic if they have the same height.

In the convolution sum

$$\sum_{V \leq W \leq U} f(V, W)g(W, U) = L(V, U) \tag{7}$$

there occur as many segments $[V, W]$ of height k as there are subspaces of dimension k of the quotient space U/V. Thus if $f, g \in R(V_n)$ i.e., $f(V, W) = a_k$ whenever $\dim(W/V) = k$ and $g(W, U) = b_{n-k}$ whenever $d(U/W) = n - k$, then equation (7) simplifies to

$$c_n = \sum_{k=0}^{n} \binom{n}{k}_q a_k b_{n-k} \tag{8}$$

where $c_n = h(V, U)$ and $d(U/V) = h$. $\{c_n\}$ is called the *Gaussian convolution* of $\{a_n\}$ and $\{b_n\}$.

When $q \to 1$ (8) reduces to the binomial convolution

$$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}.$$

Since

$$\binom{n}{k}_q = \frac{(1-q)(1-q^2)\dots(1-q^n)}{(1-q)(1-q^2)\dots(1-q^k)(1-q)(1-q^2)\dots(1-q^{n-k})},$$

we can rewrite (8) as

$$\sum_{k=0}^{n} \frac{a_k}{(1-q)(1-q^2)\dots(1-q^k)} \frac{b_{n-k}}{(1-q)(1-q^2)\dots(1-q^{n-k})}$$
$$= \frac{c_n}{(1-q)(1-q^2)\dots(1-q^n)}. \tag{9}$$

Defining an Eulerian series to be a series of the form

$$\sum_{n=0}^{\infty} \frac{a_n x^n}{(1-q)(1-q^2)\dots(1-q^n)}$$

then by (8) we have probed the following:

THEOREM 1. *The reduced incidence algebra of the lattice $L(V_\infty)$ of all finite dimensional subspaces of a countable infinite dimensional vector space over a finite field with q elements is isomorphic to the algebra of Eulerian series where multiplication is defined as formal multiplication of power series. The isomorphism maps the Eulerian series*

$$\sum_{n=0}^{\infty} \frac{a_n}{(1-q)(1-q^2)\dots(1-q^n)}$$

to the element f of $L(V_\infty)$ defined by $f(V, W) = a_k$ if $d(W/V) = k$ and $f(V, W) = 0$ if $V \not\leq W$.

In particular the zeta function corresponds to the Eulerian series

$$E_q(x) = \sum_{k=0}^{\infty} \frac{x^k}{(1-q)(1-q^2)\dots(1-q^k)} \tag{10}$$

and the Möbius function is given by the Eulerian series

$$e_q(x) = \sum_{k=0}^{\infty} \frac{(-1)^k q^{k(k-1)/2} x^k}{(1-q)(1-q^2)\dots(1-q^k)} \tag{11}$$

and since ζ is the inverse of μ in $I(V_n)$

$$E_q(x)e_q(x) \equiv 1. \tag{12}$$

In analogy to the above, the reduced incidence algebra of the lattice of finite subsets of a countable set is isomorphic to the algebra of exponential series, i.e. series of the form

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!}.$$

Note however that the Eulerian series

$$\sum_{n=0}^{\infty} \frac{a_n x^n}{(1-q)\dots(1-q^n)}$$

does not converge to the exponential series $\sum (a_n x^n/n!)$ as one might expect from previous discussion. The difficulty lies in the fact that Eulerian series as we write them refer to affine spaces, not projective spaces, i.e. $(1-q)\dots(1-q^n)$ refers to enumeration of affine points (vectors) not projective points (affine lines). If we renormalize our series and write them in the form

$$A((1-q)x) = \sum \frac{a_n x^n (1-q)^n}{(1-q)\dots(1-q^n)} = \sum \frac{a_n x^n}{\binom{1}{1}_q \binom{2}{1}_q \cdots \binom{n}{1}_q},$$

then this latter series converges to $\sum (a_n x^n)/n!$ as $q \to 1$. Since the correspondence

$$A(x) = \sum \frac{a_n x^n}{(1-q)\dots(1-q^n)} \leftrightarrow \sum \frac{a_n x^n}{\binom{1}{1}_q \cdots \binom{n}{1}_q} = A((1-q)x)$$

is an automorphism of the algebra of Eulerian series it makes no real difference if we use the affine or projective form. At present we find it more convenient to use the affine form. It should be noted, however, that it always seems to be the projective form of equation that converge to the corresponding results in sets as $q \to 1$.

As we previously remarked, there are two versions of the Möbius inversion formula, according as we sum "upwards" or "downwards". One of these remains unchanged whether the dimension is finite or infinite. The other instead becomes an infinite sum, and questions of convergence become relevant. We shall see that, in contrast to other instances of Möbius inversion (see for example Hille for a discussion of the difficult convergence questions associated with the classical Möbius inversion formula), all convergence questions here can be easily resolved by use of the q-adic norm.

In the reduced incidence algebra, the "upwards" inversion formula become:

$$g_n = \sum_{k \geq n} \binom{k}{n}_q f_k, \tag{13}$$

$$f_n = \sum_{k \geq n} \binom{k}{n}_q \mu_{k-n} g_k. \tag{14}$$

PROPOSITION. A necessary and sufficient condition that either—and hence both—of the series (13) and (14) converge in the q-adic field is that

$$\sum_{n \geq 0} f_n$$

converges q-adically.

Proof: Recall that a series converges q-adically if and only if the n^{th} term converges to zero q-adically.

A straightforward computation gives $\left\| \binom{n}{k}_q \right\|_q = 1.$

Suppose $\sum f_n < \infty$, so that $\|f_n\|_q \to 0$. Then

$$\left\| \binom{k}{n}_q f_k \right\|_q = \|f_k\|_q \to 0 \quad \text{and} \quad g_n = \sum_{k \geq n} \binom{k}{n}_q f_k < \infty.$$

But since g_k and $\binom{k}{n}_q$ are bounded and

$$\|\mu_i\|_q = \frac{1}{{}_q\binom{i}{n}} \to 0$$

we also have

$$f_n = \sum_{k \geq n} \binom{n}{k}_q \mu_{k-n} g_k < \infty.$$

Conversely, suppose that the right side of (14) converges for some n, say $n = 1$, then it converges for all n, and the partial sums of the tail end on the right side of (14) must tend to zero. Thus, $f_n \to 0$, and hence $\sum_{n \geq 0} f_n$ converges, q.e.d.

4. The incidence coalgebra

In the preceding section we studied the incidence algebra of the lattice of subspaces of a vector space as an algebra of operators acting on functions from the lattice to a commutative ring. In this section we introduce an entirely different interpretation of the incidence algebra, which will lead to a combinatorial interpretation of the convolution of two elements in the incidence algebra.

We begin with some very general notions applying to every locally finite partially ordered set, but quickly specialize to the reduced incidence algebra of $L(V_n)$. It is suggested that the reader of this section refer to the notion of *coalgebra*, as is found for example in MacLane (page 197), or in the recent survey work of Heinemann and Sweedler.

Let P be a locally finite partially ordered set. Let $V(P)$ (abbreviated V) be the module, over any ring R, spanned by the intervals $[x, y]$, where $x \leq y$. We introduce a *comultiplication* in V as follows. It is a function

$$\psi : V \to V \otimes V, \tag{1}$$

where the right side is the tensor product taken relative to the ring R (which we may as well assume to be commutative), defined on the basis elements on V as follows:

$$\psi[x, y] = \sum_{x \leq z \leq y} [x, z] \otimes [z, y], \tag{2}$$

where the summation ranges over the variable z. It is easily verified that this comultiplication is coassociative. This means that the iteration of the comultiplication leads to summations of the form

$$\psi[x, y] = \sum [x, z_1] \otimes [z_1, z_2] \otimes \cdots \otimes [z_n, y], \tag{3}$$

where the summation on the right ranges over all sequences $z_1, z_2, \ldots z_n$ such that

$$x \leq z_1 \leq z_2 \leq \cdots \leq z_n \leq y. \tag{4}$$

The counit ϵ is defined by mapping

$$\epsilon([x, y]) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

With this definition we obtain a coalgebra, as is easily verified. We call this the *incidence coalgebra* $C(P)$ of the partially ordered set P, over the ring R. Most combinatorial operations on the incidence algebra really refer to the incidence coalgebra, and in fact, when reinterpreted in terms of the incidence coalgebra, they reveal their combinatorial meaning. We shall see in a moment how this is the case. Before that, let us formally recall the relationship between the incidence coalgebra and the incidence algebra. This comes from the well-known fact that the set V^* of all linear functionals on $C(V)$ with values in R has the structure of an algebra, which is precisely the incidence algebra of the partially ordered set P.

To obtain a combinatorial interpretation of the incidence algebra, we only have to closely inspect formula (3) above. Let us call a typical summand on the right hand side of (3), that is an expression

$$[x, z_1] \otimes [z_1, z_2] \otimes \cdots \otimes [z_n, y], \qquad x \leq z_1 \leq \cdots \leq z_n \leq y \tag{6}$$

a *multichain* (or chain) of a partially ordered set P. The single entries in the multichains will be called the *links*, and the chain as displayed in (6) will be said to be of length $n + 2$. We can and will now consider a multichain of length $n + 2$ as the underlying *set* of $n + 1$ links. Note that the trivial link $[z, z]$ is also allowed. We are led to the following

Main problem

To every interval $[x, y]$ associate an element of a finite set $C(x, y)$. We are to enumerate the functions from the multichains between x and y to the set $\bigcup_{x \leq y} C(x, y)$ with the property that the first link $[x, z_1]$ is mapped into the set $C(x, z_1)$, the second link $[z_1, z_2]$ is assigned to the set $C(z_1, z_2)$, and so on.

The problem is easily visualized if one interprets the sets $C(x, y)$ as "colors" to be assigned to each link of the chain. In this way, one asks for the number of colored multichains between x and y with the property that the first link is assigned a color from a given set, the second link is assigned a color from another given set, etc.

The solution of the problem is given in the following

PROPOSITION 1. Let $f(x, y)$ be the number of elements of the set $C(x, y)$, for $x \le y$. Then the solution of the Main Problem is given by the convolution, in the sense of the incidence algebra

$$\sum f(x, z_1)f(z_1, z_2)\ldots f(z_n, y), \tag{7}$$

where the summation ranges as in (4). This gives the number of colored chains of length $n + 2$.

The proof is immediate, since the Proposition is a restating of the fact that the incidence algebra is obtained as the dual of the incidence coalgebra, when the ring R is taken to be the integers.

Thus we see that the elements of the incidence algebra, which are linear functionals on the incidence coalgebra, can be considered as the solution of problems of enumeration. This justifies the contention, first advanced in Foundations I, that the incidence algebra generalizes the notion of generating function.

In a similar vein, we can interpret the convolution of different elements of the incidence algebra. Here we need assignments $C_k(x, y)$, $k = 1, 2, \ldots$ of "colors" to segments. Letting $f_k(x, y)$ be the size of the set $C_k(x, y)$ we obtain the convolution of $f_1, f_2, \ldots, f_{n+1}$ as the number of colored chains in which the ith link is assigned a color from one of the sets $C_i(x, y)$. We spare the reader the obvious details, moving instead to more concrete applications, in the incidence algebra $L(V_\infty)$ of all finite-dimensional subspaces of an infinite-dimensional space over a field with q elements. For simplicity consider the reduced incidence algebra, where convolution in the above senses reduces to an Eulerian convolution as studied in the preceding section. This amounts to studying the Main Problem in the special case when $f(x, y) = f(u, v)$ whenever $\dim y/x = \dim v/u$.

For example, let $a_i^{(2)}$ = the number of colored chains with 2 links connecting x and y, where the height of $[x, y]$ is i, that is $\dim y$-$\dim x = i$. Then we want to color chains of the form $[x, A] \otimes [A, y]$. Since $[x, y] \cong L(V_i)$, there are $\binom{i}{k}_q$ chains such that $[x, A]$ has height k and $[A, y]$ height $i - k$. The first link can be colored in $a_k = f(x, A)$ ways and the second in $a_{n-k} = f(A, y)$ ways. Thus

$$a_i^{(2)} = \sum_{k=0}^{i} \binom{i}{k}_q a_k a_{i-k}.$$

In other words, the sequence $a_i^{(2)}$ is the Eulerian convolution of the sequence a_i with itself. If $A(x)$, $A_2(x)$ are the Eulerian generating functions of a_i and $a_i^{(2)}$ respectively, then $A_2(x) = (A(x))^2$. Similarly, if $a_i^{(k)}$ is the number of chains with k links, between x and y, where the height of $[x, y]$ is i, and $A_n(x)$ is its generating function, then $A_n(x) = (A(x))^n$.

Example 1. Let $a_i = 1$ for all i, and let us count all multichains of length 2. Setting $x = 0$ (the null space), we obtain the Eulerian convolution

$$\sum_{k \geq 0} \binom{n}{k}_q = G_n, \tag{8}$$

which gives the number of subspaces of the vector space v_n of dimension n (see Goldman–Rota for a study of these numbers, called the Galois numbers).

Example 2. Let $f(u, v) = 1$ if dim k-dim $u = 1$ and $f(u, v) = 0$ otherwise: in other words, set $a_1 = 1$, and $a_i = 0$ if $i \neq 1$. The number of chains of length $n + 2$ with this restriction is simply the total number of maximal chains with $n + 1$ non-trivial links connecting x to y. Its Eulerian generating function is

$$A_n(x) = \left(\frac{x}{1 - q}\right)^n, \tag{9}$$

where n is the difference of dimensions between the sub-spaces x and y.

Summing of all n, we obtain

$$\bar{A}(x) = \sum_{n=1}^{\infty} A_n(x) = \sum_{n=1}^{\infty} \left(\frac{x}{1 - q}\right)^n = \frac{1}{1 - (x/1 - q)}, \tag{10}$$

where the coefficients of $x^i/(1 - q)(1 - q^2) \ldots (1 - q^i)$ counts all maximal chains in a segment $[x, y]$ where dim $y/x = i$.

Next we consider some examples of convolutions of distinct functions.

Example 3. Let $C_1(x, z)$ be the family of all sets of vectors in the quotient space z/x which span the space z/x. Let $C_2(z, y)$ be a set with one element. Then obtain

$$\sum_{k=0}^{n} \binom{n}{k}_q D_k = 2^{q^n}, \tag{11}$$

where the right hand side is the total number of subsets of a vector space of dimension n, and where D_k is the number of spanning subsets for a vector subspace of dimension k. But equation (11) states that the sequence $(2^{q^0}, 2^{q^1}, \ldots)$ is the Eulerian convolution of the sequences $(1, 1, 1, \ldots)$ and $(D_0, D_1, D_2, \ldots)$. Translating this into Eulerian generating functions we get

$$E_q(x)D(x) = S(x), \tag{12}$$

where $E_q(x)$ is the zeta-function of Section 3,

$$D(x) = \sum_{n=0}^{\infty} \frac{D_n x^n}{(1 - q) \ldots (1 - q^n)} \quad \text{and} \quad S(x) = \sum_{n=0}^{\infty} \frac{2^{q^n} x^n}{(1 - q) \ldots (1 - q^n)}.$$

Solving for $D(x)$ we get

$$D(x) = \frac{1}{E_q(x)} S(x) = e_q(x)S(x), \tag{13}$$

where $e_q(x)$ is the Eulerian generating function of the Möbius function. Equating coefficients in (13) we get

$$D_n = \sum_{n=0}^{\infty} \binom{n}{k}_q (-1)^k q^{\binom{k}{2}}\ 2^{q^{n-k}},$$

a result which we could also get directly by Möbius inversion.

Example 4. We wish to count the set of all pairs (A_1, A_2), where A_1 is an atom of the lattice, A_2 is a coatom of the lattice that is, a line and a hyperplane respectively, and A_1 contained in A_2.

This is achieved by taking an Eulerian convolution according to the above prescriptions, as follows: set $a_1 = 1$; $a_i = 0$ for $i \neq 0$, and set $b_i = 1$ for all i. Then the desired number is the coefficient of

$$\frac{x^n}{(1-q)(1-q^2)\ldots(1-q^n)}$$

in the Eulerian generating function $A(x)B(x)A(x)$, where $A(x)$ is the Eulerian generating function of a_i and $B(x)$ is the Eulerian generating function of b_i.

Example 5. Suppose we have a store of a_i colors, and we wish to count the number of colored maximal chains with no trivial links between x and y. Assume again that the dimension of y/x is n. Let $f_i(n)$ be such a number. Now, the last link of such a chain is of the form $[A_{n-1}, y]$ where A_{n-1} is any $(n-1)$-dimensional subspace and each such link can receive any of the a_1 colors. Thus, we are led to the recursion

$$f_1(n) = \binom{n}{n-1}_q a_1 f_1(n-1).$$

Taking the Eulerian generating function of the sequence $f_1(n)$ we obtain for the Eulerian generating function $F_1(x)$ the recursion

$$F_1(x) = \frac{1}{1-(a_1 x)/(1-q)}, \tag{14}$$

which gives the explicity form. When $a_1 = 1$ this gives the result of the Example 2, giving $\bar{A}(x)$ in (10).

Example 6. Generalizing the preceding Example let $f_k(n)$ be the number of colored chains between x and y such that links of sizes $1, 2, \ldots k$ are allowed. The links of size i can be colored in a_i colors. Repeating the preceding argument we are led to the recursion

$$f_k(n) = \sum_{i=1}^{k} a_i \binom{n}{n-i}_q f_k(n-i), \tag{15}$$

which is the general linear q-difference equation with constant coefficients. Again, taking Eulerian generating functions $F_k(x)$ we find

$$F_k(x) = \frac{1}{1 - \dfrac{a_1 x}{1-q} - \dfrac{a_2 x^2}{(1-q)(1-q^2)} - \cdots - \dfrac{a_k x^k}{(1-q)\ldots(1-q^k)}} \tag{16}$$

$$= \sum_{n=0}^{\infty} \frac{f_k(n) x^n}{(1-q)(1-q^2)\ldots(1-q^n)}.$$

Thus, we see that the general q-difference equation has a combinatorial interpretation in terms of enumeration of multichains in the lattices of subspaces of a vector space.

We hope these few examples have given the reader an idea of the scope of the method.

5. Some eulerian identities

We shall now derive by combinatorial arguments some identities, which have traditionally been associated with the theory of partitions of a number. We use two methods: direct enumeration and Möbius inversion on the lattice of subspaces.

If X and Y are subspaces of a vector space, we write $X \dot{-} Y = (X - Y) \cup \{0\}$ for simplicity.

We begin by giving a combinatorial derivation of a very general q-analog of the binomial theorem. It includes the q-binomial theorem of Section 2 as a special case and has other cases of particular combinatorial significance.

THEOREM (*q-binomial theorem*). *Let* $P_k(x, y) = (x - y)(x - qy)\ldots(x - q^{k-1}y)$ *then*

$$P_n(x, z) = \sum_{n=0}^{n} \binom{n}{k}_q P_k(x, y)P_{n-k}(y, z). \tag{1}$$

Note: Although the variable y appears only on the right side, and cancels out when the right side is expanded, it nevertheless proves very useful to write the identity in this form. As $q \to 1$ the identity reduces to the trivial identity

$$(x - z)^n = \sum \binom{n}{k}(x - y)^k(y - z)^{n-k}.$$

Proof. Let V_n, X, Y, Z be vector spaces such that $Z \subset Y \subset X$, $\dim V_n = n$, and $\dim V_n < \dim Z$. Say X, Y, Z have x, y, z vectors respectively. Equation (1) counts in two ways the set of all one-to-one linear transformations $f: V_n \to X$ such that $f^{-1}(Z) = 0$ (or equivalently $f(V_n) \cap Z = 0$).

Indeed, let $v_1, \ldots, v_n$ be a basis for V_n. We count the ways of mapping this basis into a set of n independent vectors in X whose span intersects Z in $\{0\}$. The vector v_1 can be mapped into any vector in X not in Z i.e. in $x - z$ ways; the vector v_2 can be mapped into any of $x - qz$ vectors, namely, all vectors in X except those lying in the subspace spanned by Z together with the image of V_1; similarly, for the vector v_3 there are $x - q^2z$ choices, all vectors in X except the members of the space spanned by the images of V_1 and V_2 together with Z, and so on. Thus, the number of one-to-one linear transformations whose image doesn't intersect Z is $(y - z)(y - qz)\ldots(y - q^{n-1}z) = P_n(x, z)$.

Next, we again count the set of all one-to-one linear transformations whose image is disjoint from Z, according to the position of the image of V_n relative to Y. Let f be such a transformation. Then $f(V_n) \cap Y$ is a subspace of some dimension, say k; hence it is the image of a k-dimensional subspace of V_n. Thus we can construct such an f by first choosing an arbitrary subspace U of V_n, next mapping U into Y but outside $Z - \{0\}$, and mapping $V_n - U$ into $X \dot{-} Y$. This leads to the following enumeration: let $v_1, \ldots, v_n$ be a basis for V_n such that $v_1, \ldots, v_k$ is a basis for U. Then, as in the first part of the proof, the number of one-to-one linear maps of U into $Y \dot{-} Z$ is $P_k(y, z) = (y - z)(y - qz)\ldots(y - q^{k-1}z)$. The remaining basis vectors must map into $X \dot{-} Y$ in such a way that f is one-to-one. As above, this can be done in $P_{n-k}(x, y)$ ways since any set of independent vectors in $X \dot{-} Y$

is independent of any set of vectors in Y. Thus, for every k-dimensional subspaces U of V_n there are $P_k(y, z)P_{n-k}(x, y)$ one-to-one linear transformations of U_n into X whose image intersected with $Y \div Z$ is the image of U. Hence there are

$$\sum_{k=0}^{n} \binom{n}{k}_q P_k(y, z)P_{n-k}(x, y)$$

one-to-one linear maps of V_n into $X \div Z$. We conclude that

$$P_n(x, z) = \sum_{k=0}^{n} \binom{n}{k}_q P_k(y, z)P_{n-k}(x, y),$$

which is the desired result.

Several special cases of (1) are worth remarking.

COROLLARY 1. *Setting $z = 0, y = 1$ in* (1), *we obtain*

$$x^n = \sum \binom{n}{k}_q (x-1)(x-q)\ldots(x-q^{k-1}),$$

which is Proposition 3 *of Section* 2.

COROLLARY 2. *Set $z = 1, y = 0$ then*

$$(x-1)(x-q)\ldots(x-q^{n-1}) = \sum_{k=0}^{n} \binom{n}{k}_q x^k q^{\binom{n-k}{2}}(-1)^{n-k} \tag{2}$$

$$= \sum_{k=0}^{n} \binom{n}{k}_q (-1)^k q^{\binom{k}{2}} x^{n-k}.$$

This identity goes back to Cauchy (and probably even earlier).

(2) can also be derived directly by a Möbius inversion argument as follows: count all one-to-one linear transformations from V_n into X. The left side clearly counts this directly by the same method used in proving (1). To derive the right side let $N_=(W)$ be the number of linear transformations from V_n to X whose null space *equals* W and let $N_\geq(W)$ be the number of linear transformation from V_n to X whose null space *contains* W. Clearly

$$N_\geq(U) = \sum_{W \geq U} N_=(W)$$

for every subspace U of V_n. Hence by Möbius inversion

$$N_=(U) = \sum_{W \geq U} \mu(U, W)_\geq(W)$$

and setting $U = 0$ (0 is the zero subspace)

$$N_=(0) = \sum_{W \subseteq V_n} \mu(0, W)N_\geq(W).$$

But $N_=(0)$ counts all one-to-one linear transformations since a linear transformation is one-to-one iff its null-space is 0. Thus we count the one-to-one maps by "sieving" through all maps. We now have the identity

$$(x-1)(x-q)\ldots(x-q^{n-1}) = \sum_{W \subseteq V_n} \mu(0, W)N_\geq(W). \tag{3}$$

To compute $N_\geq(W)$, the set of all linear transformations that send W into 0, let $v_1, \ldots, v_n$ be a basis for V_n such that $v_1 \ldots v_k$, $k = \dim W$, are a basis for W. Then each of $v_1 \dots v_k$ must map into zero and the remaining $v_{k+1}, \dots, v_n$ can map into any vector in X and we can choose the images in x^{n-k} ways, i.e. $N_{\geq}(W) = x^{n-\dim W}$. Substituting this result and the value of μ into (3) we prove our identity (2).

However (3) also gives us a new method of computing the Möbius function. If we substitute the value of $N_{\geq}(W)$ into (3), observe that $\mu(0, W)$ depends only on the dimension of W (i.e. the isomorphism type of $(0, W)$) [see Rota (1964)], and equate the right hand side of equations (2) and (3), then equating coefficients of x^k we get

$$\mu(0, W) = (-1)^k q^{\binom{k}{2}}, \quad \text{where} \quad k = \dim W.$$

It is now easy to derive from (2) a famous identity due to Euler. Replace y by x^{-1} and multiply both sides by x^n, obtaining

$$(1 - x)(1 - qx)(1 - q^2x)\dots(1 - q^{n-1}x) = \sum_{k=0}^{n} (-1)^k \binom{n}{k}_q q^{k(k-1)/2} x^k.$$

This is a polynomial identity holding for all x and q. We may therefore let n tend to infinity, and obtain Euler's identity (convergence obtains in the q-adic norm, trivially):

$$\prod_{n=0}^{\infty} (1 - xq^n) = \sum_{k=0}^{\infty} \frac{(-1)^k q^{k(k-1)/2} x^k}{(1 - q)(1 - q^2)\dots(1 - q^k)} = e_q(x), \tag{4}$$

where we have used the fact that, as $n \to \infty$, in the q-adic norm

$$\binom{n}{k}_q \to \frac{1}{(1 - q)(1 - q^2)\dots(1 - q^k)},$$

and where $e_q(x)$ is the generating function of μ introduced in Section 3. Therefore the general fact that the Mobius function is the inverse of the zeta function yields at once another famous identity of Euler, namely

$$\prod_{n=0}^{\infty} \frac{1}{(1 - xq^n)} = \sum_{k=0}^{\infty} \frac{x^k}{(1 - q)(1 - q^2)\dots(1 - q^k)} = E_q(x). \tag{5}$$

COROLLARY 3. *Set $y = 0$ in* (1), *then*

$$(x - z)\dots(x - q^{n-1}z) = \sum_{k=0}^{n} \binom{n}{k}_q x^k (-1)^{n-k} q^{\binom{n-k}{2}} z^{n-k} \tag{6}$$
$$= \sum_{k=0}^{n} \binom{n}{k}_q (-1)^k q^{\binom{k}{2}} z^k x^{n-k}.$$

This identity seems a little more general than (2) but can actually be derived from it by setting $x = x/z$. We can again derive this directly by counting all one-to-one linear transformations of V_n into X such that $F(V_n) \cap Z = 0$. The left side has been proved in (2). The right side comes from setting

$N_{=}(W)$ = number of linear transformation f such that $f^{-1}(Z) = W$.

$N_{\geq}(W)$ = number of linear transformations f such that $f^{-1}(Z) \geq W$,

and proceeding by Möbius inversion as in Corollary 2.

It is interesting to note that setting $z = 1$ in (6) yields (2) algebraically. Geometrically, it corresponds to shrinking the subspace Z to 0. It would be of the utmost interest to extend this correspondence between the general algebraic and geometric theory.

To conclude, we derive an identity which does not appear to be a corollary of (1).

PROPOSITION 1.

$$(x - z)(x - qz)\dots(x - q^{n-1}z) = \sum_{k\geq 0} \binom{n}{k}_q (-1)^k q^{\binom{k}{2}} \prod_{i=0}^{k-1} (z - q^i) \prod_{i=0}^{n-k-1} (x - q^{k+i}). \tag{7}$$

Proof. As in (2) the left side counts the set of all one-to-one linear transformations f from V_n to X s.t. $f(V_n) \cap Z = 0$. We derive the right side as in (2) by Möbius inversion, but this time we sieve only through the one-to-one transformations instead of all linear transformations. Let

$N_=(W)$ be the number of one-to-one linear transformations $f: V_n \to X$ such that $f^{-1}(Z) = W$.

$N_\geq(W)$ equal the number of one-to-one linear transformations $f: V_n \to X$ such that $F^{-1}(Z) \supseteq W$.

Then $N_\geq(U) = \sum_{W\geq U} N_=(W)$. Inverting and setting $U = 0$ we get $N_=(0) = \sum_{W\subseteq V_n} \mu(0, W) N_\geq(W)$ which is the desired identity.

We next compute $N_\geq(W)$. Let $k = \dim W$. Then a linear transformation is counted by $N_\geq(W)$ whenever it is one-to-one and it sends W into Z. Therefore,

$$N_\geq(W) = \prod_{i=0}^{k-1} (z - q^i) \prod_{i=0}^{n-k-1} (x - q^{k+i}).$$

Substituting this in the formula for $N_=(0)$ we get (7), q.e.d.

Letting $n \to \infty$ in (7) we obtain the following infinite q-identity

$$\prod_{i=0}^{\infty} (x - q^i z) = \sum_{k=0}^{\infty} \frac{(-1)^k q^{\binom{k}{2}} \prod_{i=0}^{k-1} (z - q^i) \prod_{i=0}^{\infty} (y - q^{k+1})}{(1 - q)(1 - q^2)\dots(1 - q^k)}. \tag{8}$$

The particular interest case of this comes from setting $x = 1, z = 0$. This yields

$$\begin{aligned} 1 &= \sum_{k=0}^{\infty} \frac{(-1)^k q^{\binom{k}{2}} \prod_{i=0}^{k} (-q^i) \prod_{i=0}^{\infty} (1 - q^{k+i})}{(1 - q)\dots(1 - q^k)} \\ &= \sum_{k=0}^{\infty} \frac{q^{2\binom{k}{2}} \prod_{i=0}^{\infty} (1 - q^{k+i})}{(1 - q)\dots(1 - q^k)} \frac{\prod_{i=1}^{k} (1 - q^i)}{\prod_{i=1}^{k} (1 - q^i)} \\ &= \sum_{k\geq 0} \frac{q^{2\binom{k}{2}} (1 - q^k) \prod_{i=1}^{\infty} (1 - q^i)}{(1 - q)^2 \dots (1 - q^k)^2}. \end{aligned}$$

From this we arrive at the

COROLLARY

$$\frac{1}{\prod_{i=1}^{\infty}(1-q^i)} = \sum_{k=0}^{\infty} \frac{q^{2\binom{k}{2}}(1-q^k)}{(1-q)^2\ldots(1-q^k)^2}. \tag{9}$$

This generalizes Durfee's identity (Hardy and Wright p. 281) in the theory of partitions of a number, by looking at the largest $k \times (k-1)$ rectangle in the Ferrers diagram of a partition instead of the largest (Durfee) square.

6. Further work and open problems

The present paper does little more than scratch the surface of a field of research that may prove fertile in connecting various branches of mathematics. In closing, we should like to outline some of the directions in which further work might proceed, and some of the research problems that appear at present to us to be most promising.

Perhaps the most tantalizing open problem is a construction of a q-probabilistic setup within which formulas such as the Eulerian expansions (4) and (5) of the preceding section can be justified. One can see heuristically that these formulas should be related to some sort of q-Poisson distribution, as follows. From formula (2) of the preceding section we see that the following is true: Given a vector space X with x vectors, and a linear transformation from V_n to X picked at random, the probability that this linear transformation shall be one-to-one is

$$\sum_{k=0}^{n} \binom{n}{k}_q (-1)^k q^{\binom{k}{2}} x^{-k}. \tag{1}$$

This is simply the right side of (2) divided by the total number of linear transformations from the space V_n to the space X, which is x^n.

Now let $n \to \infty$ in (1). The series converges in the q-adic norm and yields precisely the Eulerian formula (4) of the preceding section, with x^{-1} replacing x. It is therefore reasonable to surmise that this should be a candidate for the probability stated above.

A deeper argument would yield the q-analog of a Poisson distribution along the same lines. However, a set-theoretic justification of this heuristic limit-taking is much more difficult to obtain. We have obtained some results in this direction using some methods drawn from Von Neumann's theory of continuous geometries, specifically, using as the analog of probability the dimension function in a continuous geometry constructed by an inductive limit of finite-dimensional vector spaces over a field with q elements. In a certain sense, this continuous geometry is the q-analog of a nonatomic probability space. The difficulty is that this geometry is not represented by subspaces of a vector space (only by ideals in a regular ring), and the development of a probability theory becomes exceedingly delicate. Nevertheless, we think that it is both possible and desirable.

In the same vein, we surmise that most of the identities that heretofore have been proved using the theory of partitions and elliptic functions are interpretable set-theoretically in terms of enumeration in finite or in infinite-dimensional vector spaces. A simple example is the theta series

$$\sum_{k=0}^{\infty} (-1)^n q^{n(n-1)/2} x^n, \tag{2}$$

which is the Eulerian generating function of the number of automorphisms of a vector spaces of dimension n over a field with q elements. The ultimate goal is a set-theoretic proof and interpretation of the Rogers–Ramanujan identities.

In this connection, one of the mysteries is the dual interpretation of various Eulerian formulas both as enumeration of subspaces or of linear transformations in vector spaces with given properties, and on the other hand as partitions of a number with given properties. A connection between these two approaches would be very revealing.

In another direction, the interpretation of linear q-difference equations with constant coefficients derived in Section 4 could be extended to more complex q-difference equations with nonconstant coefficients (for example, the q-difference equations for the q-hypergeometric function of Jackson).

The Theorem of Section 5 can be made the basis for q-analogs of the theories of Appell polynomials and basic polynomials. The first such theory would be concerned with polynomials $r_n(x, y)$ satisfying the identities

$$r_n(x, z) = \sum_{k=0}^{n} \binom{n}{k}_q r_k(x, y) P_{n-k}(y, z), \tag{3}$$

where P_n is defined in Section 5. These are the analogs of Appell polynomials. We know of several examples of polynomials in the literature that satisfy these identities. For example, the q-Hermite polynomials introduced by Carlitz.

On the other hand, a system of basic polynomials $b_n(x, z)$ satisfies the identities

$$b_n(x, y) = \sum_{k=0}^{n} \binom{n}{k}_q b_k(x, y) b_{n-k}(y, z), \tag{4}$$

an analogy with the basic polynomials such as developed for example in Foundations III. Such systems of polynomials seem to be rare. In the same vein, one can introduce the analog of shift operators and the notion of a shift-invariant operator, again in an analogy with Foundations III.

The analogy between q-identities and binomial identities can be carried further to a straight set-theoretic analogy between problems and concepts for subspaces of vector spaces and on the other hand for subsets of a set. One of the most intriguing notions to be developed is the q-analog of the notion of a partition of a set. In this connection see the recent work of Bender–Goldman.

A particularly simple analog is the classical result of Sperner regarding the maximal antichain in the lattice of subsets of a set. This theorem, as well as the well-known proof given by Lubell, carry over almost without change to the lattice of subspaces of a vector space (see Lubell, or Harper–Rota). A more intriguing conjecture is the analog of Ramsey's theorem, which has recently been studied by Graham and Rothschild. Finally, one of us has begun a study of q-analogs of the Euler characteristic, q-analogs of simplicial complexes, and, in a more speculative vein, q-analogs of homology.

All these questions point to a major problem, which is of a somewhat philosophical nature; this is the problem of explaining why q-identities relating to vector spaces tend to ordinary identities for binomial coefficients as q tends to 1. A purely set-theoretic explanation of this fact would be of great significance.

References

This bibliography does not repeat any references listed under Foundations I, II, III.

1. Bender, E. and Goldman J., "The Enumerative Uses of Generating Functions", Journal of Mathematics and Mechanics (to appear).
2. Carlitz, L., "Some Polynomials Related to Theta Functions", Ann. Math. Pura Appl., 359–373, 1956.
3. Carlitz, L., "Some Polynomials Related to Theta Functions", Duke Math. J. **24**, 521–527, 1957.
4. Crapo, H. and Rota, G.-C., "On the Foundations of Combinatorial Theory II, Combinatorial Geometries" (to appear).
5. Gauss, C. G., "Summation Quarundam Serierum Signularium", Werke **2**, 9–45, 1876.
6. Goldman, J., "The Method of Special Cases in Combinatorial Analysis and Probability", (to appear).
7. Goldman, J. and Rota, G.-C., "The Number of Subspaces of a Vector Space" in Recent Progress in Combinatorics edited by W. Tutte, Academic Press, 75–83, 1969.
8. Gould, H. W., "Combinatorial Identities, A Standardized Set of Tables Listing 500 Binomial Coefficient Summations", West Virginia University, 1959.
9. Graham, R. and Rothschild, B., "Ramsey's Theory for n-Dimensional Arrays", Bulletin of American Math. Soc., **75**, 418–422, 1969.
10. Harper, L. H. and Rota, G.-C., "Matching Theory, An Introduction," to appear in Studies in Probability ed. by P. Ney, North Holland.
11. Heyneman, R. and Sweedler, M., "Affine Hopf Algebras, I", Journal of Algebra **13**, 192–241, 1969.
12. Jackson, F. H., "Basic Intergration", Quarterly J. of Math., Oxford, **2**, 1–6, 1951.
13. MacLane, S., "Homology", Academic Press and Springer Verlag, 1963.
14. Mullin, R. and Rota, G.-C., "On the Foundations of Combinatorial Theory III, Theory of Binomial Enumeration", to appear (Academic Press).
15. Rota, G.-C., "On the Foundations of Combinatorial Theory, I", Zeitschrift fur Wahrscheinlichkeits-theorie, Band 2, Heft 4, 5 340–368, 1964.
16. Rota, G.-C., "The Number of Partitions of a Set", Amer. Math. Monthly, **71**, 498–504, 1964.
17. Rota, G.-C., "Mobius Functions and the Euler Characteristic", Rado Festschrifte, Academic Press, 1970.

Harvard University
Massachusetts Institute of Technology

(Received May 21, 1970)

Studies in Applied Mathematics, Vol. L, No. 4, December 1971. Copyright ©. 1971 by The Massachusetts Institute of Technology

On the Foundations of Combinatorial Theory V, Eulerian Differential Operators

By George E. Andrews[1]

(Communicated by G.-C. Rota)

1. Introduction

In [18], Rota and Mullin develop a theory of binomial enumeration by making an extensive study of polynomials of binomial type, that is sequences $p_0(X), p_1(X), p_2(X), \ldots$ where $p_n(X)$ is of degree n, and

$$p_n(X + Y) = \sum_{j \geq 0} \binom{n}{j} p_j(X) p_{n-j}(Y). \tag{1.1}$$

As they remark early in their paper, such sequences arise naturally in problems of enumeration. For example, if $p_n(X) = X(X - 1) \ldots (X - n + 1)$, then $p_n(X)$ enumerates the number of one-to-one mappings of a set of n elements into a set of X elements. In this instance, equation (1.1) is an obvious combinatorial assertion. Namely, $p_n(X + Y)$ is now the number of one-to-one mappings of a set of n

[1] This research was partially supported by the National Science Foundation Grant GP-9660.

I wish to express my gratitude to Professor Gian-Carlo Rota for the many valuable suggestions and comments he made concerning this work and for his invitation to include this paper in the series: On the Foundations of Combinatorial Theory.

elements into a set of $X + Y$ elements, while $\binom{n}{j} p_j(X)p_{n-j}(Y)$ is the number of such one-to-one mappings with exactly j elements mapped into the set of X elements.

In [12], p. 257, Rota and Goldman suggest the importance of a similar study for polynomials related to enumeration problems in finite vector spaces. Namely they suggest consideration of sequences of polynomials satisfying

$$b_n(X, Z) = \sum_{j \geq 0} \binom{n}{j}_q b_j(X, Y)b_{n-j}(Y, Z), \tag{1.2}$$

where $\binom{n}{j}_q$ is the Gaussian polynomial (see Section 2 for definition). They note, however, that such systems of polynomials are seemingly rare. Apparently only one example of such polynomials appears in the literature [12], p. 252, equation (1); however, as we shall see in Section 6, there are infinitely many systems satisfying (1.2) (see Theorem 7).

The object of this paper is to develop a theory for enumeration problems in finite vector spaces that is analogous to the theory Rota and Mullin [18] developed for finite sets.

In Sections 4 and 5, our theory very much parallels the work of Rota and Mullin; however in succeeding sections, the two theories are seen to go their separate ways. As it turns out the theory developed here has application not only to finite vector spaces (Section 11) but also to certain areas of classical analysis, for example, the Rogers–Ramanujan identities (Section 9).

2. Notation

The theory of basic hypergeometric series has always been plagued with a one-to-one correspondence between systems of notation and active researchers. The following table lists the most common notation.

Table 1

Notation for the Product $\prod_{j=0}^{n-1} (1 - aq^j)$

Author	Bailey	Fine	Jackson	Rota, Goldman	Slater	Watson
A work in which notation is used	[1]	[10]	[15]	[12]	[21]	[24]
Notation	$(a)_{q,n}$	$[n; aq^{-1}; q]$	$(1 - q)^n[\log_q a]_n$	$P_n(1, a)$	$(a; q)_n$	$\Pi_n(-a, q)$

Some other works in the subject use a great variety of symbols for special cases of $\prod(1 - aq^j)$ (see for example [2], p. 421).

For reasons that will become apparent as our work progresses, we shall use both the notation of Rota and Goldman as well as that of Slater. Thus the following symbols will be used throughout our work.

$$\begin{aligned} P_n(x, z) &= (x - z)(x - zq)\ldots(x - zq^{n-1}) \\ &= x^n\left(1 - \frac{z}{x}\right)\left(1 - \frac{z}{x}q\right)\ldots\left(1 - \frac{z}{x}q^{n-1}\right) \\ &= x^n\left(\frac{z}{x}; q\right)_n = x^n\left(\frac{z}{x}\right)_n. \end{aligned}$$

For the Gaussian polynomial, we use the notation of Rota and Goldman [12], p. 240:

$$\binom{n}{m}_q = \begin{cases} \dfrac{(q)_n}{(q)_m(q)_{n-m}}, & \text{if } 0 \le m \le n, \\ 0, & \text{otherwise.} \end{cases}$$

We also require some conventions concerning finite vector spaces. Upper case script Latin letters $\mathcal{N}, \mathcal{T}, \mathcal{U}, \mathcal{W}, \mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ will denote finite vector spaces; upper case Latin letters N, T, U, W, X, Y, and Z will, in this context, denote the number of elements of such spaces, and lower case Latin letters n, t, u, w, x, y, and z will, in this context, denote the dimensions of these spaces. Thus $\mathcal{X}$ is a vector space of dimension x over $GF(q)$ the finite field of q elements, and there are $X = q^x$ elements of $\mathcal{X}$.

Also every time we refer to a "map" or "mapping" we shall mean a one-to-one linear transformation of one finite vector space into another.

A comment should also be made concerning the names given to various operators, sequences, and series that arise in the course of our work. First we have decided against using "q-operator", "q-basic polynomial" etc., although q-terminology is quite extensive in the literature. Rather we shall use the adjective "Eulerian" paying tribute to the first worker in q-series [3], p. 47. However this requires that most things be given three-word names; this is necessary since terms such as "Eulerian operators", "Eulerian numbers", "Eulerian polynomials", already have been used to describe constructs much different from those in this paper (see [7], [8], and [9]).

3. Transformations of a finite vector space

Goldman and Rota [12], p. 252, have shown that $P_n(X, Z)$ is the number of one-to-one linear transformations f from $\mathcal{N}$ (an n-dimensional vector space over $GF(q)$, the finite field of q elements) into $\mathcal{X}$ such that $f(\mathcal{N}) \cap \mathcal{Z} = \{0\}$ where $\mathcal{Z}$ is a subspace of $\mathcal{X}$. They have also shown that $P_n(X, Z)$ satisfies equation (1.2).

We propose to prove an equivalent form of (1.2) for the one variable polynomials $P_n(X, 1)$. Let us count the number of one-to-one linear transformations of $\mathcal{N}$ into $\mathcal{X} \oplus \mathcal{Y}$. Since $\mathcal{X} \oplus \mathcal{Y}$ has XY elements, there are clearly $P_n(XY, 1)$ such mappings. On the other hand, since every element of $\mathcal{X} \oplus \mathcal{Y}$ is of the form $\alpha + \beta$ where $\alpha \in \mathcal{X}$ and $\beta \in \mathcal{Y}$ (α is called the $\mathcal{X}$-component and β the $\mathcal{Y}$-component), let us look at mappings for which those elements of the image with 0 as $\mathcal{X}$-component form a j-dimensional subspace of $\mathscr{X} \oplus \mathscr{Y}$. To form such maps we may choose a j-dimensional subspace of $\mathscr{N}$ in $\binom{n}{j}_q$ ways (see [20], p. 139, [11], [12], [17]) and then map it into $\mathscr{Y}$ in $P_j(Y, 1)$ ways. Since a linear transformation is completely determined by the action on a basis, we choose a basis of $\mathscr{N}$, say $b_1, b_2, \ldots, b_n$ such that $b_1, \ldots, b_j$ is a basis of the above mentioned j-dimensional subspace. I claim now that $f(b_{j+1}), \ldots, f(b_n)$ need only be chosen so that their $\mathscr{X}$-components are linearly independent in $\mathscr{X}$. This follows from the fact that if $f(b_i) = \alpha_i + \beta_i$ and there exist $C_i \in GF(q)$ not all zero such that

$$\sum_{i=j+1}^{n} C_i \alpha_i = 0,$$

then

$$\sum_{i=j+1}^{n} C_i f(b_i) = \sum_{i=j+1}^{n} C_i(\alpha_i + \beta_i) = \sum_{i=j+1}^{n} C_i \beta_i.$$

Thus $\sum_{i=j+1}^{n} C_i f(b_i)$ has 0 as $\mathscr{X}$-component and so is in the space spanned by $f(b_1), \ldots, f(b_j)$. Hence there exist $C_1, \ldots, C_j$ in $GF(q)$ such that

$$-\sum_{i=j+1}^{n} C_i f(b_i) = \sum_{i=1}^{j} C_i f(b_i).$$

Therefore

$$f\left(\sum_{i=1}^{n} C_i b_i\right) = 0,$$

and since f is a one-to-one linear transformation

$$\sum_{i=1}^{n} C_i b_i = 0$$

which is impossible since not all the C_i are zero and the b_i form a basis for $\mathscr{N}$. Conversely if the $\mathscr{X}$-components of $f(b_{j+1}), \ldots, f(b_n)$ are linearly independent, then $f(b_1), \ldots, f(b_n)$ span an n-dimensional subspace of $\mathscr{X} \oplus \mathscr{Y}$ with a j-dimensional subspace having 0 as $\mathscr{X}$-component. Thus there are $P_{n-j}(X, 1)$ ways of choosing the $\mathscr{X}$-components of $f(b_{j+1}), \ldots, f(b_n)$ and Y^{n-j} ways of choosing the $\mathscr{Y}$-components.

Consequently the total number of one-to-one linear transformations of $\mathscr{N}$ into $\mathscr{X} \oplus \mathscr{Y}$ with j-dimensional image having 0 as $\mathscr{X}$-component is

$$\binom{n}{j}_q P_j(Y, 1) Y^{n-j} P_{n-j}(X, 1).$$

Hence summing over all j, we see that

$$P_n(XY, 1) = \sum_{j=0}^{n} \binom{n}{j}_q P_j(Y, 1) Y^{n-j} P_{n-j}(X, 1),$$

and replacing j by $n - j$, we obtain

$$P_n(XY, 1) = \sum_{j \geq 0} \binom{n}{j}_q P_j(X, 1) Y^j P_{n-j}(Y, 1). \tag{3.1}$$

We remark that (3.1) is equivalent to the q-binomial theorem of Goldman and Rota [12], p. 252, equation (1), by the substitutions $x \to XY$, $y \to Y$, $z \to 1$ (to reverse $X \to x/y$, $Y \to y/z$, then multiply (3.1) by z^n). However, our derivation here closely parallels the derivation of (1.1) in the special case $p_n(x) = x(x-1)\ldots(x-n+1)$. Thus we are led to the following definition:

Definition 1. We say that $p_0(X)$, $p_1(X)$, $p_2(X), \ldots$ is an *Eulerian family* of polynomials if

(i) $p_0(X) = 1$,
(ii) $p_n(X)$ is of degree n,
(iii) for each n,

$$p_n(XY) = \sum_{j \geq 0} \binom{n}{j}_q p_j(X) Y^j p_{n-j}(Y).$$

This definition is analogous to the definition of Rota and Mullin for polynomials of binomial type [18], p. 169.

The next section is devoted to constructing a theory of operators analogous to the delta operators of Rota and Mullin [18], p. 180, and the differential operators of Berge [4], p. 73.

4. Eulerian differential operators

The role of the shift operators of [18], p. 179, used in binomial enumeration is now played by the *Eulerian shift operator*:

$$\eta^a p(X) = p(Xq^a) = p(XA),$$

where $A = q^a$.

To be consistent with our notation for finite vector spaces, we shall write $X = q^x$, $Y = q^y$, $A = q^a$, $B = q^b$, and so on; this notational convention allows us to exhibit symmetries that might otherwise be hidden. Our polynomials will all lie in algebra (over the real numbers $\mathbf{R}$) of all polynomials of one variable $X = q^x$, to be denoted by $\mathbf{P}$.

Convergence questions here are trivial and will largely be ignored. Generally we shall treat series as formal power series, and q will denote a prime power; however, the results obtained in Section 9 may be treated as analytic results valid for $|q| < 1$, or as q-adic results valid for q prime.

Definition 2. An *Eulerian differential operator* τ is a linear operator on P that satisfies the following conditions:

$$q^{-a}\tau\eta^a = \eta^a\tau, \tag{4.1}$$

and

$$\tau X^n \not\equiv 0 \quad \text{for each} \quad n > 0. \tag{4.2}$$

The most well-known Eulerian differential operator is the q-differentiation[1] operator D_q:

$$D_q = \frac{1}{X}(1 - \eta).$$

Note that

$$\begin{aligned} D_qP_n(X, 1) &= X^{-1}\{(X - 1)(X - q)\dots(X - q^{n-1}) - (Xq - 1)(Xq - q)\dots \\ &\quad \times (Xq - q^{n-1})\} \\ &= X^{-1}P_{n-1}(X, 1)\{X - q^{n-1} - q^{n-1}(Xq - 1)\} \\ &= (1 - q^n)P_{n-1}(X, 1). \end{aligned}$$

LEMMA 1. *If τ is an Eulerian differential operator, then $\tau C = 0$ for each constant C.*

Proof: Since $q^{-a}\tau\eta^a = \eta^a\tau$, we see that $q^{-a}\tau C = \eta^a\tau C$. Let $\tau C = r(X) \in \mathbf{P}$. Then we have

$$q^{-a}r(X) = r(Xq^a).$$

If $r(X) \equiv 0$, this identity is obvious. If $r(X) \not\equiv 0$, let d be the leading coefficient of r and n the degree. Hence comparing coefficients of X^n in the above identity, we find that

$$q^{-a}\, d = q^{an}\, d.$$

But since $d \neq 0$ and $n \geq 0$, this equation is impossible. Hence $\tau C = 0$.

LEMMA 2. *If τ is an Eulerian differential operator, and $p(X)$ is any polynomial of degree n, then $\tau p(X)$ is of degree $n - 1$.*

Proof: By (4.1), for each n

$$q^{-a}\tau\eta^a X^n = \eta^a\tau X^n.$$

Hence

$$q^{(n-1)a}\tau X^n = \eta^a\tau X^n.$$

Suppose $\tau X^n = r(X) = eX^j + \dots$; then

$$q^{(n-1)a}r(X) = r(Xq^a).$$

Comparing coefficients of X^j in this equation, we see that

$$q^{(n-1)a}e = q^{aj}e.$$

Since $e \neq 0$ by (4.2), we see that $j = n - 1$. By linearity $\tau p(X)$ is of degree $n - 1$.

DEFINITION 3. Let τ be an Eulerian differential operator. A sequence of polynomials $p_0(X), p_1(X), p_2(X), \dots$ is called the sequence of *Eulerian basic polynomials* for τ if:

(i) $p_0(X) = 1$,

(ii) $p_n(1) = 0, \quad$ for each $n > 0$,

(iii) $\tau p_n(X) = (1 - q^n)p_{n-1}(X)$.

[1] F. H. Jackson [14] who introduced q-differentiation actually used $(1 - q)^{-1}D_q$; the operator δ of L. J. Rogers [19] is D_q.

It is clear by mathematical induction that each $p_n(X)$ is of degree n. By Lemma 2 it is clear that we can construct a unique sequence of Eulerian basic polynomials for each τ.

THEOREM 1

(a) *If $p_n(X)$ is an Eulerian basic sequence for some Eulerian differential operator, then it is an Eulerian family of polynomials.*

(b) *If $p_n(X)$ is an Eulerian family of polynomials, then it is an Eulerian basic sequence for some Eulerian differential operator.*

Proof: (a) Iterating property (iii) of Eulerian basic polynomials, we see that

$$\tau^k p_n(X) = (q^{n-k+1})_k p_{n-k}(X),$$

and hence by property (ii)

$$[\tau^k p_n(X)]_{X=1} = \begin{cases} (q)_n, & \text{if } k = n, \\ 0, & \text{if } k < n. \end{cases}$$

Thus

$$p_n(X) = \sum_{k \geq 0} \frac{p_k(X)}{(q)_k} [\tau^k p_n(X)]_{X=1}.$$

By linearity, we see that for each $p(X) \in \mathbf{P}$,

$$p(X) = \sum_{k \geq 0} \frac{p_k(X)}{(q)_k} [\tau^k p(X)]_{X=1}.$$

Now suppose $p(X)$ is the polynomial $p_n(XY)$. Thus

$$p_n(XY) = \sum_{k \geq 0} \frac{p_k(X)}{(q)_k} [\tau^k p_n(XY)]_{X=1}$$

But since $Y = q^y$,

$$\begin{aligned} [\tau^k p_n(XY)]_{X=1} &= [\tau^k \eta^y p_n(X)]_{X=1} \\ &= [q^{ky} \eta^y \tau^k p_n(X)]_{X=1} \\ &= [q^{ky} \eta^y (q^{n-k+1})_k p_{n-k}(X)]_{X=1} \\ &= [Y^k (q^{n-k+1})_k p_{n-k}(XY)]_{X=1} \\ &= Y^k (q^{n-k+1})_k p_{n-k}(Y). \end{aligned}$$

Hence

$$\begin{aligned} p_n(XY) &= \sum_{k \geq 0} \frac{(q^{n-k+1})_k}{(q)_k} p_k(X) Y^k p_{n-k}(Y) \\ &= \sum_{k \geq 0} \binom{n}{k}_q p_k(X) Y^k p_{n-k}(Y). \end{aligned}$$

(b) Conversely suppose $p_n(X)$ is an Eulerian family of polynomials. Putting $Y = 1$ in equation (iii) of Definition 1, we see that

$$p_n(X) = \sum_{j \geq 0} \binom{n}{j}_q p_j(X) p_{n-j}(1).$$

Since this identity is valid for each $n \geq 0$ and since each $p_n(X)$ is of degree n, we see that $p_0(1) = 1$ and that $p_n(1) = 0$ for each $n > 0$. Since $p_0(X)$ is a constant, $p_0(X) = 1$. Thus properties (i) and (ii) of Definition 3 are fulfilled.

Let us now define a linear operator τ on $\mathbf{P}$ by

$$\tau p_0(X) = 0,$$
$$\tau p_n(X) = (1 - q^n)p_{n-1}(X), \qquad \text{for each } n \geq 1.$$

We need only verify that $q^{-y}\tau\eta^y = \eta^y\tau$, where $Y = q^y$.

Clearly if we replace j by $n - k$ in property (iii) of Definition 1, we see that

$$p_n(XY) = \sum_{k \geq 0} \frac{p_k(Y)}{(q)_k} Y^{n-k}\tau^k p_n(X). \tag{4.3}$$

Operating on both sides of (4.3) with τ, we first find that

$$\tau(p_n(XY)) = \tau\eta^y p_n(X),$$

while on the right hand side we have

$$\begin{aligned}
&\sum_{k \geq 0} \frac{p_k(Y)}{(q)_k} Y^{n-k}\tau^{k+1} p_n(X)\\
&= \sum_{k \geq 0} \frac{p_k(Y)}{(q)_k}(1 - q^n)\dots(1 - q^{n-k})Y^{n-k}p_{n-k-1}(X)\\
&= Y(1 - q^n)\sum_{k \geq 0} \binom{n-1}{k}_q p_k(Y)Y^{n-1-k}p_{n-1-k}(X)\\
&= Y(1 - q^n)p_{n-1}(XY)\\
&= q^y\eta^y\tau p_n(X).
\end{aligned}$$

Since the $p_n(X)$ form a basis for $\mathbf{P}$, we see by linearity that

$$\tau\eta^y = q^y\eta^y\tau$$

which is equivalent to (4.1)

5. Expansion theorems

Definition 4. If σ is a linear operator on P, we shall say that σ is an *Eulerian shift-invariant operator* if:

$$\sigma\eta^y = \eta^y\sigma$$

for all Y (recall $Y = q^y$).

Theorem 2. (Eulerian expansion theorem). *Let σ be an Eulerian shift-invariant operator, and let τ be an Eulerian differential operator with associated Eulerian family $p_n(X)$. Then*

$$\sigma = \sum_{k \geq 0} \frac{a_k}{(q)_k} X^k\tau^k,$$

where $a_k = [\sigma p_k(X)]_{X=1}$.

Proof: Since the $p_n(X)$ form the Eulerian family associated with τ, we rewrite property (iii) of Definition 1 as

$$p_n(XY) = \sum_{k \geq 0} \frac{p_k(Y)}{(q)_k} X^k \tau^k p_n(X).$$

We now apply σ as an operator on polynomials in Y to the above equation. Thus

$$\sigma\eta^x p_n(Y) = \sum_{k \geq 0} \frac{\sigma p_k(Y)}{(q)_k} X^k \tau^k p_n(X).$$

By linearity, we can extend this identity to all elements of **P**. Thus we see that

$$\sigma\eta^x p(Y) = \sum_{k \geq 0} \frac{\sigma p_k(Y)}{(q)_k} X^k \tau^k p(X).$$

Since $\sigma\eta^x = \eta^x\sigma$, we see that

$$\eta^x \sigma p(Y) = \sum_{k \geq 0} \frac{\sigma p_k(Y)}{(q)_k} X^k \tau^k p(X).$$

Consequently

$$(\sigma p)(XY) = \sum_{k \geq 0} \frac{\sigma p_k(Y)}{(q)_k} X^k \tau^k p(X).$$

Setting $Y = 1$, we obtain Theorem 2.

So far our theory is a perfect q-analog of the results of Rota and Mullin [18]. However, as is well-known in classical analysis, the q-analogs of ordinary hypergeometric series are not a mirror image of the ordinary theory. The following result exhibits the beginning of the divergence of these theories as the ring structure Rota and Mullin obtained for their operators is replaced in the analog by an additive group structure.

THEOREM 3. *Let τ be an Eulerian differential operator, and let E be the* additive group *of formal Eulerian series over* **R**. *Then there exists an isomorphism from E onto the* additive group *T of Eulerian shift invariant operators which carries*

$$f(t) = \sum_{k \geq 0} \frac{a_k t^k}{(q)_k} \quad \text{into} \quad \sum_{k \geq 0} \frac{a_k X^k \tau^k}{(q)_k}.$$

Proof: The mapping is already linear, and by Theorem 2 it is onto.

We remark that the factor X^k is what prohibits our obtaining the ring structure of formal Eulerian series. At this point we find that the corollaries that Rota and Mullin [18; p. 189, Cor. 1 and 2] easily derived from their strong Theorem 3 are not corollaries of our Theorem 3.

We shall now prove a strengthened form of Lemma 2 that will be important in future developments.

THEOREM 4. *Let τ be an Eulerian differential operator, then there exist constants $e_0 = 0, e_1, e_2, \ldots$ where $e_n \neq 0$ for each $n > 0$ such that*

$$\tau X^n = e_n X^{n-1}.$$

Conversely for any sequence of constants $e_0 = 0, e_1, e_2, \ldots$ where $e_n \neq 0$ for each $n > 0$, the linear operator τ on **P** defined by $\tau X^n = e_n X^{n-1}$ is an Eulerian differential operator.

Proof: First we assume τ is an Eulerian differential operator. Let $\tau X^n = s_n(X)$. Now

$Ys_n(XY) = Y\eta^y s_n(X) = Y\eta^y \tau X^n = \tau\eta^y X^n = Y^n \tau X^n = Y^n s_n(X)$. Setting $X = 1$, we see that

$$s_n(Y) = s_n(1)Y^{n-1}.$$

Since $\tau 1 \equiv 0$, and $\tau X^n \not\equiv 0$ for each $n > 0$, we see that $s_0(1) = 0$ and $s_n(1) \neq 0$ for each $n > 0$. The first half of the theorem now follows with $e_n = s_n(1)$.

Conversely we consider τ defined by $\tau X^n = e_n X^{n-1}$. By linearity, we see that τ is well-defined on **P**. Furthermore

$$Y\eta^y \tau X^n = Y\eta^y e_n X^{n-1} = Y e_n Y^{n-1} X^{n-1} = e_n Y^n X^{n-1} = \tau\eta^y X^n,$$

and since the X^n form a basis for **P**, we see that in general $\eta^y \tau = Y^{-1}\tau\eta^y$. Thus since also $\tau 1 \equiv 0$ and $\tau X^n \not\equiv 0$ for each $n > 0$, we see that τ is an Eulerian differential operator.

COROLLARY. *Let τ be an Eulerian differential operator with related Eulerian family of polynomials $p_n(X)$. Let C_n be the leading coefficient of $p_n(X)$, and let e_n be defined by $\tau X^n = e_n X^{n-1}$. Then*

$$e_n = \frac{(1 - q^n)C_{n-1}}{C_n}, \qquad e_0 = 0.$$

Proof: By Theorem 4 we know that $e_0 = 0$. Now

$$\begin{aligned} C_n e_n X^{n-1} + \cdots &= \tau p_n(X) \\ &= (1 - q^n)p_{n-1}(X) \\ &= (1 - q^n)C_{n-1}X^{n-1} + \cdots. \end{aligned}$$

Comparing coefficients of X^{n-1}, we obtain the desired result.

Theorem 4 and its corollary give us much information about Eulerian differential operators and Eulerian families. To obtain further information (especially an analog to Corollary 2 of Theorem 3 in [18]), we move to a full-fledged study of the relevant generating functions.

6. Generating functions

DEFINITION 5. We say that $p_0(X, Z)$, $p_1(X, Z)$, $p_2(X, Z), \ldots$ is a *homogeneous Eulerian family* of polynomials if each $p_n(X, Z)$ is a homogeneous polynomial of degree n in X and Z such that

(i) $$p_0(X, Z) = 1,$$

(ii) $$p_n(X, 0) \not\equiv 0,$$

(iii) $$p_n(X, Z) = \sum_{j \geq 0} \binom{n}{j}_q p_j(X, Y)p_{n-j}(Y, Z).$$

The relationship between homogeneous Eulerian families and ordinary Eulerian families is explicitly described in the following theorem.

THEOREM 5. *There is a one-to-one correspondence ϕ between homogeneous Eulerian families and ordinary Eulerian families given by*

$$\phi : p_n(X, Z) \to p_n(X, 1),$$
$$\phi^{-1} : p_n(X) \to Z^n p_n(X/Z) = p_n(X, Z).$$

Proof: Suppose the $p_n(X, Z)$ form a homogeneous Eulerian family. Then $p_n(X, 0) = CX^n \not\equiv 0$ by property (ii) of Definition 5. Hence $p_n(X, 1)$ is of degree n. By property (i) of Definition 5, $p_0(X, 1) = 1$. Next by homogeneity and property (iii) of Definition 5, we see that

$$\begin{aligned} p_n(XY, 1) &= \sum_{j \geq 0} \binom{n}{j}_q p_j(XY, Y) p_{n-j}(Y, 1) \\ &= \sum_{j \geq 0} \binom{n}{j}_q p_j(X, 1) Y^j p_{n-j}(Y, 1), \end{aligned}$$

which shows that the $p_n(X, 1)$ form an Eulerian family.

Conversely suppose that the $p_n(X)$ form an Eulerian family. Let

$$p_n(X) = \sum_{j \geq 0}^{n} C_{nj} X^j,$$

where $C_{nn} \neq 0$; then

$$\begin{aligned} p_n(X, Z) &= Z^n p_n(X/Z) \\ &= \sum_{j=0}^{n} C_{nj} X^j Z^{n-j}. \end{aligned}$$

Clearly $p_0(X, Z) = 1$, and

$$p_n(X, 0) = C_{nn} X^n \not\equiv 0.$$

Finally

$$\begin{aligned} p_n(X, Z) &= Z^n p_n(X/Z) = Z^n p_n\left(\frac{X}{Y} \cdot \frac{Y}{Z}\right) \\ &= Z^n \sum_{j \geq 0} \binom{n}{j}_q p_j\left(\frac{X}{Y}\right) \left(\frac{Y}{Z}\right)^j p_{n-j}\left(\frac{Y}{Z}\right) \\ &= \sum_{j \geq 0} \binom{n}{j}_q Y^j p_j\left(\frac{X}{Y}\right) Z^{n-j} p_{n-j}\left(\frac{Y}{Z}\right) \\ &= \sum_{j \geq 0} \binom{n}{j}_q p_j(X, Y) p_{n-j}(Y, Z). \end{aligned}$$

Noting that $\phi^{-1}\phi p_n(X, Z) = \phi^{-1} p_n(X, 1) = Z^n p_n(X/Z, 1) = p_n(X, Z)$, and $\phi\phi^{-1} p_n(X) = \phi Z^n p_n(X/Z) = p_n(X)$, we see that Theorem 5 is established.

The value of Theorem 5 lies in the fact that we may easily determine the form of the generating functions for the homogeneous Eulerian families (see Theorem 6).

We then set $Z = 1$ to determine the generating function for ordinary Eulerian families.

THEOREM 6. *Let $p_n(X, Z)$ be a homogeneous Eulerian family with $C_n = p_n(1, 0) \neq 0$, and let*

$$f(t) = \sum_{k \geq 0} \frac{C_k t^k}{(q)_k}.$$

Then

$$\sum_{n \geq 0} \frac{p_n(X, Z)t^n}{(q)_n} = \frac{f(Xt)}{f(Zt)}.$$

Proof: Let

$$F(X, Z; t) = \sum_{n \geq 0} \frac{p_n(X, Z)}{(q)_n} t^n$$

Then

$$\begin{aligned} F(X, Z; t) &= \sum_{n=0}^{\infty} \frac{t^n}{(q)_n} \sum_{\substack{j+k=n \\ j \geq 0, k \geq 0}} \frac{(q)_n}{(q)_j (q)_k} p_j(X, Y) p_k(Y, Z) \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{p_j(X, Y)t^j}{(q)_j} \cdot \frac{p_k(X, y)t^k}{(q)_k} \\ &= F(X, Y; t)F(Y, Z; t). \end{aligned}$$

Replace Z by 0 and then replace Y by Z; this yields

$$F(X, 0; t) = F(X, Z; t)F(Z, 0; t).$$

Hence

$$\begin{aligned} \sum_{n \geq 0} \frac{p_n(X, Z)t^n}{(q)_n} &= F(X, Z; t) \\ &= \frac{F(X, 0; t)}{F(Z, 0; t)} \\ &= \frac{\sum_{n \geq 0} \frac{C_n X^n t^n}{(q)_n}}{\sum_{n \geq 0} \frac{C_n Z^n t^n}{(q)_n}} \\ &= \frac{f(Xt)}{f(Zt)}. \end{aligned}$$

COROLLARY. *If $p_n(X)$ is an Eulerian family of polynomials, and if C_n is the leading coefficient of $p_n(X)$, then*

$$\sum_{n \geq 0} \frac{p_n(X)t^n}{(q)_n} = \frac{f(Xt)}{f(t)},$$

where

$$f(t) = \sum_{n\geq 0} \frac{C_n t^n}{(q)_n}.$$

Proof: In Theorem 6, let $p_n(X, Z) = Z^n p_n(X/Z)$. Then set $Z = 1$.

THEOREM 7. *If $C_0 = 1, C_1, C_2, \ldots$ is any sequence of non-zero real numbers, and if*

$$f(t) = \sum_{k\geq 0} \frac{C_k t^k}{(q)_k},$$

then the expressions $p_n(X, Z)$ defined by

$$\sum_{n\geq 0} \frac{p_n(X, Z)t^n}{(q)_n} = \frac{f(Xt)}{f(Zt)}$$

form a homogeneous Eulerian family of polynomials.

Proof: The argument required here merely reverses the steps in Theorem 6 so we omit it.

Finally we derive a further result for the generating functions which greatly resembles Corollary 2 of Theorem 3 in [18].

THEOREM 8. *Let the $p_n(X)$ form an Eulerian family of polynomials with C_n as the leading coefficient and let*

$$f(t) = \sum_{n\geq 0} \frac{C_n t^n}{(q)_n}.$$

Then

$$f(t) = \exp\left\{\sum_{n\geq 1} \frac{p_n'(1)t^n}{(q)_n n}\right\}.$$

Proof: By Theorem 6, if

$$F(X, Z; t) = \frac{f(Xt)}{f(Zt)},$$

then

$$F(X, Z; t) = F(X, Y; t)F(Y, Z; t)$$

Therefore replacing X by XY, then Y by X and Z by 1, we see that

$$\begin{aligned} F(XY, 1; t) &= F(XY, X; t)F(X, 1; t) \\ &= F(Y, 1; Xt)F(X, 1; t). \end{aligned}$$

Hence

$$\begin{aligned} &\frac{F(XY, 1; t) - F(X, 1; t)}{Y - 1} \\ &= F(X, 1; t)\left\{\frac{F(Y, 1; Xt) - 1}{Y - 1}\right\} \\ &= F(X, 1; t) \sum_{n\geq 1} \frac{p_n(Y)}{Y - 1} \frac{X^n t^n}{(q)_n}. \end{aligned}$$

Letting $Y \to 1$ (or $y \to \infty$ where $Y = q^y$), we obtain

$$X\frac{d}{dX}F(X,1;t) = F(X,1;t)\sum_{n\geq 1}\frac{p_n'(1)X^n t^n}{(q)_n};$$

this follows from the fact that, by Theorem 1, $p_n(1) = 0$ for each $n > 0$.

Thus $F(X,1;t)$ satisfies a first order differential equation in X, namely

$$Xy' - y\sum_{n\geq 1}\frac{p_n'(1)X^n t^n}{(q)_n} = 0.$$

The solutions of this equation are of the form

$$y = K(t)\exp\left\{\sum_{n\geq 1}\frac{p_n'(1)X^n t^n}{(q)_n n}\right\}$$

Since $F(0,1;t) = (f(t))^{-1} = K(t)$, we see that

$$\frac{f(Xt)}{f(t)} = F(X,1;t) = (f(t))^{-1}\exp\left\{\sum_{n\geq 1}\frac{p'_n(1)X^n t^n}{(q)_n n}\right\}.$$

Hence

$$f(Xt) = \exp\left\{\sum_{n\geq 1}\frac{p_n'(1)X^n t^n}{(q)_n n}\right\},$$

and this formula is clearly equivalent to the result stated in Theorem 8.

7. Further expansion theorems

In this section we shall be primarily interested in the relationship between expansions of Eulerian differential operators τ and the generating function obtained from the Eulerian family related to τ.

First we observe that any Eulerian differential operator has an expansion in terms of the q-derivative $D_q = (1/X)(1 - \eta)$.

THEOREM 9. *Let τ be any Eulerian differential operator, then*

$$\tau = \frac{1}{X}\sum_{n\geq 0}\frac{a_n X^n D_q^n}{(q)_n},$$

where

$$a_n = [\tau(X-1)(X-q)\dots(X-q^{n-1})]_{X=1}.$$

Proof: First we note that $X\tau$ is Eulerian shift-invariant. This follows from the fact that

$$\eta^y(X\tau) = XY\eta^y\tau = XYY^{-1}\tau\eta^y = (X\tau)\eta^y.$$

Hence if $\sigma = X\tau$, then by Theorem 2

$$\sigma = \sum_{n\geq 0}\frac{a_n X^n D_q^n}{(q)_n}$$

where $a_n = [\sigma p_n(X)]_{X=1}$ and where $p_n(X)$ is the Eulerian family associated with

D_q. As we observed just after Definition 2, $p_n(X) = (X - 1)\dots(X - q^{n-1})$. Thus Theorem 9 follows from the fact that $\sigma = X\tau$.

Most of the Eulerian differential operators we shall meet are expressed in terms of η rather than D_q. The following theorem relates such operators to their respective generating functions.

THEOREM 10. *Suppose τ is an Eulerian differential operator that has a Laurent series expansion in η of the form*

$$\frac{1}{X}\sum_{n=-B}^{\infty} b_n\eta^n = \frac{1}{X}L(\eta).$$

Let $p_n(X)$ be the associated Eulerian family of polynomials with C_n the leading coefficient of $p_n(X)$. Then

$$C_n = \frac{(q)_n}{\prod_{j=1}^{n} L(q^j)}.$$

Proof: We observe that

$$\tau X^m = \frac{1}{X}L(\eta)X^m = \frac{1}{X}\sum_{n=-B}^{\infty} b_n q^{mn}X^m = X^{m-1}L(q^m).$$

Thus in the notation of Theorem 4,

$$e_n = L(q^n),$$

and by the Corollary of Theorem 4

$$C_n = \frac{(1-q^n)}{L(q^n)}C_{n-1}.$$

Iterating this equation and recalling that $C_0 = 1$, we see that

$$C_n = \frac{(q)_m}{\prod_{j=1}^{m} L(q^j)}.$$

We shall now examine some further results that are related to the symbolic method utilized by Goldman and Rota in [11].

THEOREM 11. *Let τ be an Eulerian differential operator with associated Eulerian family $p_n(X)$. Let*

$$f(t) = \sum_{n\geq 0}\frac{C_n t^n}{(q)_n}$$

where C_n is the leading coefficient of $p_n(X)$. Suppose that

$$g(X,t) = \sum_{n\geq 0}\frac{\pi_n(X)t^n}{(q)_n},$$

where the $\pi_n(X)$ are polynomials in X, $\pi_0(X) = 1$, $\pi_n(1) = 0$ for each $n > 0$, and

$$\tau g(X,t) = tg(X,t).$$

Then

$$g(X, t) = f(Xt)/f(t).$$

Proof: We observe that

$$\tau g(X, t) = \sum_{n \geq 0} \frac{[\tau \pi_n(X)]t^n}{(q)_n}.$$

By by hypothesis

$$\begin{aligned} \tau g(X, t) &= tg(X, t) \\ &= \sum_{n \geq 0} \frac{\pi_n(X)t^{n+1}}{(q)_n} \\ &= \sum_{n \geq 0} \frac{(1 - q^n)\pi_{n-1}(X)t^n}{(q)_n}. \end{aligned}$$

By comparing coefficients of $t^n/(q)_n$ in our two series for $\tau g(X, t)$, we see that

$$\tau \pi_n(X) = (1 - q^n)\pi_{n-1}(X), \pi_n(1) = 0 \qquad \text{for each } n > 0,$$

and

$\pi_0(X) = 1.$

However the only family of polynomials satisfying these conditions is $p_n(X)$. Thus

$$p_n(X) = \pi_n(X).$$

Therefore

$$g(X, t) = \sum_{n=0} \frac{p_n(X)t^n}{(q)_n} = f(Xt)/f(t)$$

as asserted.

COROLLARY. *Let τ, $p_n(X)$, and $f(t)$ be defined as in Theorem* 11. *Suppose that*

$$h(t) = \sum_{n \geq 0} \frac{d_n t^n}{(q)_n}, \qquad (d_0 = 1),$$

and

$$\tau h(Xt) = th(Xt).$$

Then

$$h(t) = f(t).$$

Proof: Define

$$g(X, t) = h(Xt)/h(t).$$

Then $g(X, t)$ fulfills the conditions of Theorem 11. Therefore

$$\frac{h(Xt)}{h(t)} = g(X, t) = \frac{f(Xt)}{f(t)}.$$

Setting $X = 0$, we see that $h(t) = f(t)$.

8. Eulerian Sheffer polynomials

In [15], Rota and Kahaner extend the work in [18] to Sheffer polynomials. Let us recall that a Sheffer set relative to the delta operator Q is a sequence of polynomials $s_0(x), s_1(x), s_2(x), \ldots$ such that

$$Qs_n(x) = ns_{n-1}(x),$$

and

$$s_0(x) = 1.$$

It is then possible to prove that

$$\sum_{n\geq 0} \frac{s_n(x)t^n}{n!} = h(t)\exp\left\{x \sum_{n\geq 0} \frac{p_n'(0)t^n}{n!}\right\}$$

where $p_n(x)$ is the basic polynomial set associated with Q and $h(t)$ is a formal power series in t with $h(0) = 1$. Conversely one can show that any family of polynomials $s_n(x)$ defined by a function of the above form is a Sheffer set relative to Q. The Eulerian analogs of these facts will be important in Section 10, and so we develop them now.

DEFINITION 6. Let τ be an Eulerian differential operator. A sequence of polynomials $s_0(X), s_1(X), s_2(X), \ldots$ is called an *Eulerian Sheffer family* relative to τ if:

(i) $$s_0(X) = 1$$

(ii) $$\tau s_n(X) = (1 - q^n)s_{n-1}(X).$$

THEOREM 12. *Let τ be an Eulerian differential operator with $p_n(X)$ the associated Eulerian family. If $s_n(X)$ is an Eulerian Sheffer family relative to τ, then*

$$s_n(XY) = \sum_{j\geq 0} \binom{n}{q}_q s_j(X)X^{n-j}p_{n-j}(Y) \tag{8.1}$$

for each n. Conversely any family of polynomials satisfying (8.1) *with* $s_0(X) = 1$ *is an Eulerian Sheffer family relative to τ.*

Proof: Suppose first that the $s_n(X)$ form an Eulerian Sheffer family relative to τ. Thus if

$$S(X;t) = \sum_{n\geq 0} \frac{s_n(X)t^n}{(q)_n},$$

we see that

$$\tau\frac{S(X;t)}{S(1;t)} = \sum_{n\geq 0} \frac{(1-q^n)s_{n-1}(X)t^n}{(q)_n}(S(1,t))^{-1}$$
$$= tS(X;t)/S(1.\,t).$$

Hence by Theorem 11,

$$\frac{S(X,t)}{S(1,t)} = \sum_{n\geq 0} \frac{p_n(X)t^n}{(q)_n}.$$

Therefore

$$\sum_{n\geq 0}\frac{s_n(X)t^n}{(q)_n}=\sum_{n\geq 0}\frac{s_n(1)t^n}{(q)_n}\sum_{m\geq 0}\frac{p_n(X)t^m}{(q)_m}.$$

Comparing coefficients of t^n on each side of this equation, we see that

$$s_n(X)=\sum_{j\geq 0}\binom{n}{j}_q p_j(X)s_{n-j}(1).$$

Therefore

$$\begin{aligned}
s_n(XY)&=\sum_{j=0}^{n}\binom{n}{j}_q\sum_{r=0}^{j}\binom{j}{r}p_r(X)Y^r p_{j-r}(Y)s_{n-j}(1)\\
&=\sum_{r=0}^{n}\frac{(q)_n}{(q)_r}p_r(X)Y^r\sum_{j=r}^{n}\frac{1}{(q)_{n-j}(q)_{j-r}}p_{j-r}(Y)s_{n-j}(1)\\
&=\sum_{r=0}^{n}\binom{n}{r}_q p_r(X)Y^r\sum_{j=0}^{n-r}\binom{n-r}{j}_q p_j(Y)s_{n-r-j}(1)\\
&=\sum_{r\geq 0}\binom{n}{r}_q p_r(X)Y^r s_{n-r}(Y).
\end{aligned}$$

Conversely suppose that $s_0(X)=1$ and the $s_n(X)$ satisfy (8.1). Then by setting $X=1$ in (8.1) and then replacing Y by X, we see that

$$\begin{aligned}
\tau s_n(X)&=\sum_{j\geq 0}\binom{n}{j}_q \tau p_j(X)s_{n-j}(1)\\
&=\sum_{j\geq 0}\binom{n}{j}_q (1-q^j)p_{j-1}(X)s_{n-j}(1)\\
&=(1-q^n)\sum_{j\geq 1}\binom{n-1}{j-1}_q p_{j-1}(X)s_{n-j}(1)\\
&=(1-q^n)\sum_{j\geq 0}\binom{n-1}{j}_q p_j(X)s_{n-1-j}(1)\\
&=(1-q^n)s_{n-1}(X).
\end{aligned}$$

COROLLARY. *If $s_n(X)$ is an Eulerian Sheffer family relative to τ, then*

$$\sum_{n\geq 0}\frac{s_n(X)t^n}{(q)_n}=h(t)\sum_{n\geq 0}\frac{p_n(X)t^n}{(q)_n},$$

where $h(t)$ is a formal Eulerian series with $h(0)=1$. Conversely any family of polynomials defined by the above type of function is an Eulerian Sheffer family relative to τ.

Proof: The first part follows directly from the first part of the proof of Theorem 12. On the other hand, suppose that $s_n(X)$ is defined by the above equation. Then

$$s_0(X)=h(0)p_0(X)=1,$$

and

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{\tau s_n(X)t^n}{(q)_n} &= h(t) \sum_{n\geq 0} \frac{\tau p_n(X)t^n}{(q)_n}\\ &= h(t) \sum_{n\geq 0} \frac{(1-q^n)p_{n-1}(X)t^n}{(q)_n}\\ &= th(t) \sum_{n\geq 0} \frac{p_n(X)t^n}{(q)_n}\\ &= t \sum_{n\geq 0} \frac{s_n(X)t^n}{(q)_n}\\ &= \sum_{n\geq 0} \frac{(1-q^n)s_{n-1}(X)t^n}{(q)_n}\end{aligned}$$

Comparing coefficients of t^n, we see that

$$\tau s_n(X) = (1 - q^n)s_{n-1}(X).$$

Therefore the $s_n(X)$ from an Eulerian Sheffer family of polynomials relative to τ.

There are at least two examples of Eulerian Sheffer polynomials that have been studied extensively. First we consider

$$H_n(X) = \sum_{j\geq 0} \binom{n}{j}_q X^j,$$

the q-Hermite polynomials studied by Carlitz [5], [6] and introduced independently by Szegö [22] and Rogers [19].

$$\begin{aligned}H_0(X) &\doteq 1,\\ D_q H_n(X) &= \sum_{j\geq 0} \binom{n}{j}_q (1 - q^j)X^{j-1}\\ &= (1-q^n) \sum_{j\geq 0} \binom{n-1}{j}_q X^j\\ &= (1-q^n)H_{n-1}(X).\end{aligned}$$

Thus the $H_n(X)$ form an Eulerian Sheffer family relative to D_q. Carlitz [5] also has considered a related set of polynomials

$$q^{n(n-1)/2}G_n(-X) = q^{n(n-1)/2} \sum_{j\geq 0} \binom{n}{j}_q q^{j(j-n)}(-X)^j.$$

Now

$$q^{0(0-1)/2}G_0(-X) = 1,$$

and with $\Delta q = q/X(1 - \eta^{-1})$

$$\begin{aligned}\Delta_q q^{n(n-1)/2} G_n(-X) &= q^{n(n-1)/2} \sum_{j \geq 0} \binom{n}{j}_q q^{j(j-n)}(-1)^{j-1} X^{j-1} q^{-j+1}(1 - q^j) \\ &= q^{n(n-1)/2} (1 - q^n) \sum_{j \geq 0} \binom{n-1}{j-1}_q q^{j(j-n)-j+1}(-X)^{j-1} \\ &= (1 - q^n) q^{(n-1)(n-2)/2} \sum_{j \geq 0} \binom{n-1}{j}_q q^{j(j-n+1)}(-X)^{j-1} \\ &= (1 - q^n) q^{(n-1)(n-2)/2} G_n(-X).\end{aligned}$$

Therefore $q^{n(n-1)/2} G_n(-X)$ is an Eulerian Sheffer family relative to Δ_q.

9. Applications to basic hypergeometric series

9.1 q-Differentiation. We have already discussed $D_q = 1/X(1 - \eta)$ with related Eulerian family $P_n(X, 1) = (X - 1)\ldots(X - q^{n-1})$. Since the leading coefficient of $P_n(X, 1)$ is always 1, we see that by the Corollary to Theorem 6

$$\sum_{n \geq 0} \frac{P_n(X, 1)t^n}{(q)_n} = e(Xt)/e(t), \tag{9.1}$$

where

$$e(t) = \sum_{n \geq 0} \frac{t^n}{(q)_n}.$$

Since $[P_n'(X, 1)]_{X=1} = \lim_{X \to 1} \dfrac{P_n(X, 1)}{(X - 1)} = (q)_{n-1}$, we see that by Theorem 8

$$\begin{aligned}\sum_{n \geq 0} \frac{t^n}{(q)_n} &= e(t) \\ &= \exp\left\{\sum_{n \geq 1} \frac{(q)_{n-1} t^n}{(q)_n n}\right\} \\ &= \exp\left\{\sum_{n \geq 1} \frac{t^n}{(1 - q^n)n}\right\} \\ &= \exp\left\{-\sum_{n \geq 1} \sum_{m \geq 0} \frac{t^n q^{nm}}{h}\right\} \\ &= \exp\left\{-\sum_{m \geq 0} \log(1 - tq^m\right\} \\ &= \prod_{m \geq 0} (1 - tq^m)^{-1} = (t)_\infty^{-1},\end{aligned}$$

a well-known result due to Euler.

Equation (9.1) may now be rewritten as

$$\sum_{n \geq 0} \frac{P_n(X, 1)t^n}{(q)_n} = (t)_\infty/(tX)_\infty, \tag{9.2}$$

the well-known summation due to Heine [21], p. 92, equation (3.2.2.12).

9.2 Backwards q-differentiation. Here we consider the Eulerian differential operator

$$\Delta_q = \frac{q}{X}(1 - \eta^{-1}).$$

$$\begin{aligned}\Delta_q P_n(1, X) &= q\frac{P_n(1, X) - P_n(1, Xq^{-1})}{X}\\ &= qP_{n-1}(1, X)\left\{\frac{1 - Xq^{n-1} - 1 + Xq^{-1}}{X}\right\}\\ &= P_{n-1}(1, X)(1 - q^n).\end{aligned}$$

A repetition of the arguments used in Section 9.1 would yield

$$\sum_{n\geq 0} \frac{P_n(1, X)t^n}{(q)_n} = (Xt)_\infty/(t)_\infty, \tag{9.3}$$

a result equivalent to (9.2).

9.3 The Heine–Gauss theorem. We now examine the Eulerian differential operator

$$\gamma = \frac{1}{1 - b\eta}D_q = \frac{1}{X}\frac{1 - \eta}{1 - b\eta q^{-1}}$$

with associated Eulerian family $g_n(X)$ and generating function $G(t)$, that is

$$\sum_{n\geq 0} \frac{g_n(X)t^n}{(q)_n} = G(Xt)/G(t).$$

Now by Theorem 10

$$\begin{aligned}G(t) &= \sum_{n\geq 0} \frac{t^n}{(q)_n} \frac{(q)_n}{\prod_{j=1}^{n} \frac{(1 - q^j)}{(1 - bq^{j-1})}}\\ &= \sum_{n\geq 0} \frac{(b)_n t^n}{(q)_n}.\end{aligned}$$

Hence

$$G(t) = \frac{(bt)_\infty}{(t)_\infty},$$

by (9.3).

Therefore

$$\sum_{n\geq 0} \frac{g_n(X)t^n}{(q)_n} = \frac{(bXt)_\infty(t)_\infty}{(Xt)_\infty(bt)_\infty}. \tag{9.4}$$

Now let us expand (9.4) in the following manner

$$\sum_{n\geq 0} \frac{g_n(X)t^n}{(q)_n} = \sum_{n\geq 0} \frac{C_n(t)P_n(X, 1)}{(q)_n}. \tag{9.5}$$

That such a formal expansion exists is obvious from the fact that $P_n(X, 1)$ forms a basis for $\mathbf{P}$ over $\mathbf{R}$. We wish to determine $C_n(t)$.

Since

$$\gamma g_n(X) = (1 - q^n)g_{n-1}(X),$$

we see that

$$\begin{aligned} D_q g_n(X) &= (1 - b\eta)(1 - q^n)g_{n-1}(X) \\ &= (1 - q^n)(g_{n-1}(X) - bg_{n-1}(Xq)). \end{aligned}$$

Hence

$$\begin{aligned} \sum_{n\geq 0} \frac{C_{n+1}(t)P_n(X, 1)}{(q)_n} &= \sum_{n\geq 0} \frac{C_n(t)(1 - q^n)P_{n-1}(X)}{(q)_n} \\ &= \sum_{n\geq 0} \frac{C_n(t)D_q P_n(X)}{(q)_n} \\ &= D_q \sum_{n\geq 0} \frac{g_n(X)t^n}{(q)_n} \\ &= \sum_{n\geq 1} \frac{(g_{n-1}(X) - bg_{n-1}(Xq))t^n}{(q)_{n-1}} \\ &= t \sum_{n\geq 0} \frac{(g_n(X) - bg_n(Xq))t^n}{(q)_n} \\ &= t \sum_{n\geq 0} \frac{C_n(t)P_n(X, 1)}{(q)_n} - bt \sum_{n\geq 0} \frac{C_n(t)P_n(Xq, 1)}{(q)_n}. \end{aligned} \tag{9.6}$$

Now

$$\begin{aligned} P_n(Xq, 1) &= q^n\left(X - \frac{1}{q}\right)P_{n-1}(X, 1) \\ &= q^n(X - q^{n-1})P_{n-1}(X, 1) + (q^{2n-1} - q^{n-1})P_{n-1}(X, 1) \\ &= q^n P_n(X, 1) - q^{n-1}(1 - q^n)P_{n-1}(X, 1) \end{aligned}$$

Substituting this identity into (9.6), we see that

$$\begin{aligned} &\sum_{n\geq 0} \frac{C_{n+1}(t)P_n(X, 1)}{(q)_n} \\ &\quad = \sum_{n\geq 0} \frac{tC_n(t)(1 - bq^n)P_n(X, 1) + btq^{n-1}(1 - q^n)C_n(t)P_{n-1}(X, 1)}{(q)_n}. \end{aligned}$$

Comparing coefficients of $P_n(X, 1)/(q)_n$ on both sides of this equation we see that

$$C_{n+1}(t) = t(1 - bq^n)C_n(t) + btq^n C_{n+1}(t).$$

Therefore

$$C_{n+1}(t) = \frac{t(1 - bq^n)}{(1 - btq^n)} C_n(t). \tag{9.7}$$

By iterating (9.7) and noting that $C_0(t) = 1$, we find that

$$C_n(t) = \frac{t^n(b)_n}{(bt)_n}. \tag{9.9}$$

Substituting (9.8) into (9.5), we see that

$$\begin{aligned}\frac{(bXt)_\infty(t)_\infty}{(Xt)_\infty(bt)_\infty} &= \sum_{n\geq 0} \frac{g_n(X)t^n}{(q)_n} \\ &= \sum_{n\geq 0} \frac{C_n(t)P_n(X,1)}{(q)_n} \\ &= \sum_{n\geq 0} \frac{P_n(X,1)(b)_n t^n}{(q)_n(bt)_n}\end{aligned} \tag{9.9}$$

Equation (9.9) is the Heine–Gauss theorem [21], p. 97, equation (3.3.2.5).

9.4 The Rogers–Ramanujan Identities. Here we consider the Eulerian differential operator

$$R_q = \frac{1}{X}(\eta^{-2} - \eta^{-1}).$$

Let $r_n(X)$ denote the associated Eulerian family, and let

$$\sum_{n\geq 0} \frac{r_n(X)t^n}{(q)_n} = \frac{\rho(Xt)}{\rho(t)}.$$

By Theorem 10,

$$\begin{aligned}\rho(t) &= \sum_{n\geq 0} \frac{t^n}{(q)_n} \cdot \frac{(q)_n}{\prod_{j=1}^{n}(q^{-2j} - q^{-j})} \\ &= \sum_{n\geq 0} \frac{t^n}{\prod_{j=1}^{n} q^{-2j}(1-q^j)} \\ &= \sum_{n\geq 0} \frac{q^{n^2+n}t^n}{(q)_n}.\end{aligned}$$

Thus $\rho(t)$ is indeed one of the functions involved in the Rogers–Ramanujan identities (see [21], p. 103).

Now let us consider the following function:

$$F(t) = \sum_{n\geq 0} \frac{(-1)^n t^{2n} q^{n(5n+3)/2}(1 - tq^{2n+1})}{(q)_n(tq^{n+1})_\infty}$$

Then in the notation of Hardy and Wright [13], p. 294, equation (19.14.11)

$$F(t) = H_1(tq, q),$$

and by [13], p. 294, equation (19.14.15)

$$(\eta^{-2} - \eta^{-1})F(Xt) = XtF(Xt).$$

Thus

$$R_q F(Xt) = tF(Xt).$$

Hence by the corollary to Theorem 11,

$$\rho(t) = F(t). \tag{9.10}$$

Setting $t = q^{-1}$ in (9.10), we obtain

$$\sum_{n\geq 0} \frac{q^{n^2}}{(q)_n} = \frac{1 + \sum_{n=1}^{\infty} (-1)^n q^{n(5n-1)/2}(1 + q^n)}{(q)_\infty}$$
$$= (q; q^5)_\infty^{-1}(q^4; q^5)_\infty^{-1}, \tag{9.11}$$

by Jacobi's identity [13], p. 282.

Finally setting $t = 1$ in (9.10), we see that

$$\sum_{n\geq 0} \frac{q^{n^2+n}}{(q)_n} = \frac{\sum_{n=0}^{\infty} (-1)^n q^{n(5n+3)/2}(1 - q^{2n+1})}{(q)_\infty}$$
$$= (q^2; q^5)_\infty^{-1}(q^3; q^5)_\infty^{-1}, \tag{9.12}$$

by Jacobi's identity [13], p. 282.

Equations (9.11) and (9.12) constitute the Rogers–Ramanujan identities.

The results of this section give a small sampling of the relationship of the theory of Eulerian differential operators to the classical theory of basic hypergeometric series.

10. Applications to Eulerian Rodrigues formulae

One of the most useful results in [18] is Theorem 4 which presents several formulae for the iterative calculation of families of basic polynomials. In particular if Q is the delta operator related to the family of basic polynomials $p_n(x)$, then the Rodrigues-type formula [18; p. 194, equation (4)] may be rewritten as

$$(Qx - xQ)x^{-1}p_n(x) = p_{n-1}(x),$$

or equivalently

$$Qx^{-1}p_{n+1}(x) = nx^{-1}p_n(x). \tag{10.1}$$

Thus the Rodrigues-type formula of Rota and Mullin [18], p. 194, equation (4), is equivalent to the assertion that the family $x^{-1}p_1(x), x^{-1}p_2(x), x^{-1}p_3(x), \ldots$ is a Sheffer set (in the notation of [16] which was described in our Section 8) provided each polynomial is multiplied by $p_1'(0)$ so that $[x^{-1}p_1(x)/p_1'(0)]_{X=1} = 1$.

Hence we may prove the Rodrigues-type formula of Rota and Mullin [18], p. 194, equation (4), if we can establish that $\{x^{-1}p_{n+1}(x)/p_1'(0)\}$ is a Sheffer set relative to Q. This is possible in the following manner[2]

$$\sum_{n\geq 0} \frac{x^{-1}p_{n+1}(x)t^n}{p_1'(0)n!} = \frac{1}{xp_1'(0)}\frac{d}{dt}\sum_{n\geq 0}\frac{p_n(x)t^n}{n!}$$
$$= \frac{1}{xp_1'(0)}\frac{d}{dt}\exp\left\{x\sum_{n\geq 0}\frac{p_n'(0)t^n}{n!}\right\}$$
$$= \frac{1}{p_1'(0)}\sum_{n\geq 0}\frac{p_{n+1}'(0)t^n}{n!}\exp\left\{x\sum_{n\geq 0}\frac{p_n'(0)t^n}{n!}\right\}$$
$$= h(t)\exp\left\{x\sum_{n\geq 0}\frac{p_n'(0)t^n}{n!}\right\}$$

where $h(0) = 1$. Thus by our remarks in the beginning of Section 8, the

$$x^{-1}p_{n+1}(x)/p_1'(0)$$

do indeed form a Sheffer set relative to Q.

Our object now is to follow the q-analog of this procedure. As we shall see a simple formula like (10.1) does not hold in general for Eulerian families; however, more complicated recurrences can be obtained. We shall content ourselves with examining the polynomials $g_n(X)$ introduced in Section 9.3. Define for $n > 0$

$$G_n(X) = \frac{g_{n+1}(X) - bXg_n(X)(1 - q^n)}{X - 1}, \quad G_0(X) = 1 - b. \tag{10.2}$$

Then if $D_{q,t}$ denotes q-differentiation with respect to t,

$$\sum_{n\geq 0}\frac{G_n(X)t^n}{(q)_n} = (X-1)^{-1}(1 - bXt)\sum_{n\geq 0}\frac{g_{n+1}(X)t^n}{(q)_n}$$
$$= (X-1)^{-1}(1 - bXt)D_{q,t}\sum_{n\geq 0}\frac{g_n(X)t^n}{(q)_n}$$
$$= \frac{(X-1)^{-1}(1 - bXt)(bXtq)_\infty(tq)_\infty}{t(Xt)_\infty(bt)_\infty}$$
$$\times\{(1 - bXt)(1 - t) - (1 - Xt)(1 - bt)\}$$
$$= \frac{(1-b)}{1-t}\frac{(bXt)_\infty(t)_\infty}{(Xt)_\infty(bt)_\infty}$$
$$= \frac{(1-b)}{(1-t)}\sum_{n\geq 0}\frac{g_n(X)t^n}{(q)_n}.$$

Therefore by the corollary to Theorem 12, the $(1 - b)^{-1}G_n(X)$ form an Eulerian Sheffer family relative to γ. Hence

$$\gamma G_n(X) = (1 - q^n)G_{n-1}(X) \tag{10.3}$$

As we see (10.3) is quite a bit more complicated than (10.1), and, in general, matters are even worse. The reason is that for basic families $p_n(x)$ associated with delta

[2] I wish to thank Gian-Carlo Rota for supplying me with an equivalent form of this argument.

operators (by Corollary 2 of Theorem 3 in [18]),

$$\log \sum_{n\geq 0} \frac{p_n(x)t^n}{n!} = x \sum_{n\geq 0} \frac{p_n'(0)t^n}{n!},$$

which is a linear function of x. For Eulerian families $\pi_n(X)$ associated with Eulerian differential operators, we see by Theorem 8 that

$$\log \sum_{n\geq 0} \frac{\pi_n(X)t^n}{n!} = \sum_{n\geq 1} \frac{\pi_n'(1)X^n t^n}{(q)_n n} - \sum_{n\geq 1} \frac{\pi_n'(1)t^n}{(q)_n n},$$

and in general this is a very complicated function of X. Thus it is not surprising that recurrences among the $\pi_n(X)$ are more complicated.

In actual fact, Theorem 4 and its corollary provide very effective means for recursively defining Eulerian families.

11. Applications to finite vector spaces

Just as the theory of delta operators developed by Rota and Mullin is useful in the combinatorics of finite sets, so our theory is useful in the combinatorics of finite vector spaces.

First we remark that Rota and Goldman [12], Section 5, have studied $P_n(X, Z)$ in detail and have shown that $P_n(X, Z)$ is the number of one-to-one linear transformations f of $\mathcal{N}$ into $\mathcal{X}$ such that $f(\mathcal{N}) \cap \mathcal{Z} = \{0\}$ where $\mathcal{Z}$ is a subspace of $\mathcal{X}$. They also established combinatorially the q-binomial theorem:

$$P_n(X, Z) = \sum_{l\geq 0} \binom{n}{l}_q P_l(X, Y)P_{n-l}(Y, Z), \tag{11.1}$$

a result equivalent to our (3.1). We have already seen that the $P_n(X, 1)$ form an Eulerian family. We conclude by considering a new Eulerian family, and we show how combinatorial studies may lead to analytic identities.

DEFINITION 7. Let $\mathcal{H}_n(X, U, W)$ denote the number of one-to-one linear transformations f of $\mathcal{N}$ into $\mathcal{U} \oplus \mathcal{X}$ where all non-zero $\mathcal{U}$-components of f-images lie outside of $\mathcal{W}$ a subspace of $\mathcal{U}$.

DEFINITION 8. $h_n(X) = \mathcal{H}_n(X, U, UX^{-1})$.

Combinatorially we may think of $h_n(X)$ as being defined exactly as $\mathcal{H}_n(X, U, W)$ is with the added condition that $w = u - x$.

PROPOSITION 1. For each $n \geq 0$,

$$\mathcal{H}_n(X, U, W) = \sum_{j\geq 0} \binom{n}{j}_q P_j(X, 1)P_{n-j}(U, W)X^{n-j}.$$

Proof: Let us look at the maps f counted by $\mathcal{H}_n(X, U, W)$ for which the subspace of $f(\mathcal{N})$ with 0 as $\mathcal{U}$-component is j-dimensional. The number of such maps is obtained as follows: We can choose a j-dimensional subspace of $\mathcal{N}$ in $\binom{n}{j}_q$ ways.

We can then map this chosen subspace into $\mathscr{X}$ in $P_j(X, 1)$ ways. If $v_1, \ldots, v_j$ form a basis for this j-dimensional subspace of $\mathscr{N}$, we can extend to $v_1, \ldots, v_n$ a basis for $\mathscr{N}$. By the same argument used in Section 3, we need only choose $f(v_{j+1}), \ldots, f(v_n)$ so that their $\mathscr{U}$-components are linearly independent (and now outside of $\mathscr{W}$ as well). This can clearly be done in $X^{n-j}P_{n-j}(U, W)$ ways. Hence

$$\mathscr{H}_n(X, U, W) = \sum_{j \geq 0} \binom{n}{j}_q P_j(X, 1)P_{n-j}(U, W)X^{n-j}.$$

PROPOSITION 2. For each $n \geq 0$,

$$h_n(X) = \sum_{j \geq 0} \binom{n}{j}_q P_j(X, 1)P_{n-j}(X, 1)U^{n-j}.$$

Proof:

$$\begin{aligned} h_n(X) &= \mathscr{H}_n(X, U, UX^{-1}) \\ &= \sum_{j \geq 0} \binom{n}{j}_q P_j(X, 1)P_{n-j}(U, UX^{-1})X^{n-j} \\ &= \sum_{j \geq 0} \binom{n}{j}_q P_j(X, 1)(UX^{-1})^{n-j}P_{n-j}(X, 1)X^{n-j} \\ &= \sum_{j \geq 0} \binom{n}{j}_q P_j(X, 1)P_{n-j}(X, 1)U^{n-j}. \end{aligned}$$

PROPOSITION 3. The $h_n(X)$ form an Eulerian family of polynomials.

Proof: By Proposition 2 we see that $h_0(X) = 1$ and $h_n(X)$ is a polynomial of degree n in X for each n. Finally $h_n(XY)$ counts the number of one-to-one linear transformations f of $\mathscr{N}$ into $\mathscr{U} \oplus \mathscr{X} \oplus \mathscr{Y}$ where all non-zero $\mathscr{U}$-components of f-images lie outside of $\mathscr{T}$, a subspace of $\mathscr{U}$ with $u - t = x + y$.

Let us look at the maps counted by $h_n(XY)$ for which the subspace of $f(\mathscr{N})$ with 0 as $\mathscr{U} \oplus \mathscr{X}$-component is j-dimensional. The number of such maps is obtained as follows: We can choose a j-dimensional subspace of $\mathscr{N}$ in $\binom{n}{j}_q$ ways. We can then map this subspace into $\mathscr{Y}$ in $P_j(Y, 1)$ ways. Extending a basis $v_1, \ldots, v_j$ of this j-dimensional subspace of $\mathscr{N}$ to a basis $v_1, \ldots, v_n$ for $\mathscr{N}$, we see by the same argument used in Section 3 that we need only choose $f(v_{j+1}), \ldots, f(v_n)$ so that their $\mathscr{U} \oplus \mathscr{X}$-components are linearly independent and their non-zero $\mathscr{U}$-components are outside $\mathscr{T}$ This choice can be made in $Y^{n-j}\mathscr{H}_{n-j}(X, U, T)$ ways. Therefore

$$h_n(XY) = \sum_{j \geq 0} \binom{n}{j}_q P_j(Y, 1)Y^{n-j}\mathscr{H}_{n-j}(X, U, T)$$

Now choose $\mathscr{W}$ so that $\mathscr{U} \supset \mathscr{W} \supset \mathscr{T}$ and $u - w = x$ (consequently $w - t = y$).

Utilizing (11.1), we see that

$$\begin{aligned} h_n(XY) &= \sum_{j\ge 0} \begin{pmatrix} n \\ j \end{pmatrix}_q P_j(Y,1)Y^{n-j}\mathscr{H}_{n-j}(X,U,T) \\ &= \sum_{j\ge 0} \begin{pmatrix} n \\ j \end{pmatrix}_q P_j(Y,1)Y^{n-j} \sum_{r\ge 0} \begin{pmatrix} n-j \\ r \end{pmatrix}_q P_r(X,1)P_{n-j-r}(U,T)X^{n-j-r} \\ &= \sum_{j\ge 0} \begin{pmatrix} n \\ j \end{pmatrix}_q P_j(Y,1)Y^{n-j} \sum_{r\ge 0} \begin{pmatrix} n-j \\ r \end{pmatrix}_q P_r(X,1)X^{n-j-r} \\ &\quad \times \sum_{l\ge 0} \begin{pmatrix} n-j-r \\ l \end{pmatrix}_q P_l(U,W)P_{n-j-r-l}(W,T). \end{aligned} \tag{11.2}$$

Now if $p = r + l$, then

$$\begin{aligned} \begin{pmatrix} n \\ j \end{pmatrix}_q \begin{pmatrix} n-j \\ r \end{pmatrix}_q \begin{pmatrix} n-j-r \\ l \end{pmatrix}_q &= \frac{(q)_n}{(q)_j(q)_r(q)_l(q)_{n-j-r-l}} \\ &= \begin{pmatrix} n \\ p \end{pmatrix}_q \begin{pmatrix} p \\ r \end{pmatrix}_q \begin{pmatrix} n-p \\ j \end{pmatrix}_q. \end{aligned}$$

Hence interchanging the summations in (11.2) and replacing l by $p - r$, we see that

$$\begin{aligned} h_n(XY) &= \sum_{p\ge 0} \begin{pmatrix} n \\ p \end{pmatrix}_q Y^p \sum_{r\ge 0} \begin{pmatrix} p \\ r \end{pmatrix}_q P_r(X,1)P_{p-r}(U,W)X^{p-r} \\ &\quad \sum_{j\ge 0} \begin{pmatrix} n-p \\ j \end{pmatrix}_q P_j(Y,1)P_{n-j-p}(W,T)(XY)^{n-j-p} \\ &= \sum_{p\ge 0} \begin{pmatrix} n \\ p \end{pmatrix}_q Y^p \sum_{r\ge 0} \begin{pmatrix} p \\ r \end{pmatrix}_q P_r(X,1)P_{p-r}(X,1)U^{p-r} \\ &\quad \sum_{j\ge 0} \begin{pmatrix} n-p \\ j \end{pmatrix}_q P_j(Y,1)P_{n-j-p}(Y,1)U^{n-j-p} \\ &= \sum_{p\ge 0} \begin{pmatrix} n \\ p \end{pmatrix}_q Y^p h_p(X)h_{n-p}(Y). \end{aligned}$$

Knowing that the $h_n(X)$ form an Eulerian family, we can derive an identity of Carlitz [5], p. 361, equation (2.2). First we see by inspection of Proposition 2 that the leading coefficient of $h_n(X)$ is

$$H_n(U) = \sum_{j\ge 0} \begin{pmatrix} n \\ j \end{pmatrix}_q U^j,$$

the q-Hermite polynomial mentioned in Section 8. Furthermore for $n > 0$

$$\begin{aligned} h_n'(1) &= \lim_{X\to 1}(X-1)^{-1}h_n(X) \\ &= \lim_{X\to 1}(X-1)^{-1} \sum_{j\ge 0} \begin{pmatrix} n \\ j \end{pmatrix}_q P_j(X,1)U^{n-j}P_{n-j}(X,1) \\ &= U^n(q)_{n-1} + (q)_{n-1} \\ &= (q)_{n-1}(1 + U^n). \end{aligned}$$

These facts now allow us to establish the following result:

PROPOSITION 4. ([5], p. 361, equation (2.2))

$$\sum_{n\geq 0}\frac{H_n(U)t^n}{(q)_n}=(t)_\infty^{-1}(tU)_\infty^{-1}.$$

Proof: By Theorem 8,

$$\begin{aligned}\sum_{n\geq 0}\frac{H_n(U)t^n}{(q)_n}&=\exp\left\{\sum_{n\geq 1}\frac{h_n'(1)t^n}{(q)_n n}\right\}\\&=\exp\left\{\sum_{n\geq 1}\frac{(1+U^n)t^n}{(1-q^n)n}\right\}\\&=\exp\left\{\sum_{m\geq 0}\sum_{n\geq 1}\frac{q^{nm}(1+U^n)t^n}{n}\right\}\\&=\exp\left\{-\sum_{m\geq 0}\log(1-tq^m)-\sum_{m\geq 0}\log(1-tUq^m)\right\}\\&=\prod_{m=0}^{\infty}(1-tq^m)^{-1}(1-tUq^m)^{-1}=(t)_\infty^{-1}(tU)_\infty^{-1}.\end{aligned}$$

12. Conclusion

In light of the characterization of Eulerian differential operators given in Theorem 4, and since the factor $1-q^n$ does not appear in this characterization, we may reasonably ask what happens if the sequence $0, 1-q, 1-q^2,\ldots$ were replaced by $u_0=0, u_1, u_2,\ldots$ (where $u_n\neq 0$ for each $n>0$) in Definition 1. Actually we can prove all the theorems through Section 8 with

$$\begin{array}{lll}1-q^n & \text{replaced by} & u_n,\\ \binom{n}{r}_q & \text{replaced by} & \dfrac{u_nu_{n-1}\cdots u_{n-r+1}}{u_ru_{r-1}\cdots u_1},\\ (q)_n & \text{replaced by} & u_nu_{n-1}\cdots u_1\end{array}$$

Indeed the results would extend the work of Morgan Ward in [23]; however, such a generalization seems of little immediate value in applications (such as in Sections 3, 9, 10, and 11), and we have, therefore, not bothered to write our results in this more general form.

There are many other possible applications of our theory. For example, the series-product identity of F. H. Jackson [21], p. 96, equation (3.3.1.3)

$$1+\sum_{n\geq 1}\frac{(aq)_{n-1}(1-aq^{2n})(b)_n(c)_n(d)_n(aq/bcd)^n}{(q)_n(aq/b)_n(aq/c)_n(aq/d)_n}=\frac{(aq)_\infty(aq/bc)_\infty(aq/bd)_\infty(aq/cd)_\infty}{(aq/b)_\infty(aq/c)_\infty(aq/d)_\infty(aq/bcd)_\infty}$$

can be transformed into a generating function for a family of Eulerian polynomials by the substitutions $a=Xt$, $b=X$. Further aspects of the classical theory of basic hypergeometric series can be included in the theory of Eulerian differential operators.

As for applications to finite vector spaces, we first remark that it should be possible to extend the results of Section 11 to polynomials of the form

$$\sum_{n\geq 0} \frac{\tilde{h}_n(X)t^n}{(q)_n} = \frac{(t)_\infty(U_1t)_\infty(U_2t)_\infty\ldots(U_rt)_\infty}{(Xt)_\infty(U_1Xt)_\infty(U_2Xt)_\infty\ldots(U_rXt)_\infty}.$$

A more tantalizing problem involves a finite vector space interpretation for the $g_n(X)$ defined in (9.4) (obviously $g_n(X) = P_n(X, 1)$ if $b = 0$).

Professor L. Carlitz has drawn my attention to the paper by A. Sharma and A. Chak (The basic analog of a class of polynomials, Revista di Matematica della Universita di Parma, 5 (1954), 325–337) and to the paper by W. A. Al-Salam (q-Appell polynomials, Annali di Matematica, 77 (1967), 31–45). In the present context, the polynomials studied in these papers are essentially Eulerian Sheffer polynomials related to the Eulerian differential operator D_q. The q-differential operators L_p discussed by Al-Salam on page 43 of his paper are not Eulerian differential operators (as defined here) except when $L_p = D_q$ (cf. our Theorem 9).

Also Professor W. A. Al-Salam has drawn my attention to the extensive literature on generalized Sheffer polynomials. In particular, he pointed out the forthcoming paper by A. M. Chak (An Extension of a Class of Polynomials) in which our Eulerian Sheffer polynomials are named "Appell Polynomials to the Base c". Also Professor Al-Salam mentioned the work by Mourad El Houssieny Ismail (Classification of Polynomial Sets, M.S. Thesis, 1969, University of Alberta) in which an extensive account of generalized Sheffer polynomials is given and in which appears a list of 119 references. Presumably our results in Section 8 duplicate those of Chak; however, other than there, our results appear to be new.

References

1. W. N. Bailey, Generalized Hypergeometric Series. Cambridge Mathematical Tract No. 32, Cambridge University Press, Cambridge, 1935.
2. W. N. Bailey, "Some identities in combinatory analysis," Proc. London Math. Soc. (2), **49** (1947), 421–435.
3. R. Bellman, A Brief Introduction to Theta Functions, Holt, Rinehart, and Winston, New York, 1961.
4. C. Berge, Principles of Combinatorics, Academic Press, New York, 1971.
5. L. Carlitz, "Some polynomials related to theta functions," Ann. Math. Pura App., **11** (1956), 359–373.
6. L. Carlitz, "Some polynomials related to theta functions," Duke Math. J., **24** (1957), 521–527.
7. L. Carlitz, "Eulerian numbers and polynomials," Math. Mag., **32** (1958), 247–260.
8. L. Carlitz, "A note on the Eulerian numbers," Arch. der Math., **14** (1963), 383–390.
9. L. Carlitz, "Eulerian numbers and operators, Proc. of the Conf. on the Theory of Arithmetic Functions, Springer, New York (to appear).
10. N. J. Fine, Some Basic Hypergeometric Series (to appear).
11. J. Goldman and G.-C. Rota, "The number of subspaces of a vector space," in Recent Progress in Combinatorics edited by W. T. Tutte, Academic Press (1969), 75–83.
12. J. Goldman and G.-C. Rota, "On the foundations of combinatorial theory: IV finite vector spaces and Eulerian generating functions," Studies in Applied Mathematics, **49** (1970), 239–258.
13. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford University Press, Oxford, 1960.
14. F. H. Jackson, "q-Form of Taylor's theorem," Mess. of Math., **38** (1909), 62–64.

15. F. H. JACKSON, "On a formula relating to hypergeometric series," Mess. of Math., **37** (1908), 123–126.
16. D. KAHANER AND G.-C. ROTA, Sheffer polynomials, (to appear).
17. D. KNUTH, "Subspaces, subsets, and partitions," J. Combinatorial Theory (A), **10** (1971), 178–180.
18. R. MULLIN AND G.-C. ROTA, "On the foundations of combinatorial theory: III theory of binomial enumeration," in Graph Theory and Its Applications edited by B. Harris, Academic Press, New York (1970), 167–213.
19. L. J. ROGERS, "On the expansion of some infinite products," Proc. London Math. Soc. (1), **24** (1893), 337–352.
20. B. SEGRE, "Introduction to Galois geometries," Atti Accad. Naz. Lincei Mem. Cl. Sc. Fis. Mat. Natur. (8), **8** (1967), 135–236.
21. L. J. SLATER, Generalized Hypergeometric Functions, Cambridge University Press, Cambridge, 1966.
22. G. SZEGÖ, "Ein Bertrag zur Theorie der Thetafunktionen, Sitz. der Preuss. Akad. der Wiss.," Phys.-Math. Klasse, (1926), 242–252.
23. MORGAN WARD, "A calculus of sequences, "Amer. J. Math., **58** (1936), 255–266.
24. G. N. WATSON, "The mock theta functions (II)," Proc. London Math. Soc. (2), **42** (1936), 377–397.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PENNSYLVANIA STATE UNIVERSITY

(Received May 21, 1971)

ON THE FOUNDATIONS OF COMBINATORIAL THEORY (VI): THE IDEA OF GENERATING FUNCTION

PETER DOUBILET, GIAN-CARLO ROTA
and
RICHARD STANLEY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

1. Introduction

Since Laplace discovered the remarkable correspondence between set theoretic operations and operations on formal power series, and put it to use with great success to solve a variety of combinatorial problems, generating functions (and their continuous analogues, namely, characteristic functions) have become an essential probabilistic and combinatorial technique. A unified exposition of their theory, however, is lacking in the literature. This is not surprising, in view ot the fact that all too often generating functions have been considered to be simply an application of the current methods of harmonic analysis. From several of the examples discussed in this paper it will appear that this is not the case: in order to extend the theory beyond its present reaches and develop new kinds of algebras of generating functions better suited to combinatorial and probabilistic problems, it seems necessary to abandon the notion of group algebra (or semigroup algebra), so current nowadays, and rely instead on an altogether different approach.

The insufficiency of the notion of semigroup algebra is clearly seen in the example of Dirichlet series. The functions

$$n \to 1/n^s \tag{1.1}$$

defined on the semigroup S of positive integers under multiplication, are characters of S. They are not, however, all the characters of this semigroup, nor does there seem to be a canonical way of separating these characters from the rest (see, for example, Hewitt and Zuckerman [32]). In other words, there does not seem to be a natural way of characterizing the algebra of formal Dirichlet series as a subalgebra of the semigroup algebra (eventually completed under a suitable topology) of the semigroup S. In the present theory, however, the algebra of formal Dirichlet series arises naturally from the incidence algebra (definition below) of the lattice of finite cyclic groups, as we shall see.

The purpose of this work is to begin the development of a theory of generating functions that will not only include all algebras of generating functions used so far (ordinary, exponential, Dirichlet, Eulerian, and so on), but also provide a systematic technique for setting up other algebras of generating functions suited to particular enumerations. Our initial observation is that most families of discrete structures, while often devoid of any algebraic composition laws, are nevertheless often endowed with a natural order structure. The solution of the problem of their enumeration thus turns out to depend more often than not upon associating suitable computational devices to such order structures.

Our starting point is the notion of *incidence algebra*, whose study was briefly begun in a previous paper, and which is discussed anew here. Section 3 contains the main facts on the structure of the incidence algebra of an ordered set; perhaps the most interesting new result is the explicit characterization of the lattice of two sided ideals. It follows from recent results of Aigner, Prins, and Gleason (motivated by the present work) that for an ordered set with a unique minimal element the incidence algebra is uniquely characterized by its lattice of ideals; this assertion is no longer true if the ordered set has no unique minimal element. In particular, the lattice of two sided ideals is distributive, an unusual occurrence in a noncommutative algebra. Our characterization of the radical suggests that a simple axiomatic description of incidence algebras should be possible, and we hope someone will undertake this task.

Section 4 introduces the main working tool, namely, the reduced incidence algebra. This notion naturally arises in endowing the segments of an ordered set with an equivalence relation. Such an equivalence is usually dictated by the problem at hand, and leads to the definition of the *incidence coefficients*, a natural generalization of the classical binomial coefficients. After a brief study of the family of all equivalence relations compatible with the algebra structure, we show by examples that all classical generating functions (and their incidence coefficients) can be obtained as reduced incidence algebras. We believe this is a remarkable fact, and perhaps the most cogent argument for the use of the present techniques.

Section 5 extends the notion of reduced incidence algebras to families of ordered structures. The notion of multiplicative functions on partitions of a set and the isomorphism with the semigroup of formal power series without constant term under functional composition (Theorem 5.1) are perhaps the most important results here. Because of space limitations, we have given only a few applications, which hopefully should indicate the broad range of problems which it can solve (for example, enumeration of solutions of an equation in the symmetric group G_n, as a function of n). Pursuing the same idea, we obtain an algebra of multiplicative functions on a class of ordered structures recently studied by Dowling [19], which were suggested by problems in coding theory. Finally, we obtain the algebra of Philip Hall, arising from the enumeration of abelian groups, as a large incidence algebra.

Section 6 studies the strange phenomenon pointed out in Section 4, that the maximally reduced incidence algebra does not coincide with the algebra obtained by identifying isomorphic segments of an ordered set. The structure of such an algebra is determined.

Sections 7, 8, and 9 make a detailed study of those algebras of generating functions which are closest to the classical cases. Algebras of Dirichlet type are those where all the analogs of classical number theoretic functions can be defined, including the classical product formula for the zeta function. Algebras of binomial type are close to the classical exponential generating functions, and naturally arise in connection with certain block designs. Under mild hypotheses, we give a complete classification of such algebras.

Several applications and a host of other examples could not be treated here. Among them, we mention a general theory of multiplicative functions, and their relation to the coalgebra structure (as sketched in Goldman and Rota [25]), and large incidence algebras arising in the study of classes of combinatorial geometries closed under the operation of taking minors, in particular the *coding geometries* of R. C. Bose and B. Segre, of which the Dowling lattices are special cases.

This work was begun in Los Alamos in the summer of 1966. Since then, the notion of reduced incidence algebra was independently discovered by D. A. Smith and H. Scheid, who developed several interesting properties. The bulk of the material presented here, with the obvious exception of some of the examples, is believed to be new.

2. Notations and terminology

Very little knowledge is required to read this work. Most of the concepts basic enough to be left undefined in the succeeding sections will be introduced here.

A *partial ordering relation* (denoted by $\leqq$) on a set P is one which is reflexive, transitive, and antisymmetric (that is, $a \leqq b$ and $b \leqq a$ imply $a = b$). A set P together with a partial ordering relation is a *partially ordered set*, or simply an *ordered set*. A *segment* $[x, y]$, for x and y in P, is the set of all elements z which satisfy $x \leqq z \leqq y$. A partially ordered set is *locally finite* if every segment is finite. We shall consider locally finite partially ordered sets only.

An ordered set P is said to have a 0 or a 1 if it has a unique minimal or maximal element.

An *order ideal* in an ordered set P is a subset Z of P which has the property that if $x \in Z$ and $y \leqq x$, then $y \in Z$.

The *product* $P \times Q$ of two ordered sets P and Q is the set of all ordered pairs (p, q), where $p \in P$ and $q \in Q$, endowed with the order $(p, q) \geqq (r, s)$ whenever $p \geqq r$ and $q \geqq s$. The product of any number of partially ordered sets is defined similarly. The *direct sum* or *disjoint union* $P + Q$ of two ordered sets P and Q is the set theoretic disjoint union of P and Q, with the ordering $x \leqq y$ if and only if (i) $x, y \in P$ and $x \leqq y$ in P or (ii) $x, y \in Q$ and $x \leqq y$ in Q. Note that if $p \in P$ and $q \in Q$, then p and q are incomparable.

In an ordered set P, an element p *covers* an element q when the segment $[q, p]$ has two elements. An *atom* is an element which covers a minimal element.

A *chain* is an ordered set in which every pair of elements is comparable. A *maximal chain* in a segment $[x, y]$ of an ordered set P is a sequence $(x_0, x_1, \cdots, x_n)$, where $x_0 = x$, $x_n = y$, and x_{i+1} covers x_i for all i. The chain $(x_0, x_1, \cdots, x_n)$ is said to have *length n*. An *antichain* is an ordered set in which no two distinct elements are comparable.

The *dual* P^* of an ordered set P is the ordered set obtained from P by inverting the order.

A *lattice* is an ordered set where max and min of two elements (we call them join and meet, and write them $\vee$ and $\wedge$) are defined. A *complete lattice* is a lattice in which the join and meet of any subset exist. A *sublattice* L' of a lattice L is a subset which is a lattice with the induced order relation and in which join and meet of two elements correspond with the join and meet in L. For the definitions of *distributive*, *modular*, and *semimodular* see Birkhoff.

A *partition* of a set S is a set of disjoint nonempty subsets of S whose union is S. The subsets of S making up the partition are called the *blocks* of the partition. The *lattice of partitions* $\Pi(S)$ of a set S is the set of partitions of S, ordered by *refinement*: a partition π is less than a partition σ (or is a refinement of σ) if every block of π is contained in a block of σ. The 0 of $\Pi(S)$ is the partition whose blocks are the one element subsets of S, and the 1 of $\Pi(S)$ is the partition with one block. There is a natural correspondence between equivalence relations on a set S and partitions of S, since the equivalence classes of an equivalence relation form the blocks of a partition, and hence, there is an induced lattice structure on the family of equivalence relations of S.

At the beginning of Section 3, we define the *incidence algebra* $\mathbf{I}(P, K)$ of a locally finite ordered set P, over a field K. We assume throughout that K has characteristic 0, except for the last paragraph of Section 6 when it is explicitly stated that another characteristic is being considered. We also assume that K is a topological field, and if the topology of K is not specified, we regard K as having the discrete topology.

A certain familiarity is assumed with pp. 342–347 of *Foundations I* ([49]), when the definitions of Möbius function and zeta function are given and some elementary properties of the incidence algebra are derived.

3. Structure of the incidence algebra

3.1 *Basic identifications.* As mentioned in Section 2, we define the *incidence algebra* $\mathbf{I}(P, K)$ of a locally finite ordered set P, over a field K, as follows. The members of $\mathbf{I}(P, K)$ are K valued functions $f(x, y)$ of two variables, with x and y ranging over P and with the sole restriction that $f(x, y) = 0$ unless $x \leqq y$. The sum of two such functions, as well as multiplication by scalars, are defined as usual, and the product $f*g = h$ is defined as follows,

$$h(x, y) = \sum_{z \in P} f(x, z)g(z, y). \tag{3.1}$$

In virtue of the assumption that the ordered set P is locally finite, the variable z in the sum on the right ranges over the finite segment $[x, y]$.

It is immediately verified that this product is associative. It is also easily verified that the incidence algebra is commutative if and only if the order relation of P is trivial, that is, if and only if no two elements of P are comparable. Whenever convenient, we shall omit mention of the field K and briefly write $\mathbf{I}(P)$, with the tacit convention that K is to remain fixed throughout.

The identity element of $\mathbf{I}(P)$ will be denoted by δ, after the Kronecker delta. In addition, we use the following notation for certain elements of $\mathbf{I}(P)$. If $x \in P$, let

$$e_x(u, v) = \begin{cases} 1 & \text{if } u = v = x, \\ 0 & \text{otherwise}, \end{cases} \tag{3.2}$$

and for $x \leqq y$, let

$$\delta_{x,y}(u, v) = \begin{cases} 1 & \text{if } u = x \text{ and } v = y, \\ 0 & \text{otherwise}. \end{cases} \tag{3.3}$$

Clearly, the elements e_x are idempotent, and the $\delta_{x,y}$ are analogous to the matrix units of ring theory (see Jacobson [35]). Note that $e_x = \delta_{x,x}$.

The following easily verified identities will be used in the sequel:

$$\text{if } g = \delta_{x,y}*f, \text{ then } g(u, v) = \begin{cases} 0 & \text{if } u \neq x, \\ f(y, v) & \text{if } u = x; \end{cases} \tag{3.4}$$

$$\text{if } g = f*\delta_{z,w}, \text{ then } g(u, v) = \begin{cases} 0 & \text{if } v \neq w, \\ f(u, z) & \text{if } v = w, \end{cases} \tag{3.5}$$

$$\text{if } g = \delta_{x,y}*f*\delta_{z,w}, \text{ then } g(u, v) = \begin{cases} 0 & \text{if } u \neq x \text{ or } v \neq w \\ f(y, z) & \text{if } u = x \text{ and } v = w, \end{cases} \tag{3.6}$$

that is, $\delta_{x,y}*f*\delta_{z,w} = f(y, z)\delta_{x,w}$. In particula , $e_x*f*e_y = f(x, y)\delta_{x,y}$, and $\delta_{x,y}*\delta_{z,w} = \delta(y, z)\delta_{x,w}$.

3.2 *The standard topology.* A topology on $\mathbf{I}(P)$ is defined as follows. A generalized sequence $\{f_n\}$ converges to f in $\mathbf{I}(P)$ if and only if $f_n(x, y)$ converges to $f(x, y)$ in the field K for every x and y. We call this the *standard topology* of $\mathbf{I}(P)$.

PROPOSITION 3.1. *Let P be a locally finite ordered set. Then the incidence algebra $\mathbf{I}(P)$, equipped with the standard topology, is a topological algebra.*

PROOF. In the right side of the definition (3.1) of the product, only a finite number of terms occur for fixed x and y; this implies at once that the product $(f, g) \to f*g$ is continuous in both variables. The verification of all other properties is immediate. *Q.E.D.*

In the sequel, we shall often have occasion to use infinite sums of the form

$$f = \sum_{x, y \in P} f(x, y)\, \delta_{x,y}, \tag{3.7}$$

and we shall presently discuss the meaning that is to be attached to the right side. Let Φ be a directed set of finite subsets of $P \times P$, with the following properties: (i) Φ is ordered by inclusion; (ii) for every pair $x, y \in P$ there exists a member $A \in \Phi$ such that $(x, y) \in A$. We call such a directed set *standard*.

PROPOSITION 3.2. *Let Φ be a standard directed set. Then the set $\{f_A : A \in \Phi)$ defined as*

$$f_A = \sum_{(x, y) \in A} f(x, y)\, \delta_{x,y} \tag{3.8}$$

converges in the standard topology of the incidence algebra $\mathbf{I}(P)$ *to the element* f.

PROOF. Take $A \in \Phi$ so that $(x, y) \in A$. Then for every $B \in \Phi, B \geqq A$, we have $f_B(x, y) - f(x, y) = 0$. *Q.E.D.*

Speaking in classical language, the preceding proposition states that the "sum" on the right side of (3.7) converges to the element f together with all its "rearrangements". This justifies the use of the summation symbol on the right side of (3.7), and we shall make use of it freely from now on.

3.3. *Ideal structure.* We shall now determine the lattice of (two sided, closed) ideals of the incidence algebra $\mathbf{I}(P)$, endowed with the standard topology. For P finite, all two sided ideals are closed, so Theorem 3.1 below determines the lattice of all ideals.

Let J be a closed ideal in $\mathbf{I}(P)$, and let $\Delta(J)$ be the collection of all elements $\delta_{x,y}$ belonging to J. We call $\Delta(J)$ the *support* of the ideal J. Then, any finite or infinite linear combination of the $\delta_{x,y}$ in $\Delta(J)$ gives a member of J. Conversely, if $f \in J$, then, by 3.6 above,

$$e_x * f * e_y = f(x, y)\, \delta_{x,y}; \tag{3.9}$$

hence, if $f(x, y) \neq 0$, it follows that $\delta_{x,y} \in \Delta(J)$. This proves the following.

LEMMA 3.1. *Every closed ideal J in the incidence algebra $\mathbf{I}(P)$ consists of all functions $f \in \mathbf{I}(P)$ such that $f(x, y) = 0$ whenever $\delta_{x,y} \notin \Delta(J)$.*

Now, let $Z(J)$ be the family of all segments $[x, y]$ such that $f(x, y) = 0$ for all $f \in J$. Then we have

LEMMA 3.2. *If $[x, y] \in Z(J)$ and $x \leqq u \leqq v \leqq y$, then $[u, v] \in Z(J)$.*

The proof is immediate: Let $f \in J$. By (3.6) again,

$$\delta_{x,u} * f * \delta_{v,y} = f(u, v)\, \delta_{x,y}. \tag{3.10}$$

Thus, if $\delta_{x,y} \notin J$, then $f(u, v) = 0$, and $[u, v] \in Z(J)$.

We are now ready to state the main result.

THEOREM 3.1. *In a locally finite ordered set P, let $S(P)$ be the set of all segments of P, ordered by inclusion. Then there is a natural anti-isomorphism between the lattice of closed ideals of the incidence algebra $\mathbf{I}(P)$ and the lattice of order ideals of $S(P)$.*

PROOF. Let J be an ideal of P, and let $Z(J)$ be the family of segments defined above. Lemma 3.1 shows that $Z(J)$ uniquely determines J, and Lemma 3.2 shows that $Z(J)$ is an order ideal in $S(P)$.

Conversely, let Z be an order ideal in $S(P)$, and let J be the set of all $f \in \mathbf{I}(P)$ for which $f(x, y) = 0$, if $[x, y] \in Z$. Then J is an ideal. Indeed, if $g \in \mathbf{I}(P)$ is arbitrarily chosen, if $f \in J$, if $[x, y] \in Z$, and if $h = f * g$, then

$$h(x, y) = \sum_{x \leqq z \leqq y} f(x, z) g(z, y) = 0, \tag{3.11}$$

since all $f(x, z) = 0$ for z between x and y. The case is similar for multiplication on the left. Since we can take arbitrarily infinite sums as in (3.7), it follows that J is closed, and the proof is complete.

COROLLARY 3.1. *The lattice of closed ideals of an incidence algebra is distributive.*

COROLLARY 3.2. *The closed maximal ideals of an incidence algebra* $\mathbf{I}(P)$ *are those of the form*

$$J_x = \{f \in \mathbf{I}(P) \mid f(x, x) = 0\}, \tag{3.12}$$

where $x \in P$.

3.4. *The radical.* We recall the well-known and easily proved fact (see Smith [55], or *Foundations I*) that an element f of the incidence algebra has an inverse if and only if $f(x, x) \neq 0$ for all $x \in P$. From this it follows (Jacobson [35], p. 8, and following) that an element $f \in \mathbf{I}(P)$ is *quasiregular* if and only if $f(x, x) \neq 1$ for all $x \in P$. Hence, an element f has the property that $g * f * h$ is quasiregular for all g and h, if and only if $f(x, x) = 0$ for all $x \in P$. From Proposition 1 on page 9 of Jacobson, we make the following inference.

PROPOSITION 3.3. *The radical R of the incidence algebra* $\mathbf{I}(P)$ *of a locally finite ordered set P is the set of all* $f \in \mathbf{I}(P)$ *such that* $f(x, x) = 0$ *for all* $x \in P$.

3.5. *The incidence algebra as a functor.* We now determine a class of maps between locally finite ordered sets so that the association of the incidence algebra to such sets can be extended, in a natural way, to a functor into the category of K algebras (where K is the fixed ground ring or field). A function σ from an ordered set P to an ordered set Q will be called a *proper map* if it satisfies the following three conditions:

(a) σ is one to one;

(b) $\sigma(p_1) \leqq \sigma(p_2)$ implies $p_1 \leqq p_2$;

(c) if q_1 and q_2 are in the image of σ, and $q_1 \leqq q_2$, then the whole segment $[q_1, q_2]$ is in the image.

Note that in view of (a) and (b), condition (c) can be replaced by

(c′) if $\sigma(p_1) \leqq \sigma(p_2)$ and $q \in [\sigma(p_1), \sigma(p_2)]$, then there is a unique $p \in [p_1, p_2]$ such that $\sigma(p) = q$.

It is clear that the identity function on any partially ordered set is a proper map, and it is not hard to verify that the composition of proper maps is a proper map. Thus, ordered sets together with proper maps form a category. Let $\mathscr{A}$ be the subcategory of locally finite ordered sets together with proper maps. We then have the following proposition.

PROPOSITION 3.4. (i) *The mapping* $\mathbf{I}$ *from* $\mathscr{A}$ *to the category of* K *algebras, given by* $\mathbf{I}(P)$ = *incidence algebra of* P (*with values in* K) *and*

$$[\mathbf{I}(\sigma)(f)](p_1, p_2) = f(\sigma(p_1), \sigma(p_2)), \tag{3.13}$$

where $\sigma: P \to Q$ *and* $f \in \mathbf{I}(Q)$, *is a contravariant functor.*

(ii) *If* $\rho: P \to Q$ *is a function and* $\mathbf{I}(\rho)$ (*as defined above*) *is a homomorphism from* $\mathbf{I}(Q)$ *to* $\mathbf{I}(P)$, *then* ρ *is a proper map.*

PROOF. (i) If $f \in \mathbf{I}(Q)$ and $p_1, p_2 \in P$, then $[\mathbf{I}(\sigma)(f)](p_1, p_2) \neq 0$ implies that $f(\sigma(p_1), \sigma(p_2)) \neq 0$, which implies that $\sigma(p_1) \leqq \sigma(p_2)$ (since $f \in \mathbf{I}(Q)$) and hence (by condition (b)), that $p_1 \leqq p_2$, and so $\mathbf{I}(\sigma)(f) \in \mathbf{I}(P)$. Thus, $\mathbf{I}(\sigma)$ is a mapping from $\mathbf{I}(Q)$ to $\mathbf{I}(P)$.

It is clearly a linear map. Furthermore, $\mathbf{I}(\sigma)$ takes the identity of $\mathbf{I}(Q)$ to the identity of $\mathbf{I}(P)$, since by condition (a)

$$[\mathbf{I}(\sigma)(\delta_Q)](p_1, p_2) = \delta_Q(\sigma(p_1), \sigma(p_2)) = \delta_P(p_1, p_2). \tag{3.14}$$

Finally, $\mathbf{I}(\sigma)$ preserves multiplication, since

$$\begin{aligned} [\mathbf{I}(\sigma)(f*g)](p_1, p_2) &= f*g(\sigma(p_1), \sigma(p_2)) \\ &= \sum_{q\in[\sigma(p_1), \sigma(p_2)]} f(\sigma(p_1), q)g(q, \sigma(p_2)) \\ &= \sum_{p\in[p_1, p_2]} f(\sigma(p_1), \sigma(p))g(\sigma(p), \sigma(p_2)) \\ &= \sum_{p\in[p_1, p_2]} [\mathbf{I}(\sigma)(f)](p_1, p)\cdot[\mathbf{I}(\sigma)(g)](p, p_2) \\ &= ([\mathbf{I}(\sigma)(f)]*[\mathbf{I}(\sigma)(g)])(p_1, p_2). \end{aligned} \tag{3.15}$$

(The third equality follows from (c′).)

Thus, $\mathbf{I}(\sigma)$ is an algebra homomorphism from $\mathbf{I}(Q)$ to $\mathbf{I}(P)$. To verify that $\mathbf{I}$ is a functor, it remains to show that $\mathbf{I}(id_P) = id_{\mathbf{I}(P)}$, where id_P is the identity map on P, and where $id_{\mathbf{I}(P)}$ is the identity map on $\mathbf{I}(P)$, and that $\mathbf{I}(\sigma\circ\tau) = \mathbf{I}(\tau)\circ\mathbf{I}(\sigma)$ when the composition is defined; but these are clear.

(ii) Now, let $\rho: P \to Q$ be a function for which $\mathbf{I}(\rho)$ is a homomorphism from $\mathbf{I}(Q)$ to $\mathbf{I}(P)$. Then

$$\begin{aligned} \delta_Q(\rho(p_1), \rho(p_2)) &= [\mathbf{I}(\rho)(\delta_Q)](p_1, p_2) \\ &= \delta_P(p_1, p_2) \end{aligned} \tag{3.16}$$

since $\mathbf{I}(\rho)$ is a homomorphism, so that ρ is one to one, that is, ρ satisfies (a).

That ρ satisfies (b) follows from the fact that if $\rho(p_1) \leqq \rho(p_2)$, then $\zeta_Q(\rho(p_1), \rho(p_2)) = 1$; that is, $[\mathbf{I}(\rho)(\zeta_Q)](p_1, p_2) = 1$, and so $p_1 \leqq p_2$, since $\mathbf{I}(\rho)(\zeta_Q) \in \mathbf{I}(P)$. Finally, let $q_1 = \rho(p_1)$, $q_2 = \rho(p_2)$, $q_1 \leqq q_2$, and $q \in [q_1, q_2]$. Then we have

$$(3.17)\qquad \sum_{p\in[p_1,p_2]} \delta_{q_1,q}(q_1,\rho(p))\cdot\delta_{q,q_2}(\rho(p),q_2)$$
$$= \sum_{p\in[p_1,p_2]} [\mathbf{I}(\rho)(\delta_{q_1,q})](p_1,p)\cdot[\mathbf{I}(\rho)(\delta_{q,q_2})](p,p_2)$$
$$= ([\mathbf{I}(\rho)(\delta_{q_1,q})]*[\mathbf{I}(\rho)(\delta_{q,q_2})])(p_1,p_2)$$
$$= [\mathbf{I}(\rho)(\delta_{q_1,q}*\delta_{q,q_2})](p_1,p_2) = \delta_{q_1,q}*\delta_{q,q_2}(q_1,q_2) = 1.$$

Thus, $\rho(p) = q$ for some $p \in [p_1, p_2]$, and so ρ satisfies (c). *Q.E.D.*

We conclude with a number of examples of proper maps.

EXAMPLE 3.1. Any one to one map from an ordered set to an antichain is a proper map.

EXAMPLE 3.2. The proper maps from the integers (with the standard ordering) to themselves are those of the form $f(x) = x + k$, where k is some fixed integer.

EXAMPLE 3.3. If P is any finite or locally finite countable ordered set, a proper map onto P from a chain of integers is obtained by labeling the elements of P with integers so that $p_i < p_j$ only if $i < j$, and then taking the map $\sigma(i) = p_i$. A result of Hinrichs [33] guarantees that such a labeling of P exists.

3.6. *Isomorphic incidence algebras.* In this subsection, we prove the result of Stanley [58] that an ordered set P is uniquely determined by its incidence algebra $\mathbf{I}(P)$.

THEOREM 3.2. *Let P and Q be locally finite ordered sets. If $\mathbf{I}(P)$ and $\mathbf{I}(Q)$ are isomorphic as K algebras (even as rings), then P and Q are isomorphic.*

PROOF. We shall show how the ordered set P can be uniquely recovered from the ring $\mathbf{I}(P)$. If R is the radical of $\mathbf{I}(P)$, then $\mathbf{I}(P)/R$ is isomorphic to a direct product $\Pi_{x\in P}K_x$ of copies of the ground field $K = K_x$, one for each element x of P. The K_x are intrinsically characterized as being the *minimal components* of $\mathbf{I}(P)/R$. Note that the element e_x is an idempotent whose image in $\mathbf{I}(P)/R$ is the identity element of K_x. Moreover, the e_x are orthogonal, that is, $e_xe_y = e_ye_x$ if $x \neq y$.

Define an order relation P' on the e_x as follows: $e_x \leqq e_y$ if and only if $e_x\mathbf{I}(P)e_y \neq \{0\}$. It is clear from equation (3.6) that $e_x \leqq e_y$ if and only if $x \leqq y$ in P. Thus, $P' \simeq P$.

The proof will be complete if we can show that given any set $\{f_x | x \in P\}$ of orthogonal idempotents in $\mathbf{I}(P)$ such that the image of f_x in $\mathbf{I}(P)/R$ is the identity element of K_x, then the order relation defined on the f_x in analogy to the e_x is isomorphic to P'.

It suffices to prove that there is an automorphism σ of $\mathbf{I}(P)$ such that $\sigma(e_x) = f_x$ for all $x \in P$. We will explicitly exhibit an inner automorphism $\sigma(g) = hgh^{-1}$, for some fixed invertible $h \in \mathbf{I}(P)$, with the desired property. Define

$$(3.18)\qquad h = \sum_{x\in P} f_x e_x.$$

Clearly, h is a well-defined invertible element of $\mathbf{I}(P)$, since $h(x, y) = f_y(x, y)$.

Now by orthogonality of the e_x and the f_x, we have $he_x = f_x e_x$ and $f_x h = f_x e_x$. Hence, $he_x h^{-1} = f_x$ for all $x \in P$, and the proof is complete.

4. Reduced incidence algebras

4.1. *Order compatible relations.* In most problems of enumeration it is not the full incidence algebra that is required, but only a much smaller subalgebra of it; for example, the algebras of ordinary, exponential, Eulerian and Dirichlet generating functions are obtained by taking subalgebras of suitable incidence algebras (see Examples 4.1 through 4.12). These subalgebras are obtained by taking suitable equivalence relations on segments of a locally finite ordered set P, and then considering functions which take the same values on equivalent segments. We are therefore led to the following.

DEFINITION 4.1. *An equivalence relation* $\sim$ *defined on the segments of a locally finite ordered set* P *is said to be order compatible (or simply compatible) when it satisfies the following condition: if* f *and* g *belong to the incidence algebra* $\mathbf{I}(P)$ *and* $f(x, y) = f(u, v)$ *as well as* $g(x, y) = g(u, v)$ *for all pairs of segments such that* $[x, y] \sim [u, v]$, *then* $(f*g)(x, y) = (f*g)(u, v)$.

EXAMPLE 4.1. Set $[x, y] \sim [u, v]$ whenever the two segments are isomorphic; then $\sim$ is an order compatible equivalence relation.

There is in general no simple criterion, expressible in terms of the partial ordering, to decide when an equivalence relation on segments is order compatible. A useful sufficient criterion is the following.

PROPOSITION 4.1 (D. A. Smith). *An equivalence relation* $\sim$ *on the segments of an ordered set* P *is order compatible if whenever* $[x, y] \sim [u, v]$ *there exists a bijection* φ, *depending in general upon* $[x, y]$, *of* $[x, y]$ *onto* $[u, v]$ *such that* $[x_1, y_1] \sim [\varphi(x_1), \varphi(y_1)]$ *for all* x_1, y_1 *such that* $x \leqq x_1 \leqq y_1 \leqq y$.

The easy proof is left to the reader.

We shall be first concerned with the family of all order compatible equivalence relations on P. Its elementary structure is given by the following.

PROPOSITION 4.2. *The family of order compatible equivalence relations on a locally finite ordered set* P, *ordered by refinement, is a complete lattice* $C(P)$, *in which joins coincide with joins in the lattice* $L(P)$ *of all equivalence relations (partitions) on the segments of* P.

PROOF. In proving that joins in $C(P)$ coincide with joins in $L(P)$, it is convenient to use the language of partitions of the set of segments of P. Thus, let $\mathbf{F}$ be a family of partitions each of which defines a compatible equivalence relation. Let π be the join of $\mathbf{F}$, defining an equivalence relation $\sim$. Suppose that $f(x, y) = g(u, v)$ for all pairs of segments such that $[x, y] \sim [u, v]$. Then *a fortiori* for all $\simeq$ in $\mathbf{F}$, we shall have $f(x, y) = g(u, v)$ for all pairs of segments such that $[x, y] \simeq [u, v]$. It follows that $(f*g)(x, y) = (f*g)(u, v)$ for all such pairs of intervals. But, by definition of join of partitions, $[x, y] \sim [u, v]$ if and only if there is a sequence $\simeq_1, \simeq_2, \cdots, \simeq_n$ in $\mathbf{F}$ and segments $[x_i, y_i]$ such that $[x, y] \simeq_1 [x_1, y_1] \cdots \simeq_{n-1} [x_{n-1}, y_{n-1}] \simeq_n [u, v]$. It follows that $f(x, y) =$

$f(x_1, y_1) = \cdots$, similarly for g. Recalling that $\simeq_1$ is order compatible, we have $(f*g)(x, y) = (f*g)(x_1, y_1)$, and so forth, giving finally $(f*g)(x, y) = (f*g)(u, v)$.

The ordered set $C(P)$ has a 0, namely, the equivalence relation where no two distinct segments are equivalent, and therefore arbitrary meets exist by a simple result of lattice theory. *Q.E.D.*

Observe that meets in $C(P)$ do not in general coincide with meets in $L(P)$, so that $C(P)$ is not a sublattice of $L(P)$. Unless P is finite, it follows that $C(P)$ is not locally finite, for it is easy to stretch an infinite chain between 0 and 1 in $C(P)$ by successively "identifying" pairs of segments $[x, x]$ and $[u, u]$.

It is tempting to presume that the maximal element I of $C(P)$ is the equivalence relation described in Example 4.1, where every pair of isomorphic segments is equivalent. Surprisingly, this presumption is not generally true, even for finite ordered sets, as the following example indicates.

EXAMPLE 4.2. Let P be the ordered set obtained by taking the lattices L_1 and L_2 of subspaces of two nonisomorphic finite projective planes of the same order and identifying the top of L_1 with the bottom of L_2. Define $[x, y] \sim [u, v]$ whenever the two segments are isomorphic or whenever $[x, y] \approx L_1$, $[u, v] \approx L_2$.

4.2. *The incidence coefficients.* Let $\sim$ be an order compatible equivalence relation on P, which will remain fixed until further notice. Denote by Greek letters $\alpha, \beta, \cdots$ the nonempty equivalence classes of segments of P relative to $\sim$, and call them *types* (relative to $\sim$) for short.

Consider the set of all functions f defined on the set of types, with addition defined as usual, and multiplication $f*g = h$ defined as follows:

$$h(\alpha) = \sum \begin{bmatrix} \alpha \\ \beta, \gamma \end{bmatrix} f(\beta)g(\gamma). \tag{4.1}$$

The sum ranges over all ordered pairs β, γ of types. The brackets on the right are called the *incidence coefficients*, and are defined as follows:

$$\begin{bmatrix} \alpha \\ \beta, \gamma \end{bmatrix} \tag{4.2}$$

stands for the number of distinct elements z in a segment $[x, y]$ of type α, such that $[x, z]$ is of type β and $[z, y]$ is of type γ.

To see that the incidence coefficients are well defined, define $h_\delta \in \mathbf{I}(P)$ by

$$h_\delta(x, y) = \begin{cases} 1 & \text{if } [x, y] \text{ is of type } \delta, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

If $[u, v]$ is of type α, then clearly

$$(h_\beta * h_\gamma)(u, v) = \begin{bmatrix} \alpha \\ \beta, \gamma \end{bmatrix}. \tag{4.4}$$

Since $\sim$ is order compatible, the left side of (4.4) is independent of whichever interval $[u, v]$ of type α is chosen, so that the incidence coefficients are well defined. The incidence coefficients are a generalization of the classical *binomial coefficients*, as the examples below will show. The corresponding generalization of the algebra of generating functions is given next.

PROPOSITION 4.3. *Let P be a locally finite ordered set, together with a compatible equivalence relation $\sim$ on the segments of P. Then the set of all functions defined on types forms an associative algebra with identity, with the product defined by* (4.1), *called the* reduced incidence algebra $\mathbf{R}(P, \sim)$ modulo the equivalence relation $\sim$. *The algebra $\mathbf{R}(P, \sim)$ is isomorphic to a subalgebra of the incidence algebra of P.*

To complete the proof (much of which has already been given above), all that needs to be shown is that $\mathbf{R}(P, \sim)$ is isomorphic to a subalgebra of $\mathbf{I}(P)$ which contains δ. This will imply that the algebra $\mathbf{R}(P, \sim)$ is associative.

For $f \in \mathbf{R}(P, \sim)$, define $\hat{f} \in \mathbf{I}(P)$ as follows: $\hat{f}(x, y) = f(\alpha)$ if the segment $[x, y]$ is of type α. The only properties to be checked are that the mapping is an isomorphism and that $\delta = \hat{f}$ for some $f \in \mathbf{R}(P, \sim)$. Since each type is by definition nonempty it follows that $f \rightarrow \hat{f}$ is well defined; it is obviously one to one. Furthermore, from the definition of the incidence coefficients, we find immediately that the product is

$$\hat{h}(x, y) = \sum_{x \leqq z \leqq y} \hat{f}(x, z)\hat{g}(z, y), \tag{4.5}$$

and thus coincides with the definition (4.1) of the product in $\mathbf{R}(P, \sim)$. The fact that $\delta = \hat{f}$ for some $f \in \mathbf{R}(P, \sim)$ follows from part (i) of the following lemma.

LEMMA 4.1. *Let $\sim$ be an order compatible equivalence relation on the segments of P, and let $[x, y] \sim [u, v]$. Then*

(i) $\nu([x, y]) = \nu([u, v])$, *where $\nu([x, y])$ = number of z in $[x, y]$;*

(ii) *for every n, $[x, y]$ and $[u, v]$ have the same number of maximal chains of length n.*

PROOF. Part (i) follows from the fact that $\nu([x, y]) = \zeta^2(x, y)$ and that ζ is constant on equivalence classes of $\sim$.

From (i), it follows that the function h defined by

$$h(x, y) = \begin{cases} 1 & \text{if } \nu([x, y]) = 2, \text{ that is, } y \text{ covers } x, \\ 0 & \text{otherwise,} \end{cases} \tag{4.6}$$

is constant on equivalence classes of $\sim$; hence, so is h^n for every n, which proves (ii).

COROLLARY 4.1. *If for all types α, β, γ we have $\left[{\alpha \atop \beta, \gamma}\right] = \left[{\alpha \atop \gamma, \beta}\right]$, then the reduced incidence algebra $\mathbf{R}(P, \sim)$ is commutative.*

This follows immediately from definition (4.1) of the product.

Now let $\sim$ and $\simeq$ be two order compatible equivalence relations on the segments of P. Suppose that $[x, y] \sim [u, v]$ implies $[x, y] \simeq [u, v]$. Then, much as in the preceding proposition, $\mathbf{R}(P, \simeq)$ is isomorphic to a subalgebra of $\mathbf{R}(P, \sim)$; the isomorphism is obtained as follows: Let $\hat{\alpha}$ be a type relative to the equivalence relation $\simeq$. For $f \in \mathbf{R}(P, \simeq)$, set $\hat{f} \in \mathbf{R}(P, \sim)$ to be $\hat{f}(\alpha) = f(\hat{\alpha})$, where α is any type in $\mathbf{R}(P, \sim)$ such that the segments of type α are of type $\hat{\alpha}$ in $\mathbf{R}(P, \simeq)$.

Furthermore, $\mathbf{R}(P, \sim)$ strictly contains a natural isomorphic image of $\mathbf{R}(P, \simeq)$ unless $\sim$ equals $\simeq$, as is immediately seen by considering functions equal to one on a given type, and zero elsewhere. Thus, the lattice $C(P)$ is anti-isomorphic to the lattice of reduced incidence algebras, ordered by containment.

If $\sim$ is as in Example 4.1, then we call $\mathbf{R}(P, \sim)$ the *(standard) reduced incidence algebra* $\mathbf{R}(P)$; if $\sim$ is the maximal element of the lattice $C(P)$, we call $\bar{\mathbf{R}}(P) = \mathbf{R}(P, \sim)$ the *maximally reduced* incidence algebra.

PROPOSITION 4.4. *If $\sim$ is a finer order compatible equivalence relation than $\simeq$, and for $f \in \mathbf{R}(P, \simeq)$ the image $\hat{f}$ (as above) in $\mathbf{R}(P, \sim)$ is invertible in $\mathbf{R}(P, \sim)$, then f is invertible in $\mathbf{R}(P, \simeq)$.*

PROOF. Identify both algebras with subalgebras of $\mathbf{I}(P)$, as in the proof of Proposition 4.3, so that $f = \hat{f}$. We must show that f^{-1} is constant on $\simeq$ equivalent segments. Since f is invertible, it takes nonzero values on one point segments. Let $d \in \mathbf{I}(P)$ be the function which equals f on one point segments and is zero elsewhere. Then d is constant on $\simeq$ equivalence classes (by Lemma 4.1 (i)), and d^{-1} is also, since d^{-1} is the inverse of d on one point segments and zero elsewhere. Let $g = f - d$. Then $g \in \mathbf{R}(P, \simeq)$ and

$$\begin{aligned} f^{-1} &= (d + g)^{-1} = \big(1 + (d^{-1} * g)\big)^{-1} * d^{-1} \\ &= \big(1 - (d^{-1} * g) + (d^{-1} * g)^2 - (d^{-1} * g)^3 + \cdots\big) * d^{-1} \end{aligned} \tag{4.7}$$

which is well defined, since $d^{-1} * g$ is zero on one point segments; and hence, $f^{-1} \in \mathbf{R}(P, \simeq)$.

It follows that the *zeta function* and the *Möbius function* belong to all reduced incidence algebras.

We conclude with a simple characterization of reduced incidence algebras. In the finite case, it is purely algebraic, but in the infinite case, topological considerations come in. Recall that the *Schur product* of two elements f, g of $\mathbf{I}(P)$ is the element h defined by

$$h(x, y) = f(x, y) \cdot g(x, y) \tag{4.8}$$

for all x, y in P.

THEOREM 4.1. *Let P be a locally finite ordered set, and A a subalgebra of $\mathbf{I}(P)$ having the same identity as $\mathbf{I}(P)$. If P is finite, then A is a reduced incidence algebra of P if and only if A contains ζ and is closed under Schur multiplication. If P is infinite, then A is a reduced incidence algebra if and only if A contains ζ, is closed under Schur multiplication, and is closed in the standard topology.*

PROOF. The necessity of the conditions is evident, with the possible exception that A must be topologically closed. But if $f \in \mathbf{I}(P)$ is in the topological closure of A then it must clearly be constant on equivalence classes and so $f \in A$.

Now, assume P finite, and let A be a subalgebra of $\mathbf{I}(P)$ containing ζ and closed under Schur multiplication. Let $\sim$ be the equivalence relation on segments of P defined by $[x, y] \sim [u, v]$ if and only if $f(x, y) = f(u, v)$ for all $f \in A$. Once we have shown that the set of all functions constant on the equivalence classes of $\sim$ is precisely A, then it will follow that $\sim$ is order compatible (since A is closed under convolution) and that A is $\mathbf{R}(P, \sim)$. Let $\beta_1, \cdots, \beta_n$ be the equivalence classes of $\sim$. For each $i \neq j$, let $h_{i,j}$ be an element of A such that $h_{i,j}(\beta_i) \neq h_{i,j}(\beta_j)$, and let

$$(4.9) \qquad \bar{h}_{i,j} = \frac{h_{i,j} - c_{i,j}\zeta}{b_{i,j} - c_{i,j}},$$

where $b_{i,j} = h_{i,j}(\beta_i)$, $c_{i,j} = h_{i,j}(\beta_j)$. Then $\delta_i = \Pi_{j \neq i} \bar{h}_{i,j}$ (Schur multiplication) is in A, and δ_i is the indicator function of β_i. Now any function which is constant on equivalence classes is a linear combination of the functions δ_i, and hence is in A, proving our result.

A slight modification proves the infinite case, since δ_i is the limit of finite Schur products of $\bar{h}_{i,j}$ for $j \neq i$, and every function constant on equivalence classes is a limit of finite linear combinations of indicator functions. *Q.E.D.*

The assumption that A be topologically closed in the infinite case is necessary, as the following examples demonstrate.

EXAMPLE 4.3. Let P be an infinite locally finite ordered set in which there is a finite upper bound on the size of segments of P. Then the subset A of $\mathbf{I}(P)$ consisting of all functions which take only finitely many values is a subalgebra closed under Schur multiplication and containing ζ, but is clearly not a reduced incidence algebra, since the equivalence relation it generates is the trivial one, while A is not all of $\mathbf{I}(P)$.

EXAMPLE 4.4. Let P contain chains of arbitrarily large (finite) length, and let A be the closure under the operations of scalar multiplication, addition, convolution, and Schur multiplication of $\{\delta, \zeta\}$ in $\mathbf{I}(P)$. Then A is a subalgebra closed under Schur multiplication and containing ζ, but is not a reduced incidence algebra, since by Lemma 4.1 (ii), any reduced incidence algebra of P must have uncountable vector space dimension over the ground field, while A has countable dimension.

We now consider various examples of reduced incidence algebras and their connection with classical combinatorial theory.

EXAMPLE 4.5. *Formal power series.* Let P be the set of nonnegative integers in their natural ordering. The incidence algebra of P is evidently the algebra of upper triangular infinite matrices. On the other hand, we shall now see that the standard reduced incidence algebra $\mathbf{R}(P)$ is isomorphic to the algebra of formal power series.

Indeed, an element of $\mathbf{R}(P)$ is uniquely determined by a sequence $\{a_n : n = 0, 1, 2, \cdots\}$ of real numbers, by setting $f(i, j) = a_{j-i}$, for $i \leqq j$. The product of such an element by another element $g(i, j) = b_{j-i}$ of the same form is an element h of $\mathbf{I}(P)$ obtained by the convolution rule

$$(4.10) \qquad h(i, j) = \sum_{i \leqq k \leqq j} f(i, k)g(k, j) = \sum_{i \leqq k \leqq j} a_{k-i}b_{j-k}.$$

Setting $r = k - i$ and $j - i = n$, we obtain $h(i, j) = \Sigma_{r=0}^{n} a_r b_{n-r} = c_n$. In other words, h is the element of the reduced incidence algebra which is obtained by convoluting the sequences $\{a_n\}$ and $\{b_n\}$. It follows that the map of power series into $\mathbf{R}(P)$ defined by

$$(4.11) \qquad F(x) = \sum_{n=0}^{\infty} a_n x^n \to f(i, j) = a_{j-i}, \qquad j \geqq i$$

is an isomorphism. Under this isomorphism, the zeta function corresponds to $1/(1 - x)$, and the Möbius function corresponds to the formal power series $1 - x$. The incidence coefficients equal either 0 or 1.

EXAMPLE 4.6. *Exponential power series.* Let $B(S)$ be the family of all finite subsets of a countable set S, ordered by inclusion. We shall prove that the reduced incidence algebra of $B(S)$ is isomorphic to the algebra of exponential formal power series under formal multiplication, that is, a series of the form

$$(4.12) \qquad F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n, \qquad G(x) = \sum_{n=0}^{\infty} \frac{b_n}{n!} x^n, \qquad H(x) = \sum_{n=0}^{\infty} \frac{c_n}{n!} x^n.$$

It is immediately verified that the product $FG = H$ of two such formal power series amounts to taking the *binomial convolution* of their coefficients,

$$(4.13) \qquad c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}.$$

We obtain an isomorphism between the algebra of exponential formal power series and the reduced incidence algebra of $B(S)$ by setting

$$(4.14) \qquad F(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n \to f(A, B) = a_{\nu(B-A)}, \qquad A \subseteqq B,$$

where A and B are finite subsets of S and $\nu(B - A)$ denotes as usual the number of elements of the set $B - A$. The zeta function corresponds to e^x, and the Möbius function to e^{-x}. The Möbius inversion formula reduces to the principle of inclusion–exclusion, that is, to multiplication by e^{-x}.

The incidence coefficients coincide with the binomial coefficients, and the types naturally coincide with the integers.

EXAMPLE 4.7. Let G be the additive group of rational numbers modulo 1, and let $\mathbf{L}(G)$ be the lattice of subgroups excluding G itself. It is well known that every proper subgroup of G is finite cyclic. Let $[X, Y] \sim [U, V]$ in $\mathbf{L}(G)$ when the quotient group Y/X is isomorphic to V/U. The types correspond naturally to the positive integers; the incidence coefficients equal zero or one, and the product in $\mathbf{R}(\mathbf{L}(G), \sim)$ is given by the *Dirichlet convolution* $c_n = \Sigma_{ij=n} a_i b_j$. Thus, $\mathbf{R}(\mathbf{L}(G), \sim)$ is isomorphic to the algebra of *formal Dirichlet series*

$$(4.15) \qquad \sum_{n \geqq 1} \frac{a_n}{n^s} = f(s)$$

under ordinary multiplication.

EXAMPLE 4.8. Let P be the set of positive integers, ordered by divisibility, and let $\sim$ be the equivalence relation defined by $[a, b] \sim [m, n]$ if and only if $b/a = n/m$. Then, as in the previous example, $\mathbf{R}(P, \sim)$ is easily seen to be isomorphic to the algebra of formal Dirichlet series. The standard reduced incidence algebra is isomorphic to a subalgebra of the algebra of formal Dirichlet series, namely to those series $\Sigma_{n \geqq 1} a_n/n^s$ in which $a_k = a_n$ if $k = p_1^{a_1} p_2^{a_2} \cdots$, and $n = p_1^{b_1} p_2^{b_2} \cdots$, where $p_1, p_2, \cdots$ are the primes, the a_i and b_i are nonnegative integers, and the b_i are obtained by permuting the a_i.

EXAMPLE 4.9. Let V be a vector space of countable dimension over $GF(q)$, and let $\mathbf{L}(V)$ be the lattice of finite dimensional subspaces. Let $\sim$ be the equivalence relation defined by $[S, T] \sim [X, Y]$ if and only if $T/S \approx Y/X$, that is, $\dim T - \dim S = \dim Y - \dim X$. Then the types are in one to one correspondence with the integers, and multiplication in $\mathbf{R}(\mathbf{L}(V), \sim)$ is given by

$$(4.16) \qquad (f*g)(n) = \sum_{r=0}^{n} \frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{r-1})}{(q^r - 1)(q^r - q) \cdots (q^r - q^{r-1})} f(r) g(n - r)$$

$$= \sum_{r=0}^{n} \frac{(1 - q^n)(1 - q^{n-1}) \cdots (1 - q^{n-r+1})}{(1 - q^r)(1 - q^{r-1}) \cdots (1 - q)} f(r) g(n - r).$$

Hence, $\mathbf{R}(\mathbf{L}(V), \sim)$ is isomorphic to the algebra of *Eulerian power series*, the isomorphism being given by

$$(4.17) \qquad f \rightarrow \sum_{n \geqq 0} \frac{f(n)}{(1 - q)(1 - q^2) \cdots (1 - q^n)} x^n.$$

We now present three examples in which we arrive very simply at previously known results by using the reduced incidence algebra. Let P be a locally finite ordered set, and let $c \in \mathbf{I}(P)$ be the function which assigns to a segment $[x, y]$ the total number of chains, $x = x_0 < x_1 < \cdots < x_m = y$. Since $(\zeta - \delta)^k(x, y)$ is the number of chains, $x = x_0 < x_1 < \cdots < x_k = y$, of length k, we have

$$(4.18) \qquad \begin{aligned} c(x, y) &= \sum_{k=0}^{\infty} (\zeta - \delta)^k(x, y) \\ &= [\delta - (\zeta - \delta)]^{-1}(x, y) \\ &= (2\delta - \zeta)^{-1}(x, y). \end{aligned}$$

EXAMPLE 4.10. Let P be as in Example 4.5. Then $c(x, y)$ is the number c_n of *ordered partitions* (or *compositions*) of $n = y - x$, the chain $x = i_0 < i_1 < \cdots < i_k = y$ corresponding to the composition

$$(4.19) \qquad y - x = (i_1 - i_0) + (i_2 - i_1) + \cdots + (i_k - i_{k-1})$$

Hence,

$$(4.20) \qquad \sum_{n=0}^{\infty} c_n x^n = \frac{1}{2 - (1 - x)^{-1}} = \frac{1 - x}{1 - 2x} = 1 + \sum_{n=1}^{\infty} 2^{n-1} x^n,$$

so $c_n = 2^{n-1}$ if $n > 0$ (a well-known result).

EXAMPLE 4.11. Let P be as in Example 4.6. Then $c(x, y)$ is the number f_n of *ordered set partitions* (or *preferential arrangements*) of the set $y - x$, where n is the number of elements in $y - x$. (See Gross [29].) Hence,

$$\sum_{n=0}^{\infty} \frac{f_n}{n!} x^n = \frac{1}{2 - e^x}, \tag{4.21}$$

a basic result of Gross.

EXAMPLE 4.12. Let P be the positive integers ordered by divisibility, with $[u, v] \sim [x, y]$ if $v/u = y/x$. Then $c(x, y)$ is the number $f(n)$ of *ordered factorizations* of $y/x = n$ (into factors >1). Hence,

$$\sum_{n=1}^{\infty} f(n) n^{-s} = \frac{1}{2 - \zeta(s)}, \tag{4.22}$$

a result of Titchmarsh ([59], p. 7).

More generally, the theory of *weighted compositions*, as developed by Moser and Whitney [39] and by Hoggart and Lind [34], can be expressed in terms of the reduced incidence algebra of a chain. Thus, this theory can be extended to other ordered sets in the same way that Examples 4.11 and 4.12 extend the usual concept of composition given in Example 4.10.

5. The large incidence algebras

5.1. *Definitions.* Several enumeration problems lead not to a single ordered set, but to a family of ordered sets having some common features; for example, the family of lattices of partitions of finite sets or the family of all lattices of subgroups of finite abelian groups. It then becomes necessary to extend the notions of incidence algebra and reduced incidence algebra to these situations. Recall that we assume the ground field K to have characteristic 0. This avoids complications inherent in dividing by positive integers, such as $n!$ in exponential generating functions. We are now led to the following setup.

Two ordered sets $(P, \sim)$ and $(Q, \sim)$ each with an order compatible equivalence relation (denoted by the same symbol for convenience) are said to be *isomorphic* when there is an isomorphism ϕ of P to Q which preserves the equivalence relation, that is $[x, y] \sim [u, v]$ in P if and only if $[\phi(x), \phi(y)] \sim [\phi(u), \phi(v)]$ in Q. If S is a segment of P, then the equivalence relation $\sim$ induces a compatible equivalence relation on S. (Note that this conclusion does not hold in general if S is only assumed to be an ordered subset of P.)

Now let **C** be a category whose objects are pairs $(P, \sim)$ as above, where P is a finite ordered set with 0 and 1, and where morphisms ϕ of $(P, \sim)$ into $(Q, \sim)$ are isomorphisms of $(P, \sim)$ onto a segment of $(Q, \sim)$ with the induced equivalence relation (not all such maps need be included in the category as morphisms). It is further assumed that every segment of an object $(P, \sim)$ is in **C**, with the induced equivalence relation. Finally, it is assumed that if ϕ and ψ are two morphisms of $(P, \sim)$ into $(Q, \sim)$ having the segments $[x, y]$ and $[u, v]$ as images, then $[x, y] \sim [u, v]$ in Q.

Under these conditions, we can define the *large incidence algebra* $\mathbf{I}(\mathbf{C})$ of $\mathbf{C}$ as follows: the elements of $\mathbf{I}(\mathbf{C})$ are functions f on the isomorphism classes (in the category $\mathbf{C}$) or "types" of the objects of $\mathbf{C}$ such that $f(\alpha) = f(\beta)$, if some object $(P, \sim)$ contains $\sim$-equivalent segments of types α and β (with values, as usual, in a fixed field). The sum of two such functions is defined as usual, and the product is defined by

$$(f*g)(\alpha) = \sum \begin{bmatrix} \alpha \\ \beta, \gamma \end{bmatrix} f(\beta)g(\gamma), \tag{5.1}$$

where the brackets are taken in any object P belonging to the isomorphism class α. Our assumptions imply that the product is well defined; that is, the product remains the same if it is computed in any object of type α, and also that $f*g$ is in $\mathbf{I}(\mathbf{C})$. Thus, we obtain an algebra which is associative by Proposition 4.3. The functions ζ and δ of the ordinary incidence algebra have obvious counterparts in the large incidence algebra, and the result that a function is invertible if and only if it is nonzero on all types containing one point segments (see *Foundations I*) also carries over. Hence, the Möbius function can be defined as the inverse to the zeta function; and clearly, for each object $([0, 1], \sim)$ of the category, the value of the Möbius function on the type containing $[0, 1]$ equals $\mu(0, 1)$.

Most of the classes of incidence algebras (such as binomial type and Dirichlet type) can be trivially extended to large incidence algebras. Also note that we need make no distinction between reduced and nonreduced large incidence algebras, for the degree of reduction is built into the category itself, depending on the equivalence relations in the objects and on the morphisms.

EXAMPLE 5.1. Let L be a locally finite ordered set. Construct a category $\mathbf{C}$ as follows. The objects are all segments of L and the morphisms are the inclusion maps. The equivalence relation is the trivial one (no two distinct segments are isomorphic in $\mathbf{C}$). Note that two isomorphic segments are not in general isomorphic in $\mathbf{C}$. The large incidence algebra $\mathbf{I}(\mathbf{C})$ is isomorphic to the incidence algebra of L.

EXAMPLE 5.2. Let L be as above; let the objects of $\mathbf{C}$ be again all segments of L, but let the morphisms be all isomorphisms; and let $\sim$ be isomorphism. Then $\mathbf{I}(\mathbf{C})$ is isomorphic to the standard reduced incidence algebra of L.

EXAMPLE 5.3. Let the objects of $\mathbf{C}$ all be finite Boolean algebras; let $\sim$ be isomorphism of segments; and let morphisms all be isomorphisms. Then $\mathbf{I}(\mathbf{C})$ is isomorphic to the algebra of exponential power series of Example 5.6.

In the next three subsections, we consider situations which are better looked at from the point of view of the large incidence algebra than from that of the regular incidence algebra.

5.2. *Partition lattices.* The incidence algebra of the family of all partition lattices of finite sets can be studied by taking the lattice $\Pi(S)$ of all partitions of an infinite set S having exactly one infinite block and finitely many finite blocks, ordered by refinement. However, it is more pleasingly done in the context of the large incidence algebra, as follows.

Let the objects of a category $\mathbf{\Pi}$ be all lattices of partitions of finite sets and all segments thereof, and let the equivalence relation be an isomorphism of segments whose top elements have the same number of blocks. Let the morphisms of $\mathbf{\Pi}$ all be isomorphisms onto a segment such that the top element of a segment has the same number of blocks as the top element of the image segment. It is immediate that $\mathbf{\Pi}$ satisfies the required conditions.

The *class* of a segment $[\sigma, \pi]$ is a sequence of nonnegative numbers $(k_1, \cdots, k_n, \cdots)$, where k_i is the number of blocks in π which are the union of precisely i blocks in σ. It is clear that $k_1 + 2k_2 + 3k_3 + \cdots$ equals the number of blocks in σ and $k_1 + k_2 + k_3 + \cdots$ equals the number of blocks in π, and that a segment of class $(k_1, k_2, \cdots)$ is isomorphic to $\Pi_1^{k_1} \times \Pi_2^{k_2} \times \cdots$, where Π_i is the lattice of partitions of an i set, so it follows that two segments have the same class if and only if they are of the same type in $\mathbf{\Pi}$. We denote by $\binom{n}{k_1, \ldots, k_n}$ the number of elements τ in a segment $[\sigma, \pi]$ of type $(\delta_{0,n}, \delta_{1,n}, \cdots)$ (that is, $[\sigma, \pi]$ is isomorphic to Π_n and σ has n blocks) for which $[\sigma, \tau]$ has type $(k_1, k_2, \cdots, k_n, 0, 0, \cdots)$ (and hence, $[\tau, \pi]$ has type $(\delta_{0,m}, \delta_{1,m}, \cdots)$, where $m = k_1 + \cdots + k_n$). To compute $\binom{n}{k_1, \ldots, k_n}$, first note that any object $[\sigma, \pi]$ of type $k_1 = \cdots = k_{n-1} = 0$, $k_n = 1$ is an upper segment of some finite partition lattice; that is, $\pi = 1$ in some finite lattice of partitions. Thus, it is easy to see that

$$\binom{n}{k_1, \cdots, k_n} = \frac{n!}{1!^{k_1} k_1! \, 2!^{k_2} k_2! \cdots n!^{k_n} k_n!} \tag{5.2}$$

when $k_1 + 2k_2 + \cdots + nk_n = n$, and equals 0 when $k_1 + 2k_2 + \cdots + nk_n \neq n$.

For a partition π of some finite set S, we define the class of π to be the class of the segment $[0, \pi]$ of $\Pi(S)$, as defined above.

The fundamental concept associated with the large incidence algebra $\mathbf{I}(\mathbf{\Pi})$ is that of *multiplicative function.* A function f in $\mathbf{I}(\mathbf{\Pi})$ is said to be *multiplicative* when there is a sequence of constants $(a_1, a_2, a_3, \cdots)$ such that

$$f(\pi, \sigma) = a_1^{k_1} a_2^{k_2} a_3^{k_3} \cdots \tag{5.3}$$

when $[\pi, \sigma]$ is a segment of class $(k_1, k_2, k_3, \cdots)$. The function f is said to be determined by the sequence $(a_1, a_2, \cdots)$. Similarly, a function of one variable $F(\sigma)$ for $\sigma \in \Pi(S)$ for some finite set S is said to be multiplicative when

$$F(\sigma) = a_1^{k_1} a_2^{k_2} \cdots, \tag{5.4}$$

where $(k_1, k_2, \cdots)$ is the class of σ.

The following elementary result is fundamental.

Proposition 5.1. *The convolution of two multiplicative functions is multiplicative.*

Proof. This follows from the fact that if $[\sigma, \pi]$ *is of type* $(k_1, k_2, \cdots)$, *then* $[\sigma, \pi]$ *is isomorphic to* $\Pi_1^{k_1} \times \Pi_2^{k_2} \times \cdots$, *and that if* $f \in \mathbf{I}(P)$ *and* $g \in \mathbf{I}(Q)$ (*where P and Q are any locally finite ordered sets*) *and if* $f \times g \in \mathbf{I}(P \times Q)$ *is defined by* $f \times g((x, x'), (y, y')) = f(x, y) \cdot g(x', y')$, *then* $(f \times g) * (f' \times g') = (f * f') \times (g * g')$.

COROLLARY 5.1. *If $F(\pi)$ is multiplicative and $f(\pi, \sigma)$ is multiplicative, then so are*

$$G(\sigma) = \sum_{\pi \leqq \sigma} F(\pi) f(\pi, \sigma) \tag{5.5}$$

and

$$H(\sigma) = \sum_{\pi \geqq \sigma} f(\sigma, \pi) F(\pi), \tag{5.6}$$

where the sum is taken in the partition lattice containing σ.

EXAMPLE 5.4. The zeta function of $\mathbf{I}(\mathbf{\Pi})$ is multiplicative and is determined by the sequence $(1, 1, 1, \cdots)$. By Proposition 3 of Section 7 of *Foundations I*, the Möbius function of $\mathbf{I}(\mathbf{\Pi})$ is multiplicative, and determined by the sequence $(a_1, a_2, \cdots)$, where $a_n = (-1)^n (n - 1)!$ The delta function δ is multiplicative, determined by $(1, 0, 0, \cdots)$, but $\eta = \zeta - \delta$ is not multiplicative. Hence, the sum of multiplicative functions need not be multiplicative.

Let $\mathbf{M}(\mathbf{\Pi})$ denote the subset of $\mathbf{I}(\mathbf{\Pi})$ consisting of multiplicative functions. By Proposition 5.1, $\mathbf{M}(\mathbf{\Pi})$ is a subsemigroup of the multiplicative semigroup of $\mathbf{I}(\mathbf{\Pi})$. If f is in $\mathbf{M}(\mathbf{\Pi})$, let $f(n)$ denote $f(\Pi_n)$; that is $f(\pi, \sigma)$, where $[\pi, \sigma]$ has class $k_1 = \cdots = k_{n-1} = 0$, $k_n = 1$. Then, for $f, g \in \mathbf{M}(\mathbf{\Pi})$, we get from (5.2) that $(f * g)(n)$ is equal to

$$\sum_{k_1 + 2k_2 + \cdots + nk_n = n} \left(\frac{n!}{1!^{k_1} k_1! 2!^{k_2} k_2! \cdots n!^{k_n} k_n!} \right) f(1)^{k_1} \cdots f(n)^{k_n} g(k_1 + \cdots + k_n). \tag{5.7}$$

THEOREM 5.1. *The semigroup $\mathbf{M}(\mathbf{\Pi})$ is anti-isomorphic to the algebra of all formal exponential power series with zero contant term over K in a variable x, under the operation of composition. The anti-isomorphism is given by $f \to F_f$, where*

$$F_f(x) = \sum_{n=1}^{\infty} \frac{f(n)}{n!} x^n. \tag{5.8}$$

Thus, $F_{f * g}(x) = F_g(F_f(x))$.

PROOF. Clearly, the map defined by (5.8) is a bijection, so we need only check that multiplication is preserved. Now,

$$F_g(F_f(x)) = \sum_{j=1}^{\infty} \frac{g(j)}{j!} \left(\sum_{i=1}^{\infty} \frac{f(i)}{i!} x^i \right)^j. \tag{5.9}$$

The coefficient of x^n in the expansion of $(\Sigma_{i=1}^{\infty} (f(i)/i!) x^i)^j$ is

$$\sum_{n_1 + n_2 + \cdots + n_j = n} \frac{f(n_1) \cdots f(n_j)}{n_1! \cdots n_j!} = \sum \frac{j!}{k_1! \cdots k_n!} \frac{f(1)^{k_1} \cdots f(n)^{k_n}}{1!^{k_1} \cdots n!^{k_n}}, \tag{5.10}$$

where the summation is taken over $k_1 + 2k_2 + \cdots + nk_n = n$, $k_1 + \cdots + k_n = j$, since there are $j!/k_1! \cdots k_n!$ ways of ordering the partition $k_1 + 2k_2 + \cdots + nk_n = n$. When we multiply (5.10) by $g(j)/j!$ and sum over all j, we get (5.7), and the proof follows.

EXAMPLE 5.5. Under the isomorphism of the proposition, the zeta function corresponds to $e^x - 1$, and the delta function to x, so the Möbius function corresponds to the power series F such that $F(e^x - 1) = x$, that is, to $\log(1 + x)$. Hence, $\mu(0, 1) = (-1)^{n-1}(n - 1)!$ for $[0, 1] = \Pi_n$. This is yet another way of determining the Möbius functions for lattices of partitions.

COROLLARY 5.2. *Let f be a multiplicative function of one variable determined by the sequence $(a_1, a_2, \cdots)$. For every positive integer n, let*

$$b_n = \sum_{\pi\in\Pi_n} f(\pi), \qquad q_n = \sum_{\pi\in\Pi_n} f(\pi)\mu(\pi, 1). \tag{5.11}$$

Then

$$1 + \sum_{n=1}^{\infty} \frac{b_n x^n}{n!} = \exp\left\{a_1 x + \frac{a_2 x^2}{2!} + \frac{a_3 x^3}{3!} + \cdots\right\} \tag{5.12}$$

and

$$\sum_{n=1}^{\infty} \frac{q_n}{n!} x^n = \log\left(1 + a_1 x + \frac{a_2 x^2}{2!} + \frac{a_3 x^3}{3!} + \cdots\right). \tag{5.13}$$

PROOF. For (5.12), let $\bar{f}$ be the function in $\mathbf{M}(\mathbf{\Pi})$ determined by $(a_1, a_2, \cdots)$, and let $b = \bar{f} * \zeta$. Then $b_n = b(n)$ for all $n \geqq 1$, so

$$\begin{aligned} 1 + \sum_{n=1}^{\infty} \frac{b_n x^n}{n!} &= 1 + F_b(x) \\ &= 1 + F_{\bar{f}*\zeta}(x) = 1 + F_\zeta(F_{\bar{f}}(x)) \\ &= \exp\left\{a_1 x + \frac{a_2 x^2}{2!} + \cdots\right\}. \end{aligned} \tag{5.14}$$

For (5.13) let $q = \bar{f} * \mu$, and the proof follows as for (5.12).

We now work out various examples using the above results.

EXAMPLE 5.6 (Waring's formula). Let D and R be finite sets, and label the elements of R by different letters of the alphabet: $x, y, \cdots, z$. To every function $f\colon D \to R$, we associate a monomial $\gamma(f) = x^i y^j \cdots z^k$, where i is the number of elements of D mapped to the element of R labelled x, and so forth; and to every set E of functions from D to R, we associate a polynomial $\gamma(E)$, the sum of $\gamma(f)$ for f ranging over E; $\gamma(E)$ is called the *generating function* of the set E.

For every partition π of the set D, let $A(\pi)$ be the generating function of the set of all functions $f\colon D \to R$ whose kernel (that is, the partition of D whose blocks are the inverse images of elements of R) is π. Let $S(\pi)$ be the generating function of the set of functions whose kernel is some partition $\sigma \geqq \pi$. Clearly, we have $S(\pi) = \Sigma_{\sigma\geqq\pi} A(\sigma)$, from which, by Möbius inversion, we have $A(\pi) = \Sigma_{\sigma\geqq\pi} S(\sigma)\mu(\pi, \sigma)$; and setting $\pi = 0$, we have

$$A(0) = \sum_{\sigma\in\Pi(D)} S(\sigma)\mu(0, \sigma). \tag{5.15}$$

Now assume that D has n elements and that R is larger than D. The polynomial $A(0)$ is the generating function of the set of all one to one functions;

and hence, every term of $A(0)$ is a product of n distinct variables taken among $x, y, \cdots, z$. Furthermore, every product of n distinct variables among $x, y, \cdots, z$ appears $n!$ times as a term in $A(0)$. Thus, $A(0)$ is simply $n! \cdot a_n$, where a_n is the *elementary symmetric function* of degree n in the variables $x, y, \cdots, z$.

Next, if the partition σ has class $(k_1, k_2, \cdots, k_n)$, we claim that

$$S(\sigma) = (x + y + \cdots + z)^{k_1}(x^2 + y^2 + \cdots + z^2)^{k_2} \cdots (x^n + y^n + \cdots + z^n)^{k_n}, \tag{5.16}$$

that is, using the standard notation $s_k = x^k + y^k + \cdots + z^k$,

$$S(\sigma) = s_1^{k_1} s_2^{k_2} \cdots s_n^{k_n}. \tag{5.17}$$

To see this, let $\bar{S}(\sigma)$ be the set of all functions with kernel $\pi \geqq \sigma$, and let $B_1, \cdots, B_k$ be the blocks of the partition σ. Then $\bar{S}(\sigma)$ is the product $U_1 \times \cdots \times U_k$ of the sets U_i, where U_i is the set of all functions from B_i to R taking only one value. It follows that $S(\sigma) = \gamma(\bar{S}(\sigma)) = \gamma(U_1)\gamma(U_2) \cdots \gamma(U_k)$. The generating function $\gamma(U_i)$ is simply $x^k + y^k + \cdots + z^k$ if B_i has k elements, and this completes the verification.

We thus see that (5.15) reduces to the classical formula of Waring, expressing the elementary symmetric functions in terms of sums of powers.

EXAMPLE 5.7. Let V be a finite set of n elements ("vertices"). We count the number C_n of connected graphs whose vertex set is V. To every graph G, we can associate a partition $\pi(G)$ of the set V, the blocks of $\pi(G)$ being connected components of G. A graph is connected if and only if $\pi(G) = 1$, the partition with only one block. For every partition π of V, let $C(\pi)$ be the number of graphs G with $\pi(G) = \pi$, and let $D(\pi)$ be the number of graphs G with $\pi(G) \leqq \pi$. Let a_n be the total number of graphs whose vertex set is V; a simple enumeration gives

$$a_n = 2^{\binom{n}{2}}. \tag{5.18}$$

If $B_1, B_2, \cdots, B_k$ are the blocks of π and $D(B_i)$ is the total number of graphs on the block B_i, then clearly $D(\pi) = D(B_1)D(B_2) \cdots D(B_n)$. Hence, if the class of the partition π is $(k_1, k_2, \cdots, k_n)$, we have

$$D(\pi) = a_1^{k_1} a_2^{k_2} \cdots a_n^{k_n} = t\left(k_1\binom{1}{2} + k_2\binom{2}{2} + k_3\binom{3}{2} + \cdots + k_n\binom{n}{n}\right), \tag{5.19}$$

where $t(x) = 2^x$ and $\binom{1}{2} = 0$ by convention. Furthermore, $D(\pi) = \Sigma_{\sigma \leqq \pi} C(\sigma)$ as follows immediately from the definitions. By the Möbius inversion followed by setting $\pi = 1$, we obtain the identity

$$\begin{aligned} C_n = C(1) &= \sum_{\sigma \in \Pi_n} D(\sigma)\mu(\sigma, 1) \\ &= \sum_{k_1 + 2k_2 + \cdots + nk_n = n} t\left(k_1\binom{1}{2} + k_2\binom{2}{2} + \cdots + k_n\binom{n}{2}\right) \\ &\qquad \cdot(-1)^{k_1 + k_2 + \cdots + k_n - 1}(k_1 + k_2 + \cdots + k_n - 1)! \end{aligned} \tag{5.20}$$

which is an explicit expression for the number of connected graphs. Further, applying (5.13) to (5.20), we get

$$\sum_{n=1}^{\infty} \frac{C_n}{n!} x^n = \log\left(1 + a_1 x + \frac{a_2 x^2}{2!} + \frac{a_3 x^3}{3!} + \cdots\right). \tag{5.21}$$

From this, one can find the values of various probabilistic quantities related to connected graphs, such as the expected number of connected components, expected size of the largest component, asymptotic results, and so forth.

EXAMPLE 5.8. We now determine the number $a(n, k)$ of solutions of the equation $p^k = I$, where p is an element of the group G_n of all permutations of a set S_n of n elements, and I is the identity element of G_n. To every $p \in G_n$, we can associate the partition π of S_n whose blocks are the transitivity classes relative to the subgroup generated by p. Let $F(\pi)$ be the number of permutations p whose associated partition is π and such that $p^k = I$. Clearly, the function F is multiplicative, and so the function G, defined by $G(\sigma) = \sum_{\pi \leqq \sigma} F(\pi)$, is also multiplicative. Further,

$$G(\sigma) = a(1, k)^{k_1} a(2, k)^{k_2} \cdots \tag{5.22}$$

if $(k_1, k_2, \cdots)$ is the class of σ. Thus, if $(b_1, b_2, \cdots)$ is the sequence which determines F, then by (5.12), we obtain

$$1 + \sum_{n=1}^{\infty} \frac{a(n, k)}{n!} x^n = \exp\left\{\sum_{n=1}^{\infty} \frac{b_n}{n!} x^n\right\}. \tag{5.23}$$

Now, it is easily seen that $b_n = (n - 1)!$ if n divides k, and $b_n = 0$ otherwise, so we obtain the formula (due to Chowla, Herstein, and Scott [10])

$$\sum_{n=0}^{\infty} \frac{a(n, k)}{n!} x^n = \exp\left(\sum_{n|k} \frac{x^n}{n}\right), \tag{5.24}$$

where we take $a(0, k) = 1$.

EXAMPLE 5.9 (The number of partitions of a set). The number B_n of partitions of a set of n elements is given by $B_n = \sum_{\pi \in \Pi_n} \zeta(\pi)$. Hence, from Corollary (5.12), we get

$$\sum_{n=1}^{\infty} \frac{B_n}{n!} x^n = \exp\{e^x - 1\} - 1, \tag{5.25}$$

which is the classical generating function for B_n.

EXAMPLE 5.10. A set S of n elements splits at time t_1 into a partition π with blocks $B_1, B_2, \cdots$. At a later time $t_2 > t_1$ each block B_i splits into a partition π_i with blocks $B_{i,1}, B_{i,2}, \cdots$, and so on for N steps. Letting $E(x) = e^x - 1$, an argument much like that of the preceding example shows that the exponential generating function for the number of distinct "splittings" is $E[E(\cdots E(x)\cdots)]$, where the iteration is repeated $N + 1$ times.

5.3. *Dowling lattices.* Let F be the field of q elements (q will remain fixed throughout this subsection), and let V be a vector space over F of dimension n, with basis $b_1, \cdots, b_n$. The *Dowling lattice* $Q(V)$ is the lattice of subspaces W of V such that W has a basis whose elements are of the form b_i or $ab_j + a'b_k$, where $a, a' \in F$. Since the lattice $Q(V)$ depends up to isomorphism on the dimension of V, it will generally be denoted by Q_n.

Before attempting to study the combinatorial properties of Q_n, we will define a new lattice D_n, which is isomorphic to Q_n, in which various counting arguments become simpler. First we will state a number of definitions. The concept of *directed graph* is assumed (see Liu [38]), and we will allow loops and multiple edges between vertices. If S is any set, an S *labelled* directed graph is a directed graph $G = (V, E)$ together with a mapping from E to S in which no two edges from v to v' have the same image, for any $v, v' \in V$. The image of an edge e is called its *label*, and $v \xrightarrow{a} v'$ denotes the fact that there is an edge labelled a from v to v'. If G and G' are S labelled graphs, G is a *subgraph* of G', if both graphs have the same vertex set and if $v \xrightarrow{a} v'$ in G implies $v \xrightarrow{a} v'$ in G'. A *totally complete* S labelled directed graph G is one in which $v \xrightarrow{a} v'$ for any pair of vertices v and v' and any $a \in S$. If S consists of the nonzero elements of a field, then an S labelled directed graph G is *inverse symmetric* if $v \xrightarrow{a} v'$ implies $v' \xrightarrow{a^{-1}} v$, and is *antitransitive* if $v \xrightarrow{a} v'$ and $v' \xrightarrow{b} v''$ implies $v \xrightarrow{-ab} v''$. Finally, a *D graph* is an S labelled directed graph G, where S is the set of nonzero elements of a field in which there is at most one distinguished component which is totally complete, and every other component is simple (that is, at most one edge in each direction between two vertices), inverse symmetric, and antitransitive.

Now, let $S = F^*$ (the nonzero elements of F), and let B be a set of n elements ("vertices"). The lattice $D(B)$, or D_n, is the lattice of D graphs with vertex set B (and label set S), with $G \leqq G'$ if and only if G is a subgraph of G' and the distinguished component of G is contained in that of G'. The correspondence with the Dowling lattice Q_n is as follows. Given a Dowling lattice $Q(V)$ and a basis $B = \{b_1, \cdots, b_n\}$, to each subspace W of V in $Q(V)$ associate the graph whose vertex set is B and in which $b_i \xrightarrow{a} b_j$ if and only if $b_i + ab_j$ is in W, and in which the distinguished component is the one whose vertices are those b_i which are in W. The connected components are easily seen to be inverse symmetric and antitransitive, the distinguished component is clearly totally complete, and all other components are simple (for if $b_i \xrightarrow{a} b_j$ and $b_i \xrightarrow{a'} b_j$ with $a \neq a'$, then $b_i + ab_j \in W$, $b_i + a'b_j \in W$; hence $(a - a')b_j \in W$ and so $b_j \in W$ and $b_i \in W$, and thus b_i and b_j are in the distinguished component). This correspondence is easily seen to be a lattice isomorphism, and so D_n and Q_n are isomorphic.

EXAMPLE 5.11. It follows easily from what we have done that $Q_n \simeq \Pi_{n+1}$ if $q = 2$. The following correspondence gives an isomorphism from D_n to Π_{n+1}. Let the vertex set for D_n be $\{1, 2, \cdots, n\}$. To each element G of D_n, we associate the partition of $\{1, 2, \cdots, n, n+1\}$ whose blocks are the nondistinguished components of G as well as the distinguished component with $n + 1$ added.

Now, let $G \in D_n$. Then $[0, G]$ is isomorphic to the product of the lattices of subgraphs of the components of G (where 0 is the trivial graph with no edges and no distinguished component), and the lattice of subgraphs of a nondistinguished (and hence simple) component of G with k vertices is trivially isomorphic to Π_k. Hence, $[0, G]$ is isomorphic to $D_r \times \Pi_1^{k_1} \times \cdots \times \Pi_n^{k_n}$, where r is the size of the distinguished component of G (possibly 0) and k_i is the number of undistinguished components of G of size i. (Note that $r + \Sigma ik_i = n$ and Σk_i equals the number of undistinguished blocks in G.) Let G' be above G in D_n, and let C_1 and C_2 be distinct undistinguished components of G which are in the same undistinguished component of G'. Then all edges between vertices of C_1 and vertices of C_2 can be determined from any one such edge, using the properties of inverse symmetry and antitransitivity. Intuitively, the undistinguished components of G "act like points" in $[G, 1]$, while the distinguished component of G simply "joins with these points as they become distinguished." Using these ideas, it is not difficult to see (or to prove) that $[G, 1]$ is isomorphic to Q_m, where m is the number of undistinguished components of G, that is, to $D_{k_1+\cdots+k_n}$ (the k_i are introduced earlier in this paragraph).

We are thus led to the following definition corresponding to that in the previous subsection. The *class* of a segment $[G, G']$ of D_n is the sequence $(r; k_1, k_2, \cdots)$, where r is the number of undistinguished components of G which are contained in the distinguished component of G', and k_i is the number of undistinguished components of G' which contain exactly i components of G. (Note that $r + \Sigma ik_i$ equals the number of undistinguished components of G.) It follows from the previous paragraph that $[G, G']$ is isomorphic to $D_r \times \Pi_1^{k_1} \times \Pi_2^{k_2} \times \cdots$. The *class* of an element $G \in D_n$ is defined to be the class of $[0, G]$.

Before going any further, we will put everything preceding in the context of a large incidence algebra in which two segments are of the same type if and only if they have the same class. Let **D** be the category whose objects are the lattices $D(B)$ for all finite sets B, with two segments being equivalent if they are isomorphic and their top elements have the same number of undistinguished components (although one top element may have a distinguished component and the other not). The morphisms of **D** are all isomorphisms into in which the top element of the segment has the same number of undistinguished components as does the top element of the image segment. It is easy to see that **D** satisfies the required conditions, and also that two segments are equivalent if and only if they are of the same class.

Now, a segment $[G, G']$ in some D_n is of type $(r; 0, 0, \cdots)$ if and only if $G' = 1$ and G has r undistinguished components. We denote by $\left[{n \atop r; k_1, k_2, \cdots}\right]$ the number of elements G' in a segment $[G, 1]$ of type $(n; 0, 0, \cdots)$ such that $[G, G']$ has type $(r; k_1, k_2, \cdots)$ (and hence, $[G', 1]$ has type $[k_1 + k_2 + \cdots; 0, 0, \cdots]$). Then

$$(5.26) \qquad \left[{n \atop r; k_1, k_2, \cdots}\right] = \binom{n}{r}\binom{n-r}{k_1, k_2, \cdots}(q-1)^{k_2+2k_3+3k_4+\cdots},$$

where $\binom{n-r}{k_1, k_2, \cdots}$ is defined as in the previous subsection, as the following counting argument shows. We may assume that $[G, 1]$ is contained in D_n and that $G = 0$, that is, that $[G, 1]$ is D_n. First, choose r vertices and join edges between all pairs with all labels and distinguish the resulting component. This can be done in $\binom{n}{r}$ ways. Then choose k_1 vertices as the undistinguished one point components. This can be done in $\binom{n-r}{k_1}$ ways. Proceeding in this way, choose k_j distinct j sets of vertices. This can be done in

$$\frac{1}{k_j! j!^{k_j}} (n - r - k_1 - 2k_2 - \cdots - (j-1)k_{j-1}) \cdot (n - r - k_1 - 2k_2 - \cdots - (j-1)k_{j-1}) \cdots (n - r - k_1 - 2k_2 - \cdots - jk_j + 1) \tag{5.27}$$

ways, and each j set can be made into a labelled, simple, inverse symmetric, antitransitive component in $(q-1)^{j-1}$ ways, since the labelling is completely determined by the labels on a spanning tree, which has $j - 1$ edges (see Liu [38], pp. 185–186). This establishes (5.26).

As for lattices of partitions, the concept of *multiplicative* function is important. A function $f \in I(D)$ is multiplicative, if there is a pair of sequences $(a_1, a_2, a_3, \cdots)$, $(b_0, b_1, b_2, \cdots)$ such that

$$f(G, G') = b_r \cdot a_1^{k_1} a_2^{k_2} \cdots \tag{5.28}$$

when $[G, G']$ is of type $(r; k_1, k_2, \cdots)$, and f is said to be determined by the pair of sequences. A similar definition holds for multiplicative functions of one variable. The subset $\mathbf{M}(\mathbf{D})$ of $\mathbf{I}(\mathbf{D})$ of all multiplicative functions is closed under convolution (the proof is the same as for $\mathbf{M}(\mathbf{\Pi})$), and hence, $\mathbf{M}(\mathbf{D})$ is a semigroup. Also, if $f \in \mathbf{M}(\mathbf{D})$ and F is a multiplicative function of one variable, then K and L are also multiplicative, where

$$K(G) = \sum_{G' \leqq G} F(G') f(G', G) \tag{5.29}$$

and

$$L(G) = \sum_{G' \geqq G} f(G, G') F(G'). \tag{5.30}$$

THEOREM 5.2. *The semigroup* $\mathbf{M}(\mathbf{C})$ *is isomorphic to the set of all pairs* $(F(x), G(x))$ *of formal exponential power series in which* $F(x)$ *has zero constant term, with multiplication given by*

$$(F(x), G(x)) \cdot (F'(x), G'(x)) = \left(F'(F(x)), G(x) \cdot G'\left(\frac{F((q-1)x)}{q-1}\right)\right). \tag{5.31}$$

The isomorphism is given by $f \to (F_f^{(1)}(x), F_f^{(2)}(x))$, *where*

$$F_f^{(1)}(x) = \sum_{n=1}^{\infty} \frac{f(\Pi_n)}{n!} x^n, \tag{5.32}$$

(5.33) $$F_f^{(2)}(x) = \sum_{n=0}^{\infty} \frac{f(D_n)}{n!} x^n,$$

and where $f(\Pi_n)$ *denotes the value of* f *on a segment of type* $r = 0, k_1 = \cdots = k_{n-1} = 0, k_n = 1$, *and* $f(D_n)$ *denotes the value of* f *on a segment of type* $(n; 0, 0, \cdots)$.

PROOF. Clearly, the map defined is a bijection, so we need only check that multiplication is preserved. Let $f, g \in \mathbf{M}(\mathbf{D})$. It follows from Theorem 5.1 that $F_{f*g}^{(1)}(x) = F_g^{(1)}(F_f^{(1)}(x))$. Now, from (5.26) and denoting by Σ^* a summation taken over the set $\{r + k_1 + 2k_2 + \cdots + nk_n = n\}$ and by Σ_r^{**} a summation taken over the set $\{k_1 + 2k_2 + \cdots + nk_n = n - r\}$, we get

(5.34)
$$\begin{aligned}(f*g)(D_n) &= \sum{}^* \begin{bmatrix} n \\ r; k_1, k_2, \cdots, k_n \end{bmatrix} f(D_r) f(\Pi_1)^{k_1} \cdots f(\Pi_n)^{k_n} g(D_{k_1+\cdots+k_n}) \\ &= \sum{}^* \binom{n}{r} \binom{n-r}{k_1, k_2, \cdots, k_n} (q-1)^{k_2+2k_3+\cdots+(n-1)k_n} f(D_r) f(\Pi_1)^{k_1} \\ &\qquad \cdots f(\Pi_n)^{k_n} g(D_{k_1+\cdots+k_n}) \\ &= \sum_{r=0}^{n} \binom{n}{r} f(D_r) \Big(\sum{}_r^{**} \binom{n-r}{k_1, k_2, \cdots, k_n} f(\Pi_1)^{k_1} (f(\Pi_2)(q-1))^{k_2} \\ &\qquad \cdots (f(\Pi_n)(q-1)^{n-1})^{k_n} g(D_{k_1+\cdots+k_n}) \Big).\end{aligned}$$

Now, $f(D_r)$ is the coefficient of $x^r/r!$ in $F_f^{(2)}(x)$, and

(5.35)
$$\begin{aligned}\sum{}_r^{**} \binom{n-r}{k_1, k_2, \cdots, k_n} f(\Pi_1)^{k_1} (f(\Pi_2)(q-1))^{k_2} \\ \cdots (f(\Pi_n)(q-1)^{n-1})^{k_n} g(D_{k_1+\cdots+k_n})\end{aligned}$$

is the coefficient of $x^{n-r}/(n-r)!$ in $F_g^{(2)}(F_f^{(1)}((q-1)x)/(q-1))$, and hence the theorem follows.

COROLLARY 5.3. *The Möbius function in* D_n *is given by*

(5.36) $$\mu(0, 1) = (-1)^n \prod_{i=0}^{n-1} [1 + (q-1)i].$$

PROOF. We have $F_\zeta^{(1)}(x) = e^x - 1, F_\zeta^{(2)}(x) = e^x, F_\delta^{(1)}(x) = x$, and $F_\delta^{(2)}(x) = 1$. Now, $F_\mu^{(1)}(x) = \Sigma_{n=1}^{\infty} (\mu(\Pi_n)/n!) x^n = \log(1 + x)$ from the previous subsection. Thus,

(5.37)
$$\begin{aligned}1 = F_\delta^{(2)}(x) = F_{\mu*\zeta}^{(2)}(x) &= F_\mu^{(2)}(x) \cdot F_\zeta^{(2)}\left(\frac{F_\mu^{(1)}((q-1)x)}{q-1}\right) \\ &= F_\mu^{(2)}(x) \cdot \exp\left\{\frac{\log(1 + (q-1)x)}{q-1}\right\} \\ &= F_\mu^{(2)}(x) \cdot (1 + (q-1)x)^{1/(q-1)}.\end{aligned}$$

Hence,

$$F_\mu^{(2)}(x) = \big(1 + (q - 1)x\big)^{-1/(q-1)}, \tag{5.38}$$

from which the result follows.

COROLLARY 5.4. *Let f be a multiplicative function of one variable. For every nonnegative integer n, let* $b_n = \Sigma_{G \in D_n} f(G)$ *and* $q_n = \Sigma_{G \in D_n} f(G)\mu(G, 1)$. *Then*

$$\sum_{n=0}^{\infty} \frac{b_n}{n!} x^n = F_f^{(2)}(x) \cdot \exp\left\{\frac{F_f^{(1)}((q-1)x)}{q-1}\right\}, \tag{5.39}$$

$$\sum_{n=0}^{\infty} \frac{q_n}{n!} x^n = F_f^{(2)}(x) \cdot [1 + F_f^{(1)}((q-1)x)]^{-1/(q-1)}. \tag{5.40}$$

PROOF. The proof follows from Theorem 5.2 and Corollary 5.3 in the same way as Corollary 5.2 is proved.

Let us now return to the lattice Q_n to get an idea of what the class of a segment means in terms of the corresponding segment of vector spaces. Let $W \in Q(V)$, that is, W is a subspace of V which has a basis whose elements are of the form b_i or $ab_j + a'b_k$ (where $a, a' \in F^*$). Then it is not difficult to see that W has a basis of the form

$$\{b_{i_1}, b_{i_2}, \cdots, b_{i_r}, b_{j_1} + a_1 b_{j_2}, b_{j_2} + a_2 b_{j_3}, \cdots, b_{j_s} + a_s b_{j_{s+1}}, b_{k_1} + a'_1 b_{k^2}, b_{k_2} + a'_2 b_{k_3}, \cdots, b_{k_t} + a'_t b_{k_{t+1}}, \cdots, b_{\ell_1} + a''_1 b_{\ell_2}, \cdots, b_{\ell_u} + a''_u b_{\ell_{u+1}}\}, \tag{5.41}$$

where the a_i are nonzero and no b_i appears twice. Such a basis can be obtained by taking any basis and noting that if $ab_i + a'b_j$ and $\bar{a}b_i + \bar{a}'b_j$ both appear (and hence $a/a' \neq \bar{a}/\bar{a}'$), then b_i and b_j are in W and can replace $ab_i + a'b_j$ and $\bar{a}b_i + \bar{a}'b_j$ in the basis. The collection $\{b_{j_1} + a_1 b_{j_2}, b_{j_2} + a_2 b_{j_3}, \cdots, b_{j_s} + a_s b_{j_{s+1}}\}$ in the above basis is called an $(s + 1)$ cycle of the basis. Let k_1 equal the number of basis elements $\{b_1, b_2, \cdots, b_n\}$ which do not appear in the above basis (that is, are not among $\{b_{i_1}, b_{i_2}, \cdots, b_{i_r}, b_{j_1}, \cdots, b_{j_{s+1}}, \cdots, b_{\ell_1}, \cdots, b_{\ell_{u+1}}\}$), and for $i > 1$, let k_i be the number of i cycles in the basis. Then $(r; k_1, k_2, \cdots, k_n)$ is the class of the segment $[0, G]$ in D_n (where G is the graph corresponding to W), and $[G, 1]$ has class $[k_1 + k_2 + \cdots + k_n; 0, 0, \cdots]$. Note that it follows from this that r and $k_1, k_2, \cdots, k_n$ do not depend on the basis (of the proper type) chosen for W. The class of a general segment $[W, W']$ could also be determined from bases of the proper form chosen for W and W'. Thus, the class of a segment of Q_n could have been defined without resorting to the lattice D_n, but it then becomes necessary to prove that the class does not depend on the bases chosen.

5.4. *Abelian groups.* Let $\mathbf{C}(p)$ be the category whose objects are lattices of subgroups of finite abelian p groups (where p is a prime, fixed throughout) and all segments thereof, with the equivalence relation in each object being given by $[A, B] \simeq [G, H]$ if and only if $B/A \simeq H/G$. Morphisms in $\mathbf{C}(p)$ are all isomorphisms into such that if $[A, B]$ is the domain and $[G, H]$ the image of the isomorphism, then $B/A \simeq H/G$.

A *partition of an integer* n is a sequence $\lambda = (\lambda_1, \lambda_2, \cdots)$ of nonnegative integers arranged in decreasing order, whose sum is n. The types in the category $\mathbf{C}(p)$ above are in one to one correspondence with partitions, the type of a segment $[A, B]$ being $\lambda = (\lambda_1, \lambda_2, \cdots)$, where $B/A \simeq Z(p^{\lambda_1}) \oplus Z(p^{\lambda_2}) \oplus \cdots$. The type of a group G is defined to be the type of $[0, G]$. The incidence coefficient $\binom{\lambda}{\alpha, \beta}$ is equal to the number of members $G \in [A, B]$ (where $[A, B]$ is of type λ) such that $[A, G]$ has type α and $[G, B]$ has type β, or equivalently the number of subgroups G of a group H (where H is of type λ) such that G has type α and H/G has type β. This is precisely the "Hall polynomial" $g^{\lambda}_{\alpha,\beta}(p)$ defined in Hall [31], p. 156, and further studied by Green [27] and Klein [37]. (The Hall polynomials $g^{\rho}_{\lambda\mu\cdots\nu}(p)$ are simply the coefficients in the expression $(f*g*\cdots*h)(\rho) = \Sigma g^{\rho}_{\lambda\mu\cdots\nu}(p)f(\lambda)\cdots h(\nu)$.) Hall's algebra $A(p)$ is isomorphic to the subalgebra of $\mathbf{I}(\mathbf{C}(p))$ consisting of functions which are nonzero on only finitely many types, the isomorphism being given by linearly extending the map $\delta_\lambda \to G_\lambda(p)$, where δ_λ is the indicator function of the type λ in $\mathbf{I}(\mathbf{C}(p))$ and where $G_\lambda(p)$ is as in Hall's paper. The incidence coefficients $g^{\lambda}_{\alpha,\beta}(p)$ satisfy $g^{\lambda}_{\alpha,\beta}(p) = g^{\lambda}_{\beta,\alpha}(p)$, which follows from the well-known fact that the lattice of subgroups of a finite abelian group is self dual, and hence by Corollary 4.1, $\mathbf{I}(\mathbf{C}(p))$ is commutative. Various properties of the incidence coefficients $g^{\lambda}_{\alpha,\beta}(p)$ are worked out by Hall and extended by Klein and Green, the most basic being that $g^{\lambda}_{\alpha,\beta}(p)$ is a polynomial in p with integer coefficients. A condition for this polynomial to be identically zero, that is, for $g^{\lambda}_{\alpha,\beta}(p)$ to equal zero for all p, is given by Hall in terms of multiplication of Schur functions (see [31], p. 157).

EXAMPLE 5.12. Let $(r_1, r_2, \cdots, r_m)$ be an ordered partition of n. Then it follows from the commutativity of $\mathbf{I}(\mathbf{C}(p))$ that given any partition λ of n, the number of towers $1 \leqq H_1 \leqq H_2 \leqq \cdots \leqq H_m = G$ (where G has type λ) in which H_i/H_{i-1} has order p^{r_i} is independent of the arrangement of $(r_1, r_2, \cdots, r_m)$. This is because the number of such chains is given by $(h_{r_1} * h_{r_2} * \cdots * h_{r_m})(\lambda)$, where h_r is the function which takes the value 1 on segments $[A, B]$ in which B/A has order p^r, and is zero elsewhere (h_r is clearly constant on each type).

6. Residual isomorphism

In this section we are mainly concerned with the problem of determining when two segments are equivalent in the maximally reduced incidence algebra $\bar{\mathbf{R}}(P)$. As has been seen in Section 4, the two segments need not be isomorphic, that is, the standard reduced incidence algebra need not equal the maximally reduced incidence algebra. Until further notice, we will assume *the ground field K has characteristic* 0.

First, we give a criterion when two segments of P are equivalent in $\bar{\mathbf{R}}(P)$. Let $P = [0, 1]$ and $P' = [0', 1']$ be two finite ordered sets with unique minimal elements 0 and $0'$, and unique maximal elements 1 and $1'$, respectively. We say that P and P' are 1-*equivalent*, without imposing any other conditions on them. Define inductively P and P' to be $(n + 1)$-*equivalent* (written $P \overset{n+1}{\sim} P'$) if there exists a bijection $x \leftrightarrow x'$ between P and P' such that $[0, x] \overset{n}{\sim} [0', x']$ and $[x, 1] \overset{n}{\sim} [x', 1']$. Note that $P \overset{n}{\sim} P'$ implies $P \overset{m}{\sim} P'$ for $1 \leqq m \leqq n$. Note also that $P \overset{2}{\sim} P'$ if and only if P and P' have the same number of elements.

PROPOSITION 6.1. *Two segments $[x, y]$ and $[u, v]$ of a locally finite ordered set P are equivalent in $\bar{\mathbf{R}}(P)$ if and only if they are n equivalent for all positive integers n.*

Before proving Proposition 6.1, we show that the apparently infinite sequence of conditions that must be satisfied in order that $P \overset{n}{\sim} P'$ for every positive integer n reduces to a finite number of conditions for any given choice of P, P'.

PROPOSITION 6.2. *Let ℓ be the length of the longest chain of the two finite ordered sets $P = [0, 1]$ and $P' = [0', 1']$. Then $P \overset{n}{\sim} P'$ for every $n \geqq 1$ if and only if $P \overset{\ell}{\sim} P'$.*

PROOF. The proof is by induction on ℓ. The conclusion clearly holds when $\ell = 1$ and $\ell = 2$, since then P and P' are isomorphic. Now assume that the conclusion holds for $\ell \geqq 2$ and that the longest chain of P and P' has length $\ell + 1$ and that $P \overset{\ell+1}{\sim} P'$. Assume that $P \overset{n}{\sim} P'$ for some $n \geqq \ell + 1$. We will be done if we show $P \overset{n+1}{\sim} P'$. Since $P \overset{n}{\sim} P'$, there exists a bijection $x \leftrightarrow x'$ with $[0, x] \overset{n-1}{\sim} [0', x']$ and $[x, 1] \overset{n-1}{\sim} [x', 1']$. Clearly, $0 \leftrightarrow 0'$ and $1 \leftrightarrow 1'$, since $n \geqq 3$. If $x \neq 0, 1$, then $[0, x]$ and $[0', x']$ have no chain of length $\geqq \ell + 1$, so by the induction hypothesis $[0, x] \overset{n}{\sim} [0', x']$. Similarly, $[x, 1] \overset{n}{\sim} [x', 1']$. Hence, the bijection $x \leftrightarrow x'$ defines an $n + 1$ equivalence between P and P', and the proof is complete.

We conjecture that the following converse to Proposition 6.2 holds: for every $\ell \geqq 1$, there exist finite ordered sets $P = [0, 1]$ and $P' = [0', 1']$ with longest chain of length ℓ such that $P \overset{\ell-1}{\sim} P'$ but *not* $P \overset{\ell}{\sim} P'$. Figure 1 illustrates the validity of this conjecture for $\ell = 4$.

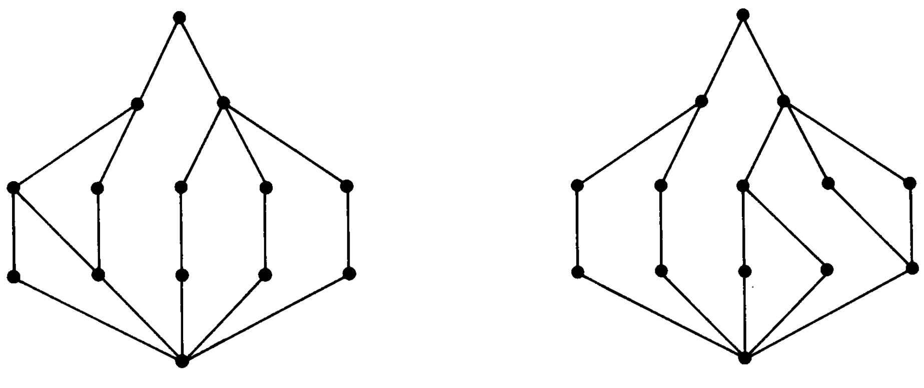


FIGURE 1
Ordered sets of length 4 which are 3-equivalent, but not 4-equivalent.

PROOF OF PROPOSITION 6.1. Define $[x, y] \sim [x', y']$ in P if and only if $[x, y] \overset{n}{\sim} [x', y']$ for all positive integers n. To prove the "if" part, we need to show that the equivalence relation $\sim$ is order compatible. It suffices to show that the coefficient $\left[{\alpha \atop \beta, \gamma}\right]$ is well defined for any equivalence classes (types) α, β, γ. Let $[x, y]$ and $[x', y']$ be two segments of P of type α. Define $r(x, y, n)$ to be the number of points $z \in [x, y]$ such that $[x, z]$ is n equivalent to a segment of type β and $[z, y]$ is n equivalent to a segment of type γ. The number $r(x, y, n)$ is independent of the particular choice of segments of type β and γ, since all such segments are n equivalent. Since $[x, y] \overset{n+1}{\sim} [x', y']$, we have $r(x, y, n) = r(x', y', n)$. But then

$$\left[\begin{matrix}\alpha \\ \beta, \gamma\end{matrix}\right] = \lim_{n \to \infty} f(x, y, n) = \lim_{n \to \infty} r(x', y', n), \tag{6.1}$$

so $\left[{\alpha \atop \beta, \gamma}\right]$ is well defined.

Conversely, suppose $[x, y] \sim [x', y']$ in $\bar{\mathbf{R}}(P)$. We prove by induction on n that $[x, y] \overset{n}{\sim} [x', y']$ for all n. Trivially $[x, y] \overset{1}{\sim} [x', y']$ for all $[x, y] \sim [x', y']$ (indeed, for *any* pair $[x, y]$, $[x', y']$). Assume $[x, y] \overset{n}{\sim} [x', y']$ for all $[x, y] \sim [x', y']$. Given any segment $[u, v]$ of P, define $f_{u,v,n} \in \mathbf{I}(P)$ by

$$f_{u,v,n}(x, y) = \begin{cases} 1 & \text{if } [x, y] \overset{n}{\sim} [u, v], \\ 0 & \text{otherwise.} \end{cases} \tag{6.2}$$

By the induction hypothesis $f_{u,v,n} \in \bar{\mathbf{R}}(P)$. Hence, $f_{u,v,n} * f_{u',v',n} \in \bar{\mathbf{R}}(P)$. But $f_{u,v,n} * f_{u',v',n}(x, y)$ is just the number of elements $z \in [x, y]$ such that $[x, z] \overset{n}{\sim} [u, v]$ and $[z, y] \overset{n}{\sim} [u', v']$. (This is where the assumption that K has characteristic 0 is needed.) Since $f_{u,v,n} * f_{u',v',n} \in \bar{\mathbf{R}}(P)$,

$$f_{u,v,n} * f_{u',v',n}(x, y) = f_{u,v,n} * f_{u',v',n}(x', y'). \tag{6.3}$$

Hence, $[x, y] \overset{n+1}{\sim} [x', y']$, and the proof is complete.

The proof of Proposition 6.1 allows us to characterize the form of functions in $\bar{\mathbf{R}}(P)$, at least when the characteristic of the ground field K is 0. If $f \in \mathbf{I}(P)$, define $\chi_f \in \mathbf{I}(P)$ by

$$\chi_f(x, y) = \begin{cases} 1 & \text{if } f(x, y) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

COROLLARY 6.1. The algebra $\bar{R}(P)$ *consists of those functions which can be obtained from* ζ *by a sequence of operations of the following three types:*

(i) *linear combination (possible infinite),*

(ii) *convolution,*

(iii) *the operation* $f \to \chi_f$.

PROOF. Clearly, all functions of the type described are in $\bar{\mathbf{R}}(P)$. The proof of Proposition 6.1 shows that for any segment $[u, v]$ of P, the function $f_{u,v,n} \in \bar{\mathbf{R}}(P)$. Proposition 6.2 shows that the sequence $f_{u,v,1}, f_{u,v,2}, \cdots$ is eventually constant, and that its limit (namely, $f_{u,v,n}$, where n is the greater of 2 and the number of elements in $[u, v]$, as is easily verified) is the characteristic function for the type of $[u, v]$ in $\bar{\mathbf{R}}(P)$. All functions in $\bar{\mathbf{R}}(P)$ are linear combinations (infinite if $\bar{\mathbf{R}}(P)$ has infinitely many types) of such characteristic functions.

Finally, it is not difficult to show (by induction on n) that $f_{u,v,n}$ is in the class of functions described for every segment $[u, v]$ and every n, so the proof is complete.

Define two locally finite ordered sets P and Q to be *residually isomorphic* (r isomorphic for short) if there is a bijection between the types of P relative to $\bar{\mathbf{R}}(P)$ and the types of Q relative to $\bar{\mathbf{R}}(Q)$ (over the same ground field K, which we still assume to have characteristic 0) inducing an isomorphism of $\bar{\mathbf{R}}(P)$ and $\bar{\mathbf{R}}(Q)$.

NOTE. It is possible for $\bar{\mathbf{R}}(P)$ and $\bar{\mathbf{R}}(Q)$ to be isomorphic as K algebras, and yet P and Q are not r isomorphic.

PROPOSITION 6.3. *Two finite ordered sets P and P', each with 0 and 1, are r isomorphic if and only if $P \overset{n}{\sim} P'$ for all $n \geqq 1$. Equivalently, two segments of a locally finite ordered set P are equivalent in $\bar{\mathbf{R}}(P)$ if and only if those segments are r isomorphic. Furthermore, α and α' are corresponding types in the isomorphism $\bar{\mathbf{R}}(P) \cong \bar{\mathbf{R}}(P')$ if and only if the segments in P of type α are r isomorphic to the segments in P' of type α'.*

PROOF. Assume P and P' are r isomorphic, and that a type α relative to $\bar{\mathbf{R}}(P)$ corresponds to a type α' relative to $\bar{\mathbf{R}}(P')$. Let Q be the disjoint union (direct sum) $P + P'$. Define an equivalence relation on segments of Q by $[x, y] \sim [x', y']$ if either (1) $[x, y] \sim [x', y']$ in $\bar{\mathbf{R}}(P)$, (2) $[x, y] \simeq [x', y']$ in $\bar{\mathbf{R}}(P')$, (3) $[x, y]$ is of type α in P and $[x', y']$ of type α' in P', or (4) $[x, y]$ is of type α' in P' and $[x', y']$ of type α in P. Clearly, this equivalence relation is order compatible. Hence, by Proposition 6.1 segments of type α are n equivalent to segments of type α' for all $n \geqq 1$, and in particular $P \overset{n}{\sim} P'$.

Conversely, if $P \overset{n}{\sim} P'$ for all $n \geqq 1$, define a bijection $\alpha \leftrightarrow \alpha'$ between types α relative to $\bar{\mathbf{R}}(P)$ and types α' relative to $\bar{\mathbf{R}}(P')$ by requiring that segments of type α be n equivalent to segments of type α' for all $n \geqq 1$. It follows easily from Proposition 6.1 that this bijection induces an isomorphism between $\bar{\mathbf{R}}(P)$ and $\bar{\mathbf{R}}(P')$, and the proof is complete.

COROLLARY 6.2. *Two finite r isomorphic ordered sets $P = [0, 1]$ and $P' = [0', 1']$ have the following properties in common:*

(i) *number of maximal chains of a given length,*

(ii) *number of elements a given minimum length from the bottom (or top); consequently, total number of elements, number of atoms, and number of dual atoms.*

PROOF. It follows from Corollary 6.1 that the function $\eta = \chi_{\zeta^2 - \zeta}$ is in $\bar{\mathbf{R}}(P)$ and $\bar{\mathbf{R}}(P')$. Note that

$$\eta(x, y) = \begin{cases} 1 & \text{if } \; y \text{ covers } x, \\ 0 & \text{otherwise}, \end{cases} \tag{6.5}$$

so that $\eta^r(x, y)$ is the number of maximal chains of $[x, y]$ of length r. By Proposition 6.3, $\eta^r(0, 1) = \eta^r(0', 1')$, since P and P' are r isomorphic. This proves (i).

Similarly one can find functions in $\bar{\mathbf{R}}(P)$ and $\bar{\mathbf{R}}(P')$, explicitly expressed in the form given by Corollary 6.1, which enumerate the quantities in (ii). The details we omit.

PROPOSITION 6.4. *Let P be an ordered set with 0 and 1 with $\leqq 7$ elements, and let Q be any finite ordered set with 0 and 1. Then P and Q are r isomorphic if and only if they are isomorphic.*

The proof is essentially by inspection of all possibilities, and will be omitted. Figure 2 shows two r isomorphic nonisomorphic ordered sets with eight elements. Another example of r isomorphic nonisomorphic ordered sets is the lattice of subspaces of two nonisomorphic finite projective planes of the same order.

FIGURE 2
Residually isomorphic nonisomorphic ordered sets.

We say that a finite ordered set P with 0 and 1 is *residually self dual* (r self dual for short) if it is r isomorphic to its dual. The next proposition uses this concept to characterize those P for which $\bar{\mathbf{R}}(P)$ is commutative.

PROPOSITION 6.5. *Let P be a locally finite ordered set. Then $\bar{\mathbf{R}}(P)$ is commutative if and only if every segment of P is r self dual.*

PROOF. Suppose $\bar{\mathbf{R}}(P)$ is commutative. This means that $\left[{\alpha \atop \beta, \gamma}\right] = \left[{\alpha \atop \gamma, \beta}\right]$ for all types α, β, γ. If δ is the type of a segment, let δ^* be the type of its dual. If $[x, y]$ is a segment of type α, consider the bijection $\delta \leftrightarrow \delta^*$ between types of segments in $[x, y]$ and types in the dual $[x, y]^*$. Then

$$\left[\begin{matrix}\alpha^* \\ \beta^*, \gamma^*\end{matrix}\right] = \left[\begin{matrix}\alpha \\ \gamma, \beta\end{matrix}\right] = \left[\begin{matrix}\alpha \\ \beta, \gamma\end{matrix}\right], \tag{6.6}$$

so the bijection $\delta \leftrightarrow \delta^*$ induces an isomorphism between $\bar{\mathbf{R}}([x, y])$ and $\bar{\mathbf{R}}([x, y]^*)$, that is, $[x, y]$ is r self dual.

Conversely, suppose every segment $[x, y]$ of P is r self dual. Since $[x, y]$ is r self dual, the number of elements $z \in [x, y]$ such that $[x, z]$ is of type β and $[z, y]$ of type γ is equal to the number of elements $z' \in [x, y]$ such that $[x, z']$ is of type γ^* and $[z', y]$ of type β^*. But $\beta = \beta^*$ and $\gamma = \gamma^*$, since every segment of these types is r self dual. Hence, if $[x, y]$ is of type α, then $\left[{\alpha \atop \beta, \gamma}\right] = \left[{\alpha \atop \gamma, \beta}\right]$ and $\bar{\mathbf{R}}(P)$ is commutative. This completes the proof.

Figure 3 illustrates an r self dual ordered set P which is not self dual. For this ordered set, $\bar{\mathbf{R}}(P)$ is equal to the standard reduced incidence algebra. This answers a question of Smith ([55], p. 632) on the existence of such ordered sets.

REMARKS. *On characteristic p.* Proposition 6.1 and its consequences are false if the characteristic of the ground field is not 0. For example, whenever the

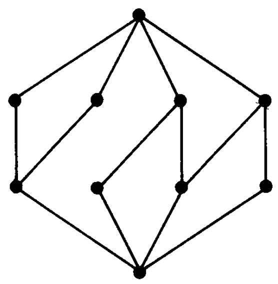


FIGURE 3
A residually self dual ordered set which is not self dual.

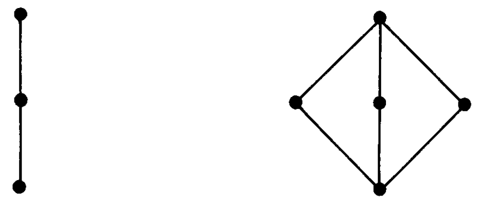


FIGURE 4
Equivalent segments in characteristic 2 which are not residually isomorphic.

two ordered sets of Figure 4 occur as segments of a locally finite ordered set P, then they are equivalent in $\bar{\mathbf{R}}(P)$ over a ground field of characteristic 2. It is not difficult, however, to modify the results of this section to get corresponding results for characteristic p, basically by replacing all concepts by the corresponding concepts modulo p. We will not go into the details here.

7. Algebras of Dirichlet type

7.1. *Definitions.* Let P be a locally finite ordered set, having a unique minimal element 0. Let $\mathbf{R}(P, \sim)$ be a reduced incidence algebra whose types are in one to one correspondence with a subset of the positive integers, the type of a segment $[x, y]$ being denoted by $O(x, y)$. Suppose the function O satisfies the following property: if $x \leqq y \leqq z$ in P, then $O(x, z) = O(x, y)O(y, z)$.

We then call $\mathbf{R}(P, \sim)$ an *algebra of Dirichlet type*. The bracket $\left[{n \atop k,\ell}\right]$ stands for the number of points y in a segment $[x, z]$ of type n such that $O(x, y) = k$ and $O(y, z) = \ell$. Clearly, $\left[{n \atop k,\ell}\right] = 0$ unless $n = k\ell$. Hence, it makes sense to define the *brace* $\left\{{n \atop k}\right\} = \left[{n \atop k, n/k}\right]$. The reduced incidence algebra $\mathbf{R}(P, \sim)$ is isomorphic to the algebra of all sequences a_n, $n = 1, 2, \cdots$, where $a_n \neq 0$ only if there is a segment of type n in P. The convolution of two such sequences is

$$c_n = \sum_{k|n} \left\{{n \atop k}\right\} a_k b_{n/k}. \tag{7.1}$$

EXAMPLE 7.1. Let P be the set of all positive integers, ordered by divisibility. Set $O(k, n) = n/k$, for $k, n \in P$. This gives the reduced incidence algebra mentioned at the beginning of Example 4.8. The braces are identically equal to one, the convolution is commutative, and it reduces to the classical *Dirichlet convolution*

$$c_n = \sum_{k|n} a_k b_{n/k}. \tag{7.2}$$

The reduced incidence algebra $\mathbf{R}(P, \sim)$ is isomorphic to the algebra of formal Dirichlet series. The zeta function is mapped into the *Riemann zeta function*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{7.3}$$

and the Möbius function goes into the function

$$\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, \tag{7.4}$$

where $\mu(n)$ is the classical Möbius function, as has already been sketched in *Foundations I*.

Algebras of Dirichlet type satisfy the following fundamental recursion:

$$\begin{Bmatrix} n \\ m \end{Bmatrix} \begin{Bmatrix} m \\ k \end{Bmatrix} = \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{Bmatrix} n/k \\ m/k \end{Bmatrix}. \tag{7.5}$$

This is obtained by counting in two ways the number of subsegments $[x_1, y_1]$ of a segment $[x, y]$ of type n such that $O(x, x_1) = k$, $O(x, y_1) = m$. There are $\{{n \atop m}\}$ ways of choosing y_1, and for each such choice there are $\{{m \atop k}\}$ ways of choosing x_1 below it. On the other hand, there are $\{{n \atop k}\}$ ways of choosing x_1, and for each such choice there are $\{{n/k \atop m/k}\}$ ways of choosing y_1 above it. This establishes (7.5).

There are three kinds of algebras $\mathbf{R}(P, \sim)$ of Dirichlet type of special importance.

(A) The algebra $\mathbf{R}(P, \sim)$ is *commutative* if and only if $\{{n \atop k}\} = \{{n \atop n/k}\}$ for all types n and all $k|n$.

(B) The algebra $\mathbf{R}(P, \sim)$ is said to be *full*, if whenever n is a type and $k|n$, then $\{{n \atop k}\} \neq 0$.

(C) The algebra $\mathbf{R}(P, \sim)$ is said to be of *binomial type*, if there is a prime p such that all types are powers of p. We then write

$$\begin{Bmatrix} p^b \\ p^a \end{Bmatrix} = \begin{bmatrix} b \\ a \end{bmatrix}. \tag{7.6}$$

The recursion (7.5) becomes

$$\begin{bmatrix} c \\ b \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} c \\ a \end{bmatrix} \begin{bmatrix} c - a \\ b - a \end{bmatrix}. \tag{7.7}$$

An algebra of binomial type is simply the additive analogue of an algebra of Dirichlet type. We shall always speak of algebras of binomial type in an additive sense, so a segment of type n in an algebra $\mathbf{R}(P, \sim)$ of binomial type is of type p^n when $\mathbf{R}(P, \sim)$ is regarded as an algebra of Dirichlet type.

There are, *a priori*, eight kinds of algebras of Dirichlet type obtained by specifying which of (A), (B), (C) hold or do not hold. It is easy to construct examples of seven of these kinds; in the next section, we shall see that every algebra of full binomial type is commutative.

7.2. *Full commutative algebras of Dirichlet type.* In this section, we show that if $\mathbf{R}(P, \sim)$ is a full commutative algebra of Dirichlet type, then there is an isomorphism of $\mathbf{R}(P, \sim)$ into formal Dirichlet series.

LEMMA 7.1. *Let $\mathbf{R}(P, \sim)$ be a full commutative algebra of Dirichlet type. Then the segments of P of type* 1 *are precisely the one point segments, and a segment has a prime type if and only if it is a two point segment. Further, P satisfies the Jordan–Dedekind chain condition, that is, in all segments of P, all maximal chains have the same length.*

PROOF. If $[x, x]$ has type k, then $k^2 = k$, so $k = 1$. Conversely, if $[x, y]$ has type 1, it follows from Lemma 4.1 that $x = y$.

If $[x, y]$ has prime type p, then $x \neq y$ (by the above), and if $[x, y]$ contained a third point z, then $p = O(x, z) \cdot O(z, y)$, which is impossible. Conversely, if $[x, y]$ is a two point segment and has type n, then n must be prime, for if it had a nontrivial factor k, then since $\mathbf{R}(P, \sim)$ is full there would be an element $z \in [x, y]$ such that $[x, z]$ would have type k. Finally, it follows from this that for any segment $[x, y]$, the length of any maximal chain is the number of primes in the prime decomposition of $O(x, y)$. This completes the proof.

Let $[x, y]$ be a segment of P of type n, and let C be a maximal chain of $[x, y]$, say $x = x_0 < x_1 < x_2 < \cdots < x_m = y$. If p_i is the type of $[x_{i-1}, x_i]$, then $n = p_1 p_2 \cdots p_m$ is an ordered factorization of n into primes; we call it the *factorization of n induced by C*, or more briefly, the *factorization of C*.

LEMMA 7.2. *Let $\mathbf{R}(P, \sim)$ be a full commutative algebra of Dirichlet type and $[x, y]$ a segment of type n. Let $n = p_1 p_2 \cdots p_m$ be any ordered factorization of n into primes. The number of maximal chains of $[x, y]$ with factorization $p_1 p_2 \cdots p_m$ is given by*

$$B(n) = \begin{Bmatrix} n \\ p_1 \end{Bmatrix} \begin{Bmatrix} n/p_1 \\ p_2 \end{Bmatrix} \begin{Bmatrix} n/p_1 p_2 \\ p_3 \end{Bmatrix} \cdots \begin{Bmatrix} n/p_1 \cdots p_{m-1} \\ p_m \end{Bmatrix} \tag{7.8}$$

and this number depends only on n, not on the factorization chosen.

Hence, if $n = q_1^{a_1} q_2^{a_2} \cdots q_r^{a_r}$ is the canonical factorization of n, then

$$B(n) = \frac{M(n) a_1! \, a_2! \cdots a_r!}{(a_1 + a_2 + \cdots + a_r)!}, \tag{7.9}$$

where $M(n)$ is the number of maximal chains in $[x, y]$.

PROOF. The number of maximal chains with factorization $p_1 p_2 \cdots p_m$ is obviously the expression on the right side of (7.8). By the commutativity relation

$\left\{{s \atop r}\right\} = \left\{{s \atop s/r}\right\}$ and the recursion (7.5),

$$\left\{{n/p_1p_2\cdots p_{k-1} \atop p_k}\right\}\cdot\left\{{n/p_1p_2\cdots p_k \atop p_{k+1}}\right\} \tag{7.10}$$
$$= \left\{{n/p_1p_2\cdots p_{k-1} \atop n/p_1p_2\cdots p_k}\right\}\cdot\left\{{n/p_1p_2\cdots p_k \atop p_{k+1}}\right\}$$
$$= \left\{{n/p_1p_2\cdots p_{k-1} \atop p_{k+1}}\right\}\cdot\left\{{n/p_1p_2\cdots p_{k-1}p_{k+1} \atop n/p_1p_2\cdots p_kp_{k+1}}\right\}$$
$$= \left\{{n/p_1p_2\cdots p_{k-1} \atop p_{k+1}}\right\}\cdot\left\{{n/p_1p_2\cdots p_{k-1}p_{k+1} \atop p_k}\right\}.$$

Hence, $B(n)$ is not changed when p_k and p_{k+1} are interchanged. Since all permutations of $p_1, \cdots p_m$ are generated by such interchanges, the proof follows.

PROPOSITION 7.1. *Let* $\mathbf{R}(P, \sim)$ *be a full algebra of Dirichlet type with types* $n_1 = 1, n_2, \cdots$. *If* $f \in \mathbf{R}(P, \sim)$, *then the map*

$$f \to \sum_k \frac{f(n_k)}{B(n_k)n_k^s} \tag{7.11}$$

of $\mathbf{R}(P, \sim)$ *into formal Dirichlet series is an isomorphism, if when we multiply Dirichlet series we ignore all* β^{-s} *terms when* β *is not some* n_k.

PROOF. Let $[x, y]$ be of type n. For any type $\ell|n$, let $n = p_1 \cdots p_m$ be any factorization with $p_1p_2 \cdots p_k = \ell$. Exactly $B(\ell)$ maximal chains with the factorization $p_1p_2 \cdots p_k$ connect x with a fixed point z such that $[x, z]$ is of type ℓ. Exactly $B(n/\ell)$ maximal chains with the factorization $p_{k+1} \cdots p_m$ connect z with y. Thus, the number of such z is

$$\left\{{n \atop \ell}\right\} = \frac{B(n)}{B(\ell)B(n/\ell)}, \tag{7.12}$$

and the isomorphism follows.

REMARK. As we will see in the next section, when $\mathbf{R}(P, \sim)$ is of full binomial type we know that we can write $B(n) = A(1)A(2) \cdots A(n)$, where $A(n) = \left\{{n \atop 1}\right\}$ is the number of points covered by y in an interval $[x, y]$ of length n. The analogy for full commutative algebras of Dirichlet type is formula (7.8). Here

$$A(k) = \left\{{n/p_1p_2\cdots p_{m-k} \atop p_{m-k+1}}\right\} \tag{7.13}$$

depends on the particular ordered factorization of n into the primes chosen. A canonical choice of $A(k)$ can be specified by the requirement $p_1 \leqq p_2 \leqq \cdots \leqq p_m$.

In certain cases it is possible to know considerably more about the structure of P and $\mathbf{R}(P, \sim)$.

PROPOSITION 7.2. *Let* $\mathbf{R}(P, \sim)$ *be a full commutative algebra of Dirichlet type. Suppose the function* B *is "multiplicative if defined," that is, if* $(m, n) = 1$ *and if* mn *is a type, then* $B(mn) = B(m)B(n)$. *Let* $[x, y]$ *be a segment of type*

$n = p_1^{a_1}p_1^{a_2}\cdots p_m^{a_m}$ and let $[x, x_1], \cdots, [x, x_m]$ be segments of $[x, y]$ of types $p_1^{a_1}, \cdots, p_m^{a_m}$, respectively. Then $[x, y]$ is the product, $[x, y] = [x, x_1] \times [x, x_2] \times \cdots \times [x, x_m]$, and $\mathbf{R}(P, \sim)$ *restricted to $[x, y]$ is given by the tensor product (over κ), $\mathbf{R}([x, y], \sim) = \mathbf{R}([x, x_1], \sim) \otimes \cdots \otimes \mathbf{R}([x, x_m], \sim)$. Each of the algebras $\mathbf{R}([x, x_i], \sim)$ is of full binomial type.*

PROOF. If $1 \leqq i \leqq m$, we have

$$\left\{ n \atop p_i^{a_i} \right\} = \frac{B(n)}{B(p_i^{a_i})B(n/p_i^{a_i})} = \frac{B(n)}{B(n)} = 1. \tag{7.14}$$

Thus, the segments $[x, x_i]$ are unique. If $z \in [x, y]$ and $[x, z]$ is of type $\ell = p_1^{b_1}p_2^{b_2}\cdots p_m^{b_m}$, then as above $\left\{ \ell \atop p_i^{b_i} \right\} = 1$, and z lies above a unique point $z_i \in [x, x_i]$ with $[x, z_i]$ of type $p_i^{b_i}$. Hence, we have a mapping $z \to (z_1, \cdots, z_m)$. Now, the number of $z \in [x, y]$ such that $[x, z]$ is of type $p_1^{b_1}\cdots p_m^{b_m}$ equals

$$\begin{aligned} \left\{ p_1^{a_1}\cdots p_m^{a_m} \atop p_1^{b_1}\cdots p_m^{b_m} \right\} &= \frac{B(p_1^{a_1}\cdots p_m^{a_m})}{B(p_1^{b_1}\cdots p_m^{b_m})B(p_1^{a_1-b_1}\cdots p_m^{a_m-b_m})} \\ &= \frac{B(p_1^{a_1})\cdots B(p_m^{a_m})}{B(p_1^{b_1})\cdots B(p_m^{b_m})B(p_1^{a_1-b_1})\cdots B(p_m^{a_m-b_m})} \\ &= \left\{ p_1^{a_1} \atop p_1^{b_1} \right\} \cdot \left\{ p_2^{a_2} \atop p_2^{b_2} \right\} \cdots \left\{ p_m^{a_m} \atop p_m^{b_m} \right\} \end{aligned} \tag{7.15}$$

which is the number of m-tuples $(z_1, \cdots, z_m)$ with $[x, z_i]$ of type $p_i^{b_i}$. Further, the mapping is injective, as the following argument shows. Suppose z and $\bar{z}$ are distinct elements of $[x, y]$ with $[x, z]$ and $[x, \bar{z}]$ of type $p_1^{b_1}\cdots p_m^{b_m}$, and suppose both z and $\bar{z}$ lie over $z_1, \cdots, z_m$, where $O(x, z_i) = p_i^{b_i}$. Take $w_1 \in [z, y]$ with $O(z, w_1) = p_1^{a_1-b_1}$, $w_2 \in [w_1, y]$ with $O(w_1, w_2) = p_2^{a_2-b_2}, \cdots, w_n \in [w_{n-1}, y]$ with $O(w_{n-1}, w_n) = p_m^{a_m-b_m}$, and similarly take elements $\bar{w}_1, \cdots, \bar{w}_m$ above $\bar{z}$. Note that $w_n = \bar{w}_n = y$, since

$$\begin{aligned} O(x, w_n) &= O(x, z)O(z, w_1)O(w_1, w_2)\cdots O(w_{n-1}, w_n) \\ &= p_1^{a_1}\cdots p_m^{a_m}. \end{aligned} \tag{7.16}$$

However, we show by induction that $w_j \neq \bar{w}_j$ for $1 \leqq j \leqq m$, which gives the desired contradiction.

For $j = 1$, if $w_1 = \bar{w}_1$, then

$$\left\{ p_1^{b_1-a_1}p_2^{b_2}p_3^{b_3}\cdots p_m^{b_m} \atop p_2^{b_2}p_3^{b_3}\cdots p_m^{b_m} \right\} > 1 \tag{7.17}$$

(since $z, \bar{z} \in [z_1, w_1]$), which is not the case as B is multiplicative. Assume $w_{j-1} \neq \bar{w}_{j-1}$ for $j \leqq m$. If $w_j = \bar{w}_j$, then

$$\left\{ p_1^{b_1}\cdots p_{j-1}^{b_{j-1}}p_{j+1}^{b_{j+1}}\cdots p_m^{b_m}p_1^{a_1-b_1}\cdots p_{j-1}^{a_{j-1}-b_{j-1}}p_j^{a_j-b_j} \atop p_1^{b_1}\cdots p_{j-1}^{b_{j-1}}p_{j+1}^{b_{j+1}}\cdots p_m^{b_m}p_1^{a_1-b_1}\cdots p_{j-1}^{a_{j-1}-b_{j-1}} \right\} > 1 \tag{7.18}$$

(since $w_{j-1}, \bar{w}_{j-1} \in [z_j, w_j]$) which is not the case, as B is multiplicative.

Thus the mapping $z \to (z_1, \cdots, z_m)$ is the desired isomorphism $[x, y] \simeq [x, x_1] \times \cdots \times [x, x_m]$, and the rest of the proof follows easily.

As a converse to the above proposition, suppose $\bar{\mathbf{R}}(P_1)$, $\bar{\mathbf{R}}(P_2), \cdots$ are full algebras of binomial type. Let $p_1, p_2, \cdots$ be distinct primes, and let $[x, y] = [(x_1, x_2, \cdots), (y_1, y_2, \cdots)]$ be a segment of $P_1 \times P_2 \times \cdots$, where $[x_i, y_i]$ is of type a_i in $\bar{\mathbf{R}}(P_i)$. Then defining $O(x, y) = p_1^{a_1} p_2^{a_2} \cdots$ gives a full commutative algebra of Dirichlet type such that B is multiplicative if defined.

REMARK. The condition that B is "multiplicative if defined" is equivalent to saying that the Dirichlet series corresponding to the zeta function $\zeta \in \mathbf{R}(P, \sim)$ has an Euler product in the sense that the Dirichlet series

$$\sum_k \frac{1}{B(n_k) n_k^s} - \prod_p \sum_a \frac{1}{B(p^a) p^{as}} \tag{7.19}$$

for some $n_k = p^a$ vanishes at all terms m^{-s} whenever m is a type.

All the usual number theoretic functions such as the Euler totient function ϕ, the number of divisors d, the sum of the divisors σ, and so forth, have analogues in full commutative algebras of Dirichlet type (even in any algebra of Dirichlet type, although some of their properties do not carry over). For instance, if $O(x, y) = n$, we define

$$\begin{aligned} \phi(n) &= \mu * O(n) = \sum_{z \in [x, y]} \mu(x, z) O(z, y), \\ d(n) &= \zeta^2(n) \quad = \sum_{z \in [x, y]} 1, \\ \sigma(n) &= O * \zeta(n) = \sum_{z \in [x, y]} O(x, z), \end{aligned} \tag{7.20}$$

and so on.

These functions, along with μ, will be "multiplicative if defined" if and only if B is also.

PROBLEM. It is easy to construct examples of infinite noncommutative Dirichlet algebras. For instance, let P be the lattice of positive integers under $\leqq$ (a discrete chain). If $m \leqq n$, define

$$O(m, n) = \begin{cases} 2^{n-m} & \text{if } 1 < m, \\ 3 \cdot 2^{n-m-1} & \text{if } 1 = m < n, \\ 1 & \text{if } m = n = 1. \end{cases} \tag{7.21}$$

The corresponding Dirichlet algebra $\mathbf{R}(P, \sim)$ is infinite, that is, there are infinitely many values of $O(m, n)$, and noncommutative.

Suppose, however, we require $\mathbf{R}(P, \sim)$ to have the following properties:

(a) $\mathbf{R}(P, \sim)$ is a full algebra of Dirichlet type, and

(b) any two elements of P have an upper bound. We know of no infinite noncommutative algebras $\mathbf{R}(P, \sim)$ satisfying (a) and (b).

7.3. *Abelian groups.* Suppose G is an abelian group whose lattice P of subgroups gives a Dirichlet algebra $\mathbf{R}(P, \sim)$ if we take $O(x, y)$ to be the order of the quotient group y/x. Then G is finite (since P must be locally finite and every infinite group has infinitely many subgroups), and every Sylow subgroup of G is either cyclic or elementary abelian. Conversely, any such G gives rise to such a Dirichlet algebra, which in fact is of full Dirichlet type whose zeta function has an Euler product.

PROOF. Suppose a Sylow p subgroup of G is not cyclic or elementary abelian. Then it contains a subgroup isomorphic to $Z(p) \oplus Z(p^2)$, where $Z(n)$ denotes the cyclic group of order n. The segments $[0, Z(p^2)]$ and $[0, Z(p) \otimes Z(p)]$ both have type p^2 but are not residually isomorphic, so $\mathbf{R}(P, \sim)$ cannot be of Dirichlet type.

That the converse is true is a straightforward verification.

8. Algebras of full binomial type

8.1. *Structure.* Recall from the previous section that $\mathbf{R}(P, \sim)$ is an algebra of full binomial type if the types are in one to one correspondence with a subset of the nonnegative integers, the type of a segment $[x, y]$ being denoted $O(x, y)$, satisfying:

(A) if $x \leqq z \leqq y$, then $O(x, y) = O(x, z) + O(z, y)$;

(*B*) if n is a type and if $k \leqq n$, then $\left[{n \atop k}\right] \neq 0$, where $\left[{n \atop k}\right]$ is the number of elements z in a segment $[x, y]$ of type n for which $O(x, z) = k$ (and hence, $O(z, y) = n - k$). We then had the following relation

$$\begin{bmatrix} n \\ m \end{bmatrix} \begin{bmatrix} m \\ k \end{bmatrix} = \begin{bmatrix} n \\ k \end{bmatrix} \begin{bmatrix} n - k \\ m - k \end{bmatrix}. \tag{8.1}$$

PROPOSITION 8.1. *Every algebra of full binomial type is commutative.*

PROOF. Suppose $\mathbf{R}(P, \sim)$ is a full algebra of binomial type. We prove by induction on n that $\left[{n \atop m}\right] = \left[{n \atop n-m}\right]$ when $0 < m < n$. Since $\left[{n \atop 0}\right] = \left[{n \atop n}\right] = 1$, it will follow that $\mathbf{R}(P, \sim)$ is commutative.

The statement is clear for $n = 0, 1, 2$. Assume it is true for all $n_0 < n$. Suppose $0 < m < n \leqq 2m$. From the relation (8.1), we have

$$\begin{bmatrix} n \\ m \end{bmatrix} \begin{bmatrix} m \\ n - m \end{bmatrix} = \begin{bmatrix} n \\ n - m \end{bmatrix} \begin{bmatrix} m \\ 2m - n \end{bmatrix}. \tag{8.2}$$

Since $0 < m < n \leqq 2m$ and $\mathbf{R}(P, \sim)$ is full, we have $\left[{m \atop n-m}\right] \neq 0$, $\left[{m \atop 2m-n}\right] \neq 0$. By induction $\left[{m \atop n-m}\right] = \left[{m \atop 2m-n}\right]$. Hence, $\left[{n \atop m}\right] = \left[{n \atop n-m}\right]$. If $0 < m < n$ but $2m < n$, then $0 < n - m < n$ and $n \leqq 2(n - m)$, so again $\left[{n \atop n-m}\right] = \left[{n \atop m}\right]$, and the proof is complete.

LEMMA 8.1. *Let* $\mathbf{R}(P, \sim)$ *be an algebra of binomial type. Then the segments of type* 0 *are precisely the points of P. Moreover, if* $\mathbf{R}(P, \sim)$ *is of full binomial type, then the segments of type* 1 *are those segments of P which contain exactly two points.*

PROOF. If $[x, x]$ is of type n, then $n + n = n$, so $n = 0$. Conversely, if $[x, y]$ is of type 0, it follows from Lemma 4.1 that $x = y$.

If $\mathbf{R}(P, \sim)$ is full and $[x, y]$ is a two point segment of type $n > 0$, then since $\left[{n \atop 1}\right] \neq 0$, we must have $n = 1$. Conversely, by Lemma 4.1 any segment of type 1 contains exactly two points. This completes the proof.

We assume for the rest of this subsection that $\mathbf{R}(P, \sim)$ is a full algebra of binomial type. Let N be the largest type of any segment of P (or $N = \infty$ if there is no largest type). Since $\mathbf{R}(P, \sim)$ is full, we have

$$(8.3)\qquad \begin{bmatrix} 0 \\ 1 \end{bmatrix} = A(0) = 0, \begin{bmatrix} n \\ 1 \end{bmatrix} = A(n) \neq 0, \qquad 1 \leqq n \leqq N \text{ (except } n = \infty).$$

Define $B(n) = A(1)A(2) \cdots A(n)$, with $B(0) = 1$.

Setting $k = 1$ in (8.1) and iterating, we find

$$(8.4)\qquad \begin{bmatrix} n \\ m \end{bmatrix} = \frac{A(n)A(n-1)\cdots A(n-m+1)}{A(m)A(m-1)\cdots A(1)} = \frac{B(n)}{B(m)B(n-m)}, \quad 0 \leqq m \leqq n \leqq N \text{ (except } n = \infty),$$

where we have used the obvious fact that

$$(8.5)\qquad \begin{bmatrix} n \\ 0 \end{bmatrix} = 1, \qquad 0 \leqq n \leqq N, n \neq \infty.$$

We have therefore shown that a full algebra of binomial type is isomorphic to an algebra of formal power series, taken modulo z^{N+1}, the isomorphism being given by

$$(8.6)\qquad f \to \sum_{n=0}^{N} \frac{f(n)}{B(n)} z^n \pmod{z^{N+1}},$$

where $f(n)$ denotes the value $f \in \mathbf{R}(P, \sim)$ takes at any segment of type n. The converse to this statement, and a characterization of full algebras of binomial type, is provided by the next theorem.

THEOREM 8.1. *Suppose P is a locally finite ordered set and $\mathbf{R}(P, \sim)$ a reduced incidence algebra of P with types labelled* 0, 1, 2, $\cdots$, N, $1 \leqq N \leqq \infty$, *such that* (8.6) *is an isomorphism of $\mathbf{R}(P, \sim)$ onto formal power series modulo z^{N+1}. The isomorphism* (8.6) *can be "normalized" by setting $z' = (1/B(1))z$, so we can assume $B(1) = 1$. Then $\mathbf{R}(P, \sim)$ is a full algebra of binomial type and the following hold:*

(i) *P satisfies the Jordan–Dedekind chain condition;*

(ii) *all segments of P of length n have the same number of maximal chains;*

(iii) *a segment of length n is of type n;*

(iv) *in the isomorphism* (8.6) (*normalized to* $B(1) = 1$), *$B(n)$ is the number of maximal chains in a segment of length n and N is the length of P;*

(v) $\mathbf{R}(P, \sim) = \bar{\mathbf{R}}(P)$.

Conversely, if P is a locally finite ordered set satisfying (i) *and* (ii), *then $\bar{\mathbf{R}}(P)$ is a full algebra of binomial type given by* (iii) *and* (iv).

PROOF. Suppose $\mathbf{R}(P, \sim)$ satisfies the hypothesis of the theorem. Then (A) follows from the isomorphism (8.6) using the law of exponents $z^{m+n} = z^m z^n$. Hence, $\mathbf{R}(P, \sim)$ is of binomial type. By the hypothesis that the isomorphism (8.6) is onto, it follows that $\mathbf{R}(P, \sim)$ is a full algebra of binomial type.

Define $g \in \mathbf{R}(P, \sim)$ by

$$g(x, y) = \begin{cases} 1 & \text{if } [x, y] \text{ is of type 1,} \\ 0 & \text{otherwise.} \end{cases} \tag{8.7}$$

Using Lemma 8.1, we see $g^n(x, y)$ is the number of maximal chains of $[x, y]$ of length n. By (8.6), $g^n(x, y) \neq 0$ if and only if $[x, y]$ is of type n. Hence, every maximal chain of $[x, y]$ has length n. Since

$$\left(\frac{z}{B(1)}\right)^n = \frac{B(n)}{B(1)^n} \frac{z^n}{B(n)}, \tag{8.8}$$

then by (8.6), $B(n)$ is the number of maximal chains in an interval of type n when we take $B(1) = 1$. Clearly, N is the length of P. By Lemma 4.1 (ii), $\mathbf{R}(P, \sim) = \bar{\mathbf{R}}(P)$. This proves (i), (ii), (iii), (iv), (v) of the theorem.

Conversely, suppose P satisfies (i) and (ii). (Actually, (i) and (ii) follow easily from the slightly weaker condition that all segments of P of the same minimum length contain the same number of maximal chains.) Let $B(n)$ be the number of maximal chains in a segment of length n. Then each segment $[x, y]$ of length n contains $B(n)/B(k)B(n - k)$ points of height k, since $B(k)B(n - k)$ maximal chains in $[x, y]$ pass through a point of height k. Thus, if $f, g \in \mathbf{I}(P)$ depend only on the length n of any segment $[x, y]$, we have

$$(f*g)(n) = (f*g)(x, y) = \sum_{k=0}^{n} \frac{B(n)}{B(k)B(n - k)} f(k)g(n - k), \tag{8.9}$$

which is a function of n only. Thus, specifying all segments of the same length to be of the same type gives a reduced incidence algebra $\mathbf{R}(P, \sim)$, which by Lemma 4.1 (ii) must be $\bar{\mathbf{R}}(P)$. The isomorphism (8.6) now follows immediately from (8.9). We have proved that if (8.6) holds, then $\mathbf{R}(P, \sim)$ is of full binomial type, so the proof is complete.

COROLLARY 8.1. *If P is a locally finite ordered set and if every segment of P of the same minimum length is of the same type in $\bar{\mathbf{R}}(P)$, then $\bar{\mathbf{R}}(P)$ is a full algebra of binomial type.*

PROOF. By Lemma 4.1, all segments of P of the same minimum length contain the same number of maximal chains, since they are of the same type. We have already remarked that it is easy to prove from this that P satisfies the Jordan–Dedekind chain condition. The proof now follows from Theorem 8.1.

REMARK. Suppose $\mathbf{R}(P, \sim)$ is of full binomial type. By the previous theorem any two segments of P of the same length are r isomorphic. Moreover, any segment of P is "r self dual", that is, is r isomorphic to its dual, since $\mathbf{R}(P, \sim) \simeq \mathbf{R}(P^*, \sim)$, when P is of full binomial type and P^* is the dual of P.

A further characterization of full algebras $\mathbf{R}(P, \sim)$ of binomial type, at least when P does not have arbitrarily large chains, is given by the next proposition.

PROPOSITION 8.2. *Suppose $\mathbf{R}(P, \sim)$ is a reduced incidence algebra of a locally finite ordered set P with 0 which when considered as an algebra with identity over the ground field (which we have been assuming has characteristic 0) is generated* *by a single function f. Then $\mathbf{R}(P, \sim)$ is a full algebra of binomial type, and there is an integer N such that the longest chain in P has length N.*

Conversely, if $\mathbf{R}(P, \sim)$ is a full algebra of binomial type and if the longest chain in P has finite length N, then $\mathbf{R}(P, \sim)$ is generated by any function $f \in \mathbf{R}(P, \sim)$ not vanishing on segments of length 1 (for example, $f = \zeta$).

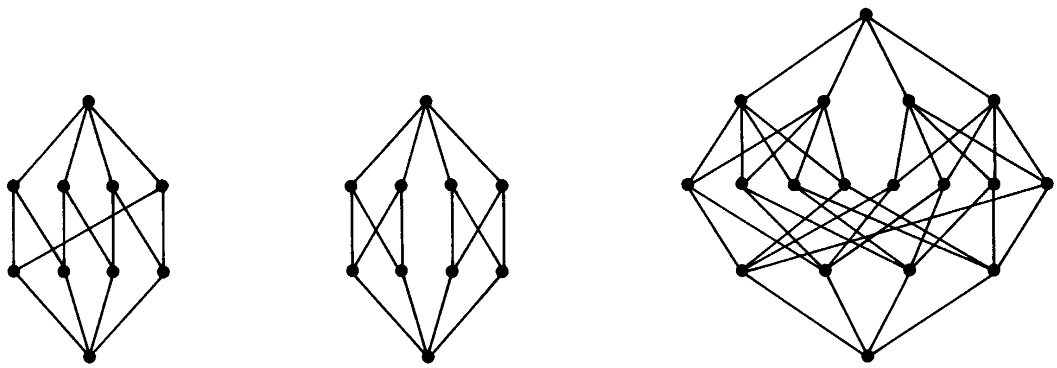
PROOF. Suppose f generates $\mathbf{R}(P, \sim)$. We first show that all points of P belong to the same equivalence class relative to $\sim$. Otherwise, since P has a 0, there is a two point segment $[x, y]$ of P such that $[x, x]$ and $[y, y]$ are not equivalent. Hence $\mathbf{R}(P, \sim)$, when restricted to $[x, y]$, has dimension three as a vector space. But if $f(x, x) = a$ and $f(y, y) = b$, then $(f - a)(f - b)$ vanishes on all three subsegments of $[x, y]$ and hence f generates, together with the identity, a vector space of dimension $\leqq$ two when restricted to $[x, y]$. This contradiction shows $[x, x] \sim [y, y]$ and hence all points of P are equivalent.

If P contains arbitrarily long chains, then $\mathbf{R}(P, \sim)$ has uncountable dimension as a vector space, while f generates a vector space of countable dimension. Hence, there is an integer N such that the longest chain in P has length N. The preceding paragraph shows that f is constant on points, say $f(x, x) = a$. Then $(f - a)^{N+1} = 0$. Hence, f, together with the identity, generates a vector space of dimension $\leqq N + 1$. Since two segments of different maximum lengths must be of different types, it follows that any two segments of the same maximum length are of the same type (because the dimension of $\mathbf{R}(P, \sim)$ is equal to the number of types). It then follows from Corollary 8.1 that $\mathbf{R}(P, \sim)$ is a full algebra of binomial type.

The converse is a trivial consequence of the isomorphism (8.6), and the proof is complete.

8.2. *Lattices of full binomial type.* An ordered set P is said to be of full binomial type if it satisfies (i) and (ii) of Theorem 8.1.

Examples of ordered sets P of full binomial type are discrete chains with 0, lattices of finite subsets of a set, and lattices of finite subspaces of a projective space. Various other examples are given in Figure 5.



(A) (B) (C)

FIGURE 5
Ordered sets of full binomial type.

The ordered sets (A) and (B) have isomorphic reduced incidence algebras of full binomial type, although they are not isomorphic as ordered sets. In fact, (A) is a lattice and (B) is not. The ordered set (C) has two interesting properties: not all its segments of the same length are isomorphic (it has 3 segments isomorphic to (A) and (B)), and its Möbius function (see *Foundations I*) does not alternate in sign.

We now prove some results relating the structure of P to the numbers $B(1)$, $B(2), \cdots$.

PROPOSITION 8.3. *Let P be of full binomial type. An n segment of P is a chain if and only if $B(n) = 1$.*

The proof is obvious.

PROPOSITION 8.4. *Let L be a* lattice *of full binomial type. Every element of L is the join of atoms (that is, L is* atomic*) if and only if $A(2) > 1$.*

PROOF. If $A(2) = 1$, then a 2 segment is a chain; hence any element of L of height 2 is not the join of atoms.

Conversely, suppose L is not atomic and let y be an element of L of minimum height $n > 1$ which is not the join of atoms. Let x be an element of height $n - 2$ lying below y. Then $[x, y]$ is a chain of length 2, so $A(2) = 1$.

PROPOSITION 8.5. *Let L be a lattice of full binomial type and $[x, y]$ an n segment of L. The join of any two distinct atoms of $[x, y]$ is of height 2 if and only if*

$$A(k) = 1 + (A(2) - 1) + (A(2) - 1)^2 + \cdots + (A(2) - 1)^{k-1} \tag{8.10}$$

when $1 \leqq k \leqq n$.

PROOF. Suppose the join of any two distinct atoms of $[x, y]$ has height 2. Then the same is true for $[x, y']$, where y' is any point of $[x, y]$ of height $k \leqq n$. Now any element of $[x, y']$ of height 2 lies above $A(2)$ atoms and $\binom{A(2)}{2}$ pairs of atoms. But $[x, y']$ contains $\left[{k \atop 2}\right]$ elements of height 2 and $\binom{A(k)}{2}$ pairs of atoms. Hence,

$$\binom{A(k)}{2} = \begin{bmatrix} k \\ 2 \end{bmatrix} \binom{A(2)}{2}, \tag{8.11}$$

which implies $A(k) = A(k - 1)(A(2) - 1) + 1$. By induction

$$A(k) = 1 + (A(2) - 1) + (A(2) - 1)^2 + \cdots + (A(2) - 1)^{k-1}, \qquad 1 \leqq k \leqq n. \tag{8.12}$$

Conversely, if two atoms of $[x, y']$ have join of height > 2, then the above argument yields

$$\binom{A(k)}{2} > \begin{bmatrix} k \\ 2 \end{bmatrix} \binom{A(2)}{2}. \tag{8.13}$$

Consequently,

$$A(k) > 1 + (A(2) - 1) + \cdots + (A(2) - 1)^{k-1}, \tag{8.14}$$

and the proof is complete.

LEMMA 8.2. *Let L be a lattice of full binomial type such that the join of any two distinct atoms of L has height 2. Then L is modular.*

PROOF. Let x, y be two elements of L such that x and y cover $x \wedge y$. (If no such x, y exist, then L is a chain and hence modular.) Let n be the length of $[x \wedge y, x \vee y] = L'$. Then L' is a lattice of full binomial type whose invariants $B(1), B(2), \cdots B(n)$ are the same as those for L. Hence, by Proposition 8.5, the join of any two distinct atoms of L' has height 2; in particular, $x \vee y$ has height 2 and thus covers x and y. This means L is upper semimodular. Dually, if x and y are covered by $x \vee y$, then the same argument applied to the dual of $[x \wedge y, x \vee y]$ shows that L is lower semimodular. Hence, L is modular and the proof is complete.

Finally, we come to the main theorem of this subsection.

THEOREM 8.2. *Let L be a lattice of full binomial type, such that the join of any two atoms of L has height 2. Then L is isomorphic to either:*

(i) *a chain;*

(ii) *the lattice of finite subsets of a set; or to*

(iii) *the lattice of finite subspaces of a projective space.*

PROOF. Suppose L is not a chain. Then L, being of binomial type, has two distinct atoms whose join has height 2; hence, $A(2) > 1$. By Proposition 8.4, L is atomic. By Lemma 8.2, L is modular. Thus, every segment $[x, y]$ of L is a modular geometric lattice. By Birkhoff (Theorem IV–10), $[x, y]$ is the product of a Boolean algebra with projective geometries. The only such products which are of full binomial type are the single factor ones, that is, $[x, y]$ is of the type (ii) or (iii). Since every segment $[0, x]$ of L is of the type (ii) or (iii), so is L, and the proof is complete.

9. Algebras of triangular type

In this section, we investigate locally finite ordered sets P with 0 which have a reduced incidence algebra $\mathbf{R}(P, \sim)$ which is isomorphic, in a natural way, to the algebra of all upper triangular $N \times N$ matrices (possibly $N = \infty$) over the ground field of $\mathbf{R}(P, \sim)$. First we describe a class of such P. Let P be a locally finite ordered set with 0 satisfying the Jordan–Dedekind chain condition. If $[x, y]$ is a segment of P with x of height m and y of height n, we call $[x, y]$ an (m, n) segment. Suppose that for all m, n any two (m, n) segments contain the same number $B(m, n)$ of maximal chains. (By convention $B(n, n) = 1$ if P contains an element of height n.) We then call P an ordered set of *triangular type*. Geometric lattices of triangular type are considered by Edmonds, Murty, and Young [20] under a different name.

PROPOSITION 9.1. *The equivalence relation on the segments of an ordered set P of triangular type defined by $[x, y] \sim [x', y']$, if and only if $[x, y]$ and $[x', y']$ are both (m, n) segments for some m, n, gives a reduced incidence algebra $\mathbf{R}(P, \sim)$. If $f(m, n)$ denotes the value that $f \in \mathbf{R}(P, \sim)$ takes on an (m, n) segment, then the* *mapping*

$$
(9.1)\qquad f \to \begin{pmatrix} \frac{f(0,0)}{B(0,0)} & \frac{f(0,1)}{B(0,1)} & \frac{f(0,2)}{B(0,2)} & \cdots \\ 0 & \frac{f(1,1)}{B(1,1)} & \frac{f(1,2)}{B(1,2)} & \cdots \\ 0 & 0 & \frac{f(2,2)}{B(2,2)} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}
$$

is an isomorphism of $\mathbf{R}(P, \sim)$ *onto the algebra of all upper triangular* $N \times N$ *matrices, where* N *is the height of* P *(possibly* ∞*).*

PROOF. Let $[x, y]$ be an (m, n) segment. The number of points $z \in [x, y]$ such that $[x, z]$ is an (m, m') segment and $[z, y]$ is an (m', n) segment is given by $B(m, n)/B(m, m')B(m', n)$. Thus, if f, g are constant on equivalence classes relative to $\sim$, we have

$$
(9.2)\qquad (f*g)(x, y) = \sum_{m'=m}^{n} \frac{B(m, n)}{B(m, m')B(m', n)} f(m, m')g(m', n),
$$

which is a function of m, n only. Hence, $\sim$ gives a reduced incidence algebra, and (9.2) is the condition for (9.1) to be an isomorphism.

The converse of Proposition 9.1 is provided by the next proposition.

PROPOSITION 9.2. *Let* P *be a locally finite ordered set with* 0 *and* $\mathbf{R}(P, \sim)$ *a reduced incidence algebra whose types can be labeled by ordered pairs* (m, n), $0 \leqq m \leqq n$, *such that whenever* (m, n) *is a type and* $0 \leqq m' \leqq n' \leqq n$, *then* (m', n') *is a type. Suppose there are numbers* $B(m, n)$ *for every type* (m, n) *such that the mapping* (9.1) *is an isomorphism of* $\mathbf{R}(P, \sim)$ *onto the algebra of all upper triangular* $N \times N$ *matrices for some* $N \leqq \infty$. *Then the following hold:*

(i) *P satisfies the Jordan–Dedekind chain condition;*

(ii) $B(n, n) = 1$ *whenever* (n, n) *is a type;*

(iii) *we can take new values of* $B(m, n)$ *preserving the isomorphism* (9.1) *such that* $B(n, n+1) = 1$ *whenever* $(n, n+1)$ *is a type;*

(iv) *every* (m, n) *segment of* P *contains the same number of maximal chains, and when the isomorphism* (9.1) *is normalized by* (iii), *then* $B(m, n)$ *is the number of maximal chains in an* (m, n) *segment.*

PROOF. Define $h_{m,n} \in \mathbf{R}(P, \sim)$ by

$$
(9.3)\qquad h_{m,n}(x, y) = \begin{cases} 1 & \text{if } [x, y] \text{ is an } (m, n) \text{ segment}, \\ 0 & \text{otherwise}. \end{cases}
$$

It follows from (9.1) that

$$
(9.4)\qquad h_{m,k}h_{k,n} = \frac{B(m, n)}{B(m, k)B(k, n)} h_{m,n}.
$$

Thus, if $[x, y]$ is an (m, k) segment and $[y, z]$ a (k, n) segment, then $[x, z]$ is an (m, n) segment. Conversely, if $[x, z]$ is an (m, n) segment and $m \leqq k \leqq n$, then there is a point $y \in [x, z]$ such that $[x, y]$ is an (m, k) segment and $[y, z]$ a (k, n) segment. It follows that points are (n, n) segments for some n and that two point segments are $(n, n + 1)$ segments for some n. Moreover, every maximal chain in an (m, n) segment has length $n - m$, which proves (i).

Since the identity of $\mathbf{R}(P, \sim)$ goes into the identity matrix under (9.1), we have $B(n, n) = 1$ whenever (n, n) is a type, proving (ii).

If $B(m, n)$ is replaced by

$$\frac{B(m, n)}{B(m, m + 1)B(m + 1, m + 2) \cdots B(n - 1, n)}, \tag{9.5}$$

then the isomorphism (9.1) is preserved and $B(n, n + 1)$ is replaced by 1. This proves (iii).

Hence, suppose each $B(n, n + 1) = 1$ whenever $(n, n + 1)$ is a type. Let $\eta \in \mathbf{R}(P, \sim)$ be the function which is 1 on two point segments and 0 elsewhere, that is,

$$\eta(x, y) = \begin{cases} 1 & \text{if } [x, y] \text{ is an } (n, n + 1) \text{ segment for some } n, \\ 0 & \text{otherwise.} \end{cases} \tag{9.6}$$

If $[x, y]$ is an (m, n) segment, then $\eta^{n-m}(x, y)$ is the number of maximal chains in $[x, y]$, so that this number depends only on m and n. Using (9.1), $\eta^{n-m}(x, y) = B(m, n)$, so (iv) is proved.

Propositions 9.1 and 9.2 give a characterization of ordered sets P which have a reduced incidence algebra isomorphic to the algebra of all upper triangular $N \times N$ matrices, namely P is of triangular type. If we assume P is a lattice, then some of the structure of P can be inferred from the numbers $B(m, n)$.

PROPOSITION 9.3. *Let L be a lattice of triangular type. Set $T(n) = B(n, n + 2) - 1$.*

(i) *If $T(n) \neq 0$ for every type $(n, n + 2)$, then L is atomic (that is, every element of L is the join of atoms); the converse is true if L is semimodular;*

(ii) *L is upper semimodular if and only if for all types (m, n),*

$$\frac{B(m, n)}{B(m + 1, n)} = 1 + T(m) + T(m)T(m + 1) + T(m)T(m + 1)T(m + 2) + \cdots + T(m)T(m + 1) \cdots T(n - 2); \tag{9.7}$$

(iii) *L is lower semimodular if and only if for all types (m, n),*

$$\frac{B(m, n)}{B(m, n - 1)} = 1 + T(n - 2) + T(n - 2)T(n - 3) + \cdots + T(n - 2)T(n - 3) \cdots T(m). \tag{9.8}$$

PROOF. For (i), suppose L is not atomic, and let $y \in L$ be a join irreducible of L of height $n + 2 > 1$. If x is any element of height n lying below y, then $[x, y]$ is a chain, so $T(n) = 0$. The converse will be proved after (ii) and (iii).

For (ii), L is upper semimodular if and only if whenever x and y cover $x \wedge y$, then $x \vee y$ covers x and y; that is, if and only if in every (m, n) segment, the join of any two distinct atoms has height 2. Now an (m, n) segment has $B(m, n)/B(m + 1, n) = A(m, n)$ atoms and $\binom{A(m,n)}{2}$ pairs of distinct atoms. Moreover, each element of height 2 in an (m, n) segment covers $B(m, m + 2)$ atoms, and hence $\binom{B(m,m+2)}{2}$ pairs of distinct atoms. Since an (m, n) segment has $B(m, n)/B(m, m + 2)B(m + 2, n)$ elements of height 2, we see that L is upper semimodular if and only if

$$(9.9)\qquad \binom{A(m, n)}{2} = \frac{B(m, n)}{B(m, m + 2)B(m + 2, n)}\binom{B(m, m + 2)}{2}$$

for all types (m, n). Simplifying (9.9) gives

$$\begin{aligned}(9.10)\qquad \frac{B(m, n)}{B(m + 1, n)} &= 1 + T(m)\frac{B(m + 1, n)}{B(m + 2, n)}\\ &= 1 + T(m)\big(1 + T(m + 1)\big)\frac{B(m + 2, n)}{B(m + 3, n)}\\ &\ \ \vdots\\ &= 1 + T(m) + T(m)T(m + 1) + \cdots\\ &\qquad + T(m)T(m + 1)\cdots T(n - 2).\end{aligned}$$

Case (iii) is the dual of (ii).

We now prove the second part of (i); that is, if L is semimodular and $T(m) = 0$ for some type $(m, m + 2)$, then L is not atomic. Say L is upper semimodular. (The dual argument works when L is lower semimodular.) We show that there is only one element of L of height $m + 1$. Suppose there are two elements of L of height $m + 1$, and let $n > m + 1$ be the height of their join. We prove by "descending induction" on k that

$$(9.11)\qquad \frac{B(k, n)}{B(k, m + 1)B(m + 1, n)} = 1$$

when $0 \leqq k \leqq m + 1$. The case $k = 0$ asserts that a $(0, n)$ segment has only one element of height $m + 1$, a contradiction.

Clearly, (9.11) holds for $k = m + 1$. Assume it holds for $k + 1$ with $0 < k + 1 \leqq m + 1$. By (ii),

$$\begin{aligned}(9.12)\qquad &\frac{B(k, n)}{B(k, m + 1)B(m + 1, n)}\\ &= \frac{B(k + 1, n)}{B(k + 1, m + 1)B(m + 1, n)}\\ &\qquad\cdot\frac{\big(1 + T(k) + T(k)T(k + 1) + \cdots + T(k)\cdots T(n - 2)\big)}{\big(1 + T(k) + T(k)T(k + 1) + \cdots + T(k)\cdots T(m - 1)\big)}.\end{aligned}$$

By assumption,

$$B(k+1,n) \Big/ \ldots \quad (9.13) \qquad \frac{B(k+1,n)}{B(k+1,m+1)B(m+1,n)} = 1.$$

Since $T(m) = 0$,

$$(9.14) \qquad 1 + T(k) + \cdots + T(k)\cdots T(n-2) = 1 + T(k) + \cdots + T(k)\cdots T(m-1).$$

Hence, $B(k, n)/B(k, m+1)B(m+1, n) = 1$, and the proof follows.

If L is a, say, upper semimodular lattice of triangular type, then Proposition 9.3 (ii) expresses $B(m, n)$ in terms of $B(m+1, n)$ and the $T(k)$. By iteration, we can in fact express $B(m, n)$ in terms of the $T(k)$ only, namely

$$(9.15) \qquad B(m,n) = \prod_{i=0}^{n-m-2} [1 + T(m+i) + T(m+i)T(m+i+1) + \cdots + T(m+i)\cdots T(n-2)], \qquad n \geqq m+2.$$

A dual formula holds for lower semimodularity.

The proof of the second part of Proposition 9.3(i) reduces the theory of semimodular lattices of triangular type to that of atomic semimodular lattices. In fact, we have the following theorems.

THEOREM 9.1. *Let L be an upper semimodular lattice of triangular type. Then there are geometric lattices (that is, upper semimodular atomic lattices of finite length) $L_1, L_2, \cdots, L_r$ of triangular type and an upper semimodular atomic lattice L_{r+1} of triangular type, such that L is isomorphic to the lattice obtained by identifying the top of L_i with the bottom of L_{i+1} for $1 \leqq i \leqq r$.*

THEOREM 9.2. *If L is a modular lattice of triangular type, then the lattices $L_1, \cdots, L_{r+1}$ of Theorem 9.1 are either Boolean algebras or projective geometries.*

Theorem 9.2 follows from the well-known structure theorem for modular geometric lattices (Birkhoff, Theorem IV–10). Any such lattice is the product of a Boolean algebra and projective geometries, and it is easily seen that this is of triangular type if and only if the product has only one factor.

EXAMPLE 9.1. *Chains.* Discrete chains with 0 are modular lattices of triangular type. Each lattice L_i of Theorem 9.1 consists of two points. Here $B(m, n) = 1$ whenever (m, n) is a type, or equivalently $T(n) = 0$ whenever $(n, n+2)$ is a type.

EXAMPLE 9.2. *Projective geometries.* The lattice of finite subspaces of a projective geometry with $q + 1$ points on a line is a modular lattice of triangular type with $T(n) = q$ whenever $(n, n+2)$ is a type.

EXAMPLE 9.3. *Boolean algebras.* These are modular lattices of triangular type with $T(n) = 1$ whenever $(n, n+2)$ is a type.

Examples 9.1, 9.2, and 9.3 all have the property that $B(m, n)$ depends only on $n - m$ when (m, n) is a type. Such ordered sets are of full binomial type defined in the previous section. It is proved there that a semimodular lattice of full binomial type is one of Example 9.1, 9.2, 9.3.

EXAMPLE 9.4. *Affine geometries.* The lattice of finite affine subspaces of an affine space with q points on a line is an upper semimodular (but not modular unless there is only one line) lattice of triangular type with $T(0) = q - 1$, $T(n) = q, n > 0$, when $(n, n + 2)$ is a type.

EXAMPLE 9.5. Various ways of putting together and taking apart ordered sets of triangular type give other ordered sets of triangular type. The simplest examples are (a) segments, (b) identifying all elements above or below a certain level to a single element (called upper or lower truncation), and (c) identifying the top of an ordered set of triangular type with 1 with the bottom of an ordered set of triangular type. All of these operations except lower truncation preserve upper semimodularity.

EXAMPLE 9.6. *Block designs.* Let L be a geometric lattice of triangular type of height 3. If we regard the atoms of L as objects and the co-atoms as blocks containing the atoms they cover, then L determines a balanced incomplete block design with parameters

$$\begin{aligned} v &= 1 + T(0) + T(0)T(1) = 1 - B(1, 3) + B(0, 2)B(1, 3), \\ b &= \frac{B(1, 3)}{B(0, 2)} v, \\ k &= B(0, 2), \\ r &= B(1, 3), \\ \lambda &= 1. \end{aligned} \tag{9.16}$$

Conversely, any balanced incomplete blocks design with $\lambda = 1$ determines a geometric lattice of triangular type of height 3. Thus, geometric lattices of triangular type can be regarded as generalizations of $\lambda = 1$ block designs. Proposition 9.3 is then the generalization of the well-known relations $bk = vr$ and $r(k - 1) = v - 1$ holding in any block design with $\lambda = 1$ (see Hall, [30], Chapter 10).

EXAMPLE 9.7. Miscellaneous other examples and a method for classifying them, can be found in the paper of Edmonds, Murty, and Young [20].

REFERENCES

This bibliography does not repeat any references listed in, *On the Foundations of Combinatorial Theory*, I, II, III, and IV.

[1] S. A. AMITSUR, "On arithmetic functions," *J. Analyse Math.*, Vol. 5 (1956–1957), pp. 273–314.

[2] C. ARF, K. İMRE, and E. ÖZİZMİR, "On the algebraic structure of the cluster expansion in statistical mechanics," *J. Math. Physics*, Vol. 6 (1965), pp. 1179–1188.

[3] E. A. BENDER, "Numerical identities in lattices with an application to Dirichlet products," *Proc. Amer. Math. Soc.*, Vol. 15 (1964), pp. 8–13.

[4] E. A. BENDER and J. R. GOLDMAN, "Enumerative uses of generating functions," *Indiana Univ. Math. J.*, Vol. 20 (1971), pp. 753–765.

[5] R. BUSCHMAN, "The convolution ring of sequences," *Amer. Math. Monthly*, Vol. 74 (1967), pp. 284–286.

[6] L. Carlitz, "Rings of arithmetic functions," *Pacific J. Math.*, Vol. 14 (1964), pp. 1165–1171.
[7] ———, "Arithmetic functions in an unusual setting," *Amer. Math. Monthly*, Vol. 74 (1966), pp. 582–590.
[8] E. D. Cashwell and C. J. Everett, "The rings of number-theoretic functions," *Pacific J. Math.*, Vol. 9 (1959), pp. 975–985.
[9] ———, "Formal power series," *Pacific J. Math.*, Vol. 13 (1963), pp. 45–64.
[10] S. Chowla, I. N. Herstein, and W. R. Scott, "The solutions of $x^d = 1$ in symmetric groups," *Norske Vid. Selsk. Forh. (Trondheim)*, Vol. 25 (1952), pp. 29–31.
[11] S. Chowla, I. N. Herstein, and W. K. Moore, "On recursions connected with symmetric groups, I," *Canad. J. Math.*, Vol. 3 (1951), pp. 328–334.
[12] E. Cohen, "Arithimetical functions associated with the unitary divisors of an integer," *Math. Z.*, Vol. 74 (1960), pp. 66–80.
[13] ———, "Unitary products of arithmetical functions," *Acta Arith.*, Vol. 7 (1961), pp. 29–38.
[14] H. H. Crapo, "The Möbius function of a lattice," *J. Combinatorial Theory*, Vol. 1 (1966), pp. 126–131.
[15] ———, "Möbius inversion in lattices," *Arch. Math.*, Vol. 19 (1968), pp. 595–607.
[16] H. H. Crapo and G.-C. Rota, "On the foundations of combinatorial theory, II: combinatorial geometries," *Studies in Appl. Math.*, Vol. 49 (1970), pp. 109–133.
[17] ———, *On the Foundations of Combinatorial Theory: Combinatorial Geometries*, Cambridge, Massachusetts Institute of Technology Press, Preliminary edition, 1970.
[18] T. M. R. Davidson, "On arithmetic convolutions, " *Canad. Math. Bull.*, Vol. 9 (1966), pp. 287–296.
[19] T. Dowling, "F-partition lattices," to appear.
[20] J. Edmonds, U. S. R. Murty, and P. Young, "Equicardinal matroids and matroid-designs," *University of Waterloo Notes*, Ontario, 1970.
[21] R. W. Frucht and G.-C. Rota, "La function de Möbius para partitiones de un conjuncto," *Scientia (Valparaíso)*, Vol. 122 (1963), pp. 111–115.
[22] ———, "Polonomios de Bell y partitiones de conjunctos finitos," *Scientia (Valparaíso)*. Vol. 126 (1965), pp. 5–10.
[23] M. Gessley, "A generalized arithmetic convolution," *Amer. Math. Monthly*, Vol. 74 (1967), pp. 1216–1217.
[24] A. A. Gioia, "The *R*-product of arithmetic functions," *Canad. J. Math.*, Vol. 17 (1965), pp. 970–976.
[25] J. Goldman and G.-C. Rota, "On the foundations of combinatorial theory, IV: finite-vector spaces and Eulerian generating functions," *Studies in Appl. Math.*, Vol. 49 (1970), pp. 239–258.
[26] D. L. Goldsmith, "On the multiplicative properties of arithmetic functions," *Pacific J. Math.*, Vol. 27 (1968), pp. 238–304.
[27] J. A. Green, "The characters of the finite general linear groups." *Trans. Amer. Math. Soc.*, Vol. 80 (1955), pp. 402–447.
[28] ———, "Symmetric functions and p-modules," Lecture Notes, 1965.
[29] O. A. Gross, "Preferential arrangements," *Amer. Math. Monthly*, Vol. 69 (1962), pp. 4–8.
[30] M. Hall, Jr., *Combinatorial Theory*, Waltham, Blaisdell, 1967.
[31] P. Hall, "The algebra of partitions," *Proceedings of the Fourth Canadian Mathematics Congress, 1957, Banff*, 1959, pp. 147–159.
[32] E. Hewitt and H. S. Zuckerman, "The L_1-algebra of a commutative semigroup," *Trans. Amer. Math. Soc.*, Vol. 83 (1956), pp. 77–97.
[33] L. A. Hinrichs, "A note on locally finite refinements of order," to appear.
[34] V. E. Hoggart, Jr. and D. A. Lind, "Fibonacci and binomial properties of weighted compositions," *J. Combinatorial Theory*, Vol. 4 (1968), pp. 121–124.
[35] N. Jacobson, *Structure of Rings*, Providence, American Mathematical Society Colloquium Publication, Vol. 37, 1956.
[36] D. A. Klarner, "A ring of sequences generated by rational functions," *Amer. Math. Monthly*, Vol. 74 (1967), pp. 813–816.

[37] T. Klein, "The Hall polynomial," *J. Algebra*, Vol. 12 (1969), pp. 61–78.
[38] C. L. Liu, *Introduction to Combinatorial Mathematics*, New York, McGraw-Hill, 1968.
[39] L. Moser and E. L. Whitney, "Weighted compositions," *Canad. Math. Bull.*, Vol. 4 (1961), pp. 39–43.
[40] R. Mullin and G.-C. Rota, "On the foundations of combinatorial theory, III: theory of binomial enumeration," *Graph Theory and Its Applications* (edited by B. Harris), New York, Academic Press, 1970, pp. 167–213.
[41] W. Narkievica, "A class of arithmetical convolutions," *Colloq. Math.*, Vol. 10 (1963), pp. 81–94.
[42] S. Newman, "Measure algebras and functions of bounded variation on idempotent semigroups," to appear.
[43] I. Niven, "Formal power series," *Amer. Math. Monthly*, Vol. 76 (1969), pp. 871–889.
[44] I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers*, New York, Wiley, 1960.
[45] G. N. Raney, "Functional composition patterns and power series reversion," *Trans. Amer. Math. Soc.*, Vol. 94 (1960), pp. 441–450.
[46] D. Rearick, "Semi-multiplicative functions," *Duke Math. J.*, Vol. 33 (1966), pp. 623–628.
[47] ———, "Correlation of semi-multiplicative functions," *Duke Math. J.*, Vol. 33 (1966), pp. 623–628.
[48] ———, "Operators on algebras of arithmetic functions," *Duke Math. J.*, Vol. 35 (1968), pp. 761–766.
[49] G.-C. Rota, "On the foundations of combinatorial theory, I: theory of Möbius functions," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 2 (1964), pp. 340–368.
[50] H. Scheid, "Einige Ringe zahlentheoretischer Funktionen," *J. Reine Angew. Math.*, Vol. 237 (1969), pp. 1–13.
[51] ———, "Über ordnungstheoretische Funktionen," *J. Reine Angew. Math.*, Vol. 238 (1969), pp. 1–13.
[52] ———, "Funktionen über lokal endlichen Halbordnungen, I," *Monatsh. Math.*, Vol. 74 (1970), pp. 336–347.
[53] ———, "Über die Inzidenzalgebra einer lokal endlichen Halbordnung," Habilitationsschrift, Mainz, 1970.
[54] H. Scheid and R. Sivaramakrishnan, "Certain classes of arithmetic functions and the operation of additive convolution," *J. Reine Angew. Math.*, Vol. 245 (1970), pp. 201–207.
[55] D. A. Smith, "Incidence functions as generalized arithmetic functions, I," *Duke Math. J.*, Vol. 34 (1967), pp. 617–634.
[56] ———, "Incidence functions as generalized arithmetic functions, II," *Duke Math. J.*, Vol. 36 (1969), pp. 15–30.
[57] ———, "Incidence functions as generalized arithmetic functions, III," *Duke Math. J.*, Vol. 36 (1969), pp. 343–368.
[58] R. P. Stanley, "Structure of incidence algebras and their automorphism groups," *Amer. Math. Soc. Bull.*, Vol. 76 (1970), pp. 1236–1239.
[59] E. C. Titchmarsh, *The Theory of the Riemann-Zeta Function*, Oxford, Oxford University Press, 1951.
[60] R. Vaidyanathaswamy, "The theory of multiplicative arithmetical functions," *Trans. Amer. Math. Soc.*, Vol. 33 (1931), pp. 579–662.
[61] M. Ward, "Arithmetic functions on rings," *Ann. of Math.*, Vol. 38 (1937), pp. 725–732.

Studies in Applied Mathematics, Vol. LI, No. 4, December 1972.

On the Foundations of Combinatorial Theory. VII: Symmetric Functions through the Theory of Distribution and Occupancy*

By Peter Doubilet
(Communicated by G.-C. Rota)

1. Introduction

Our purpose in this paper is to derive many of the known results about symmetric functions, and a few new ones, using techniques involving the lattice of partitions of a set. There are a number of advantages to be obtained from this approach. The results come out in a more simple and elegant manner than in more standard approaches, thereby giving, it is hoped, more insight into their meaning. It also gives new interpretations to various formulas and statements. One further possibility is that the present line of attack could be extended to deal with a class of symmetric functions not discussed here, the so-called Schur functions, and in doing this develop the theory of the linear representations of the symmetric group in a very beautiful manner.

In Section 3 we develop the tools to study the monomial symmetric functions, the elementary symmetric functions, and the power sum symmetric functions. Among the results we prove is what is sometimes called the fundamental theorem of symmetric functions, namely that the elementary symmetric functions form a basis for the vector space of all symmetric functions. The standard proof of this fact is an induction argument, but we prove it neatly here by Möbius inversion. We also show that the matrix of coefficients expressing the elementary symmetric functions in terms of the monomial symmetric functions is symmetric.

In Section 4 we extend our techniques to allow us to deal with the complete homogeneous symmetric functions and a new type of symmetric functions which are "dual" to the monomial symmetric functions in the same sense that the elementary symmetric functions are dual to the homogeneous symmetric functions. One of the results proved in this section is one which we believe to be new, namely that under the isometry θ of the space of symmetric functions introduced by Philip Hall, every monomial in the image under θ of a monomial symmetric function appears with the same sign, either positive or negative.

Noting that many of the results obtained in Sections 3 and 4 hold in a stronger form than the statements about symmetric functions they imply, in Section 5 we

* This work appears as part of the author's doctoral thesis.

make this precise and thereby set up a system in which we, in the next two sections, study the inner product of Philip Hall and the Kronecker inner product, which arises from the theory of representations of the symmetric group. One of the results obtained can be interpreted to yield an interesting fact about permutation representations of S_n, but this interpretation will not be discussed here.

We would like to thank Dr. Gian-Carlo Rota, whose idea it was to study symmetric functions from the present point of view, for his many helpful suggestions and discussions on the topic.

2. Terminology and assumed results

We assume the reader to have a certain familiarity with symmetric functions, at least to the extent of knowing the definitions of the following basic symmetric functions:

(i) The monomial symmetric functions, denoted k_λ.
(ii) The elementary symmetric functions a_λ.
(iii) The complete homogeneous symmetric functions h_λ.
(iv) The power sum symmetric functions s_λ.

All symmetric functions dealt with have coefficients in Q (the field of rational numbers), involve infinitely many indeterminates $x_1, x_2, \ldots$, and are of a fixed homogeneous degree n. The vector space (over Q) of symmetric functions of degree n is denoted $\mathscr{S}_n$, or simply $\mathscr{S}$. The relevant definitions are to be found in [10] or [19].

Along with the above we assume a knowledge of the definitions and notations used in [19] for partitions of an integer n. A number of these notations are so important that we shall list them here:

(i) $\lambda \vdash n$ denotes that λ is a partition of n.
(ii) $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)$ or $\lambda = (\lambda_1, \lambda_2, \ldots)$ means that the parts of λ are $\lambda_1, \lambda_2, \ldots$ in non-increasing order, and $\lambda_p > 0$.
(iii) $\lambda = (1^{r_1}2^{r_2}\ldots)$, denotes the partition of n with r_1 parts equal to $1, r_2$ parts equal to 2, etc.

We also use the following notations:

(i) $\lambda! \equiv \lambda_1!\lambda_2!\ldots = 1!^{r_1}2!^{r_2}\ldots$ if $\lambda = (\lambda_1, \lambda_2, \ldots) = (1^{r_1}2^{r_2}\ldots)$
(ii) $|\lambda| \equiv r_1!r_2!\ldots$ if $\lambda = (1^{r_1}2^{r_2}\ldots)$
(iii) $\operatorname{sign} \lambda \equiv (-1)^{r_2+2r_3+3r_4+\cdots}$.

We further assume the reader to be familiar with the notions of (finite) partially ordered sets (posets) and lattices, of segments and direct products of these, and of Möbius inversion over posets (see [18]).

In this paper we deal with the lattice $\prod(D)$ of partitions of a finite set D (also denoted $\prod_n$ if D has n elements). A partition π of D is a family of disjoint subsets $B_1, B_2, \ldots, B_b$, called blocks, whose union is D. The set of partitions of D is ordered by putting $\sigma \leq \pi$ if every block of σ is contained in a block of π. It is easily verified that this is a partial ordering relation and that $\prod(D)$ is in fact a lattice. We assume knowledge of the result that if $\sigma \leq \pi$ in $\prod_n$, the segment $[\sigma, \pi]$ is isomorphic to the direct product of r_1 copies of $\prod_1$, r_2 copies of $\prod_2$, etc., where r_i is the number of blocks of π which are composed of i blocks of σ, and that $\mu(\sigma, \pi) = \prod_i (i-1)!^{r_i}$. To the segment $[\sigma, \pi]$ we assign the partition $\lambda(\sigma, \pi) = (1^{r_1}2^{r_2}\cdots)$, ($\lambda(\sigma, \pi) \vdash m$ for some $m \leq n$), called the type of $[\sigma, \pi]$, and to $\pi \in \prod_n$ we assign the partition $\lambda(\pi) = \lambda(0, \pi)$ of n (where 0 is the minimal element in $\prod(D)$, namely the partition whose blocks are the one point subsets of D), called the type of π. If $\lambda(\pi) = \mu$ we sometimes write $\pi \in \mu$. To $[\sigma, \pi]$ we assign the number sign $(\sigma, \pi) = (-1)^{r_2 + 2r_3 + \cdots} = \text{sign}\, \lambda(\sigma, \pi)$ and to π we assign the number sign $\pi = \text{sign}\,(0, \pi)$. It is not difficult to see that sign $(\sigma, \pi) = (\text{sign}\, \sigma)(\text{sign}\, \pi)$ and that $\mu(\sigma, \pi) = \text{sign}\,(\sigma, \pi) \cdot |\mu(\sigma, \pi)|$. For more details and some proofs see [2], [6], or [18].

It is useful to know a few simple facts about the symmetric group S_n, the group of permutations of the set $\{1, 2, \ldots, n\}$. Any permutation can be written as a product of disjoint cycles (see [3], page 133, for definitions), and hence to each permutation σ in S_n we can associate the partition $\lambda = (1^{r_1}2^{r_2}\ldots)$ of n, where r_i is the number of cycles of length i in the cycle decomposition of σ. λ is called the type of σ. The number of elements of S_n of type λ, denoted $[{n \atop \lambda}]$, is easily computed to be

$$\frac{n!}{1^{r_1} r_1! 2^{r_2} r_2! \cdots}$$

3. The functions k_λ, a_λ, and S_λ

Let D be a set with n elements, $X = \{x_1, x_2, \ldots\}$, and let $F = \{f : D \to X\}$. For $f \in F$, its generating function $\gamma(f)$ is $\prod_{d \in D} f(d)$, i.e., $\prod_i x_i^{|f^{-1}(x_i)|}$. For $T \subset F$, the generating function $\gamma(T)$ is $\sum_{f \in T} \gamma(f)$. To any $f \in F$, we assign a partition ker f of D, by putting d_1 and d_2 in the same block of ker f if $f(d_1) = f(d_2)$. ker f is called the kernel of f.

We now define three types of subsets of F. If $\pi \in \prod(D)$, let

$$\mathscr{K}_\pi = \{f \in F | \ker f = \pi\} \tag{1}$$

$$\mathscr{S}_\pi = \{f \in F | \ker f \geq \pi\} \tag{2}$$

$$\mathscr{A}_\pi = \{f \in F | \ker f \wedge \pi = 0\} \tag{3}$$

Let $k_\pi = \gamma(\mathscr{K}_\pi)$, $s_\pi = \gamma(\mathscr{S}_\pi)$, $a_\pi = \gamma(\mathscr{A}_\pi)$. We now compute k_π, s_π, a_π.

THEOREM 1.

(i) $k_\pi = |\lambda(\pi)| k_{\lambda(\pi)}$, (4)

where k_λ is the monomial symmetric function.

(ii) $s_\pi = s_{\lambda(\pi)}$ (5)

(iii) $a_\pi = \lambda(\pi)! a_{\lambda(\pi)}$. (6)

Proof. Let $\lambda(\pi) = (1^{r_1}2^{r_2}) \cdots = (\lambda_1, \lambda_2, \ldots)$.

(i) For $f \in \mathscr{K}_\pi$, i.e. ker $f = \pi$, it is clear that $\gamma(f) = x_{i_1}^{\lambda_1} x_{i_2}^{\lambda_2} \cdots$ for some choice of distinct indices $i_1, i_2, \ldots$.

Further, each such monomial can arise from $r_1! r_2! \cdots$ functions $f \in \mathscr{K}_\pi$.

Therefore $k_\pi = \gamma(\mathscr{K}_\pi) = r_1! r_2! \ldots k_\lambda = |\lambda(\pi)| k_{\lambda(\pi)}$

(ii) $\mathscr{S}_\pi = \{f \in F | \ker f \geq \pi\} = \{f \in F | f \text{ is a constant on blocks of } \pi\}$.

$$\therefore \quad \gamma(\mathcal{T}_\pi) = \sum_{f\in\mathcal{T}_\pi} \gamma(f) = \sum_{f\in\mathcal{T}_\pi} \gamma(f|B_1)\gamma(f|B_2)\cdots$$

$$= \sum_{\substack{f:f|B_i \text{ is} \\ \text{constant } \forall i}} \gamma(f|B_1)\gamma(f|B_2)\cdots$$

$$= \prod_i \left(\sum_{\substack{f:B_i\to X \\ \text{constant}}} \gamma(f) \right)$$

$$= \prod_i (x_1^{|B_i|} + x_2^{|B_i|} + \cdots)$$

$$= s_{\lambda(\pi)}.$$

(iii) $\mathcal{A}_\pi = \{f \in F | \ker f \wedge \pi = 0\} = \{f \in F | f \text{ is 1-1 on the blocks of } \pi\}$

$$\gamma(\mathcal{A}_\pi) = \sum_{f\in\mathcal{A}_\pi} \gamma(f) = \sum_{f\in\mathcal{A}_\pi} \gamma(f|B_1)\gamma(f|B_2)\ldots$$

$$= \sum_{\substack{f\in F: \\ f|B_i \text{ is 1-1 } \forall i}} \gamma(f|B_1)\gamma(f|B_2)\ldots$$

$$= \prod_i \left(\sum_{\substack{f:B_i\to X \\ f\,1-1}} \gamma(f) \right).$$

Now,

$$\sum_{\substack{f:B\to X \\ f\,1-1}} \gamma(f)$$

is the sum of monomials of $|B|$ distinct terms, and each such monomial can arise from $|B|!$ functions $f:B \to X$. Therefore

$$\sum_{\substack{f:B\to X \\ f\,1-1}} \gamma(f) = |B|!a_{|B|}.$$

Therefore $a_\pi = \gamma(\mathcal{A}_\pi) = |B_1|!|B_2|!\cdots a_{|B_1|}a_{|B_2|}\cdots = \lambda(\pi)!a_{\lambda(\pi)}$.

By formulas (1), (2), and (3), it follows that

$$s_\pi = \sum_{\sigma\geq\pi} k_\sigma \tag{7}$$

$$a_\pi = \sum_{\sigma:\sigma\wedge\pi=0} k_\sigma \tag{8}$$

Formula (7) can be inverted by Möbius inversion, as can (8) by the following lemma.

LEMMA. *Let L be a finite lattice on which $\mu(0, x) \neq 0$ for all x in L, where μ is the Möbius function on L. Then*

$$f(x) = \sum_{y:y\wedge x=0} g(y) \leftrightarrow g(x) = \sum_{y\geq x} \frac{\mu(x, y)}{\mu(0, y)} \sum_{z\leq y} \mu(z, y)f(z) \tag{9}$$

Proof: Let

$$f(x) = \sum_{y:y\wedge x=0} g(y)$$

Then

$$f(x) = \sum_{y} \left(\sum_{z \leq x \wedge y} \mu(0, z) \right) g(y)$$

$$= \sum_{z \leq x} \mu(0, z) \sum_{y \geq z} g(y) \tag{10}$$

By Möbius inversion of (10),

$$\mu(0, x) \sum_{y \geq x} g(y) = \sum_{z \leq x} \mu(z, x) f(z) \tag{11}$$

Thus

$$\sum_{y \geq x} g(y) = \frac{1}{\mu(0, x)} \sum_{z \leq x} \mu(z, x) f(z) \tag{12}$$

since $\mu(0, x) \neq 0$, so by Möbius inversion again,

$$g(x) = \sum_{y \geq x} \frac{\mu(x, y)}{\mu(0, y)} \sum_{z \leq y} \mu(z, y) f(z)$$

The converse is proved by reversing the steps.

THEOREM 2.

(i) $k_\pi = \sum_{\sigma \geq \pi} \mu(\pi, \sigma) s_\sigma$ (13)

(ii) $k_\pi = \sum_{\sigma \geq \pi} \frac{\mu(\pi, \sigma)}{\mu(0, \sigma)} \sum_{\tau \leq \sigma} \mu(\tau, \sigma) a_\tau$ (14)

Proof:
(i) follows from (7) by Möbius inversion
(ii) follows from (8) and the preceding lemma.

COROLLARY. *$\{s_\lambda\}$ and $\{a_\lambda\}$ are bases for $\mathscr{S}_n$, the vector space of symmetric functions of homogeneous degree n.*

Proof: $\{k_\lambda\}$ is a basis for $\mathscr{S}_n$, and Theorem 2 shows that $\{s_\lambda\}$ and $\{a_\lambda\}$ generate $\mathscr{S}_n$. Since all three sets have the same number of elements (namely, the number of partitions of n), the result follows.

We now determine the relationship between the elementary and the power sum symmetric functions.

THEOREM 3.

(i) $a_\pi = \sum_{\sigma \leq \pi} \mu(0, \sigma) s_\sigma$ (15)

(ii) $s_\pi = \frac{1}{\mu(0, \pi)} \sum_{\sigma \leq \pi} \mu(\sigma, \pi) a_\sigma$ (16)

(iii) $1 + \sum_{n \geq 1} a_n t^n = \exp\left(\sum_{n \geq 1} \frac{(-1)^{n-1}}{n} s_n t^n \right)$ (17)

Proof:

(i) $$a_\pi = \sum_{\sigma:\sigma \wedge \pi = 0} k_\sigma$$
$$= \sum_\sigma \left(\sum_{\tau \leq \sigma \wedge \pi} \mu(0, \tau) \right) k_\sigma$$
$$= \sum_{\tau \leq \pi} \mu(0, \tau) \sum_{\sigma \geq \tau} k_\sigma$$
$$= \sum_{\tau \leq \pi} \mu(0, \tau) s_\tau$$

(ii) Apply Möbius inversion to (15)

(iii) Follows directly from (i) and the corollary to Theorem 4 in [6], since $\mu(0, \sigma)s_\sigma$ is a multiplicative function of σ (in the sense of [6]).

Waring's formula follows from Theorem 3 if we put $\pi = 1$ (the maximum element in $\prod_n$) in (16) using the fact (proved in [6]) that the number of $\pi \in \prod_n$ of type $\lambda = (1^{r_1}2^{r_2}\cdots)$, which we will denote $\binom{n}{\lambda}$, is equal to

$$\frac{n!}{1!^{r_1}r_1!2!^{r_2}r_2!\ldots} = \frac{n!}{\lambda!|\lambda|}$$

If we generalize the formula for a_π in (15) and consider the functions $k_{[\tau,\pi]} = \sum_{\sigma\in[\tau,\pi]} \mu(\tau, \sigma)s_\sigma$ it is not difficult to verify that $k_{[\tau,\pi]} = \sum_{\sigma:\sigma \wedge \pi = \tau} k_\sigma$ and that $k_{[\tau,\pi]}$ is the product $\prod_B k_{\sigma_B}$, where the product is over the blocks B of π and where σ_B is the partition of B whose blocks are those blocks of σ contained in B. We can use this to express the product of monomial symmetric functions as a linear combination of any of the symmetric functions discussed here.

We now prove a well-known result concerning the elementary symmetric functions.

THEOREM 4. *Let* $a_\lambda = \sum_\mu c_{\lambda\mu}k_\mu$. *The* $c_{\lambda\mu} = c_{\mu\lambda}$.

Proof: By (8) and Theorem 1, the coefficient of $|\mu|k_\mu$ in $\lambda!a_\lambda$ is $\sum_{\sigma\in\mu} \delta(o, \sigma \wedge \pi)$ where π is some fixed partition of type λ. Thus the coefficient of k_μ in a_λ is

$$\frac{|\mu|}{\lambda!} \sum_{\sigma\in\mu} \delta(0, \sigma \wedge \pi) = \frac{|\mu|}{\lambda!} \frac{1}{\binom{n}{\lambda}} \sum_{\sigma\in\mu,\pi\in\lambda} \delta(0, \sigma \wedge \pi)$$
$$= \frac{|\lambda||\mu|}{n!} \sum_{\sigma\in\mu,\pi\in\lambda} \delta(0, \sigma \wedge \pi)$$

which is symmetric in λ and μ.

4. The functions h_λ and f_λ

We now generalize the notions introduced at the beginning of Section 3 in a way that will allow us to obtain the complete homogeneous symmetric functions h_n. Consider the domain set D to be a set of "balls", and $x_1, x_2, \ldots$ to be "boxes". By a placing p we mean an arrangement of the balls in the boxes in which the balls in each box may be placed in some configuration. The kernel ker p is the partition

$\pi \in \prod(D)$ such that d_1 and d_2 are in the same block of π if they are in the same box. The generating function $\gamma(p)$ of a placing p is the monomial $\prod_i x_i$ (number of balls in x_i), or equivalently $\gamma(p) = \prod_{d\in D}$(box in which d lies). The generating function $\gamma(P)$ of a set P of placings is $\sum_{p\in P}\gamma(p)$.

The notion of a placing is a generalization of a function, since to a function $f:D \to X$ we can associate the placing $\bar{f}$ in which ball d is in box $f(d)$, and the balls in each box are in no special configuration. (Actually placings are similar to reluctant functions, a generalization of functions defined in [15]). It is important to note that $\ker f = \ker \bar{f}$ and $\gamma(f) = \gamma(\bar{f})$. If p is a placing, by the "underlying function" of p we mean the mapping $f:D \to X$ given by: $f(d) = x_i$ if ball d is in box x_i in the placing p.

Using this terminology, the definitions at the beginning of Section 3 can be restated as follows:

$\mathscr{K}_\pi$ = {placings p with no configuration and kernel π}
= {ways of placing the blocks of π into distinct boxes}

$\mathscr{A}_\pi$ = {placings p with no configuration such that no two balls from the same block of π go into the same box}

$\mathscr{T}_\pi$ = {placings p with no configuration such that balls from the same block of π go into the same box}.

We now define two new families:

$\mathscr{H}_\pi$ = {placings p such that within each box the balls from the same block of π are linearly ordered.}.

$\mathscr{F}_\pi$ = {ways of placing the blocks of π into the boxes and within each box linearly ordering the blocks appearing}.

Put $h_\pi = \gamma(\mathscr{H}_\pi)$, $f_\pi = \gamma(\mathscr{F}_\pi)$. We now determine h_π.

THEOREM 5.

$$h_\pi = \lambda(\pi)!h_{\lambda(\pi)} \tag{18}$$

Proof: Since a placing $p \in \mathscr{H}_\pi$ can be obtained by first placing the balls from B_1 into the boxes and linearly ordering within each box, then placing the balls from B_2 and linearly ordering again within each box (independently of how the balls from B_1 are ordered), etc. (where $B_1, B_2, \ldots$ are the blocks of π in some order), it follows that h_π is the product $\prod_{\text{blocks } B}\gamma(\mathscr{H}_B)$, where $\mathscr{H}_B$ is the set of placings of the balls from B into the boxes and linearly ordering them within each box.

In $\gamma(\mathscr{H}_B)$, each monomial $x_{i_1}^{u_1}x_{i_2}^{u_2}\cdots$ of degree $|B|$ arises $|B|!$ times, since there is a one-one correspondence between $\{p \in \mathscr{H}_B | \gamma(p) = x_{i_1}^{u_1}x_{i_2}^{u_2}\cdots\}$ and linear orderings of B, namely to each such placing associate the linear ordering obtained by taking first the balls in x_{i_1} in the given order, then those in x_{i_2}, and so on.

Hence

$$\gamma(\mathscr{H}_B) = |B|!h_{|B|}, \qquad \text{so } \gamma(\mathscr{H}_\pi) = |B_1|!|B_2|!\cdots h_{|B_1|}h_{|B_2|}\cdots$$

i.e.,

$$h_\pi = \lambda(\pi)!h_{\lambda(\pi)}.$$

The relationship between the h's and the k's is given by the following.

THEOREM 6.

$$h_\pi = \sum_\sigma \lambda(\sigma \wedge \pi)!k_\sigma. \tag{19}$$

Proof: For each function $f:D \to X$ with kernel σ, the number of $p \in \mathscr{H}_\pi$ with underlying function f is the number of ways of putting balls D in boxes X as prescribed by f and then linearly ordering within each box the elements from the same block of π, i.e., the number of ways of independently linearly ordering the elements within each block of $\sigma \wedge \pi$. This number is just $\lambda(\sigma \wedge \pi)!$. Since this depends only on σ (for fixed π) and not on f, we have

$$h_\pi = \sum_\sigma \lambda(\sigma \wedge \pi)!k_\sigma.$$

COROLLARY. *If* $h_\lambda = \sum_\mu d_{\lambda\mu}k_\mu$, *then* $d_{\lambda\mu} = d_{\mu\lambda}$.

Proof: same as the proof of Theorem 4, replacing the function $\delta(0, \sigma \wedge \pi)$ with $\lambda(\sigma \wedge \pi)!$, both of which are symmetric in σ and π.

Formula (19) can be better dealt with using the following result.

LEMMA.

$$\sum_{\sigma\in[\tau,\pi]} |\mu(\tau, \sigma)| = \lambda(\tau, \pi)! \tag{20}$$

Proof. Let

$$f(\tau, \pi) = \sum_{\sigma\in[\tau,\pi]} |\mu(\tau, \sigma)|.$$

It is clear that $f(\tau, \pi) = \prod_i (\sum_{\sigma\in\Pi(B_i)}|\mu(0, \sigma)|)$ where $B_1, B_2, \ldots$ are the relative blocks of $[\tau, \pi]$. Hence it suffices to show that $\sum_{\sigma\in\Pi_n}|\mu(0, \sigma)| = n!$ But

$$\begin{aligned}
\sum_{\sigma\in\pi} |\mu(0, \sigma)| &= \sum_{\lambda\vdash n} \binom{n}{\lambda} 0!^{r_1}1!^{r_2}\cdots(i-1)!^{r_i}\cdots \\
&= \sum_{\lambda\vdash n} \frac{n!}{1^{r_1}r_1!2^{r_2}r_2!\cdots} \\
&= \sum_{\lambda\vdash n} (\#\text{ of permutations } \sigma \text{ in the symmetric group } S_n \text{ of type } \lambda) \\
&= |S_n| \quad (\text{where } |A| = \text{number of elements in } A) \\
&= n!
\end{aligned}$$

Note: The lemma also follows easily from the results on the lattice of partitions in [6].

We now obtain the relationship between the h's and the k's, a's, and s's.

THEOREM 7.

$$\text{(i) } k_\pi = \sum_{\sigma\geq\pi} \frac{\mu(\pi, \sigma)}{|\mu(0, \sigma)|} \sum_{\tau\leq\sigma} \mu(\tau, \sigma)h_\tau \tag{21}$$

$$\text{(ii) } h_\pi = \sum_{\sigma\leq\pi} |\mu(0, \sigma)|s_\sigma \tag{22}$$

(iii) $s_\pi = \frac{1}{|\mu(0,\pi)|} \sum_{\sigma \le \pi} \mu(\sigma, \pi) h_\sigma$ (23)

(iv) $a_\pi = \sum_{\sigma \le \pi} (\text{sign } \sigma) \lambda(\sigma, \pi)! h_\sigma$ (24)

(v) $h_\pi = \sum_{\sigma \le \pi} (\text{sign } \sigma) \lambda(\sigma, \pi)! a_\sigma$ (25)

(vi) $1 + \sum_{n \ge 1} h_n t^n = \exp\left(\sum_{n \ge 1} \frac{1}{n} s_n t^n \right).$ (26)

Proof:

(i)
$$h_\pi = \sum_\sigma \lambda(\sigma \wedge \pi)! k_\sigma = \sum_\sigma \left(\sum_{\tau \le \sigma \wedge \pi} |\mu(0, \tau)| \right) k_\sigma = \sum_{\tau \le \pi} |\mu(0, \tau)| \sum_{\sigma \ge \tau} k_\sigma$$

Invert twice to obtain (21).

(ii) We showed in (i) that $h_\pi = \sum_{\tau \le \pi} |\mu(0, \tau)| \sum_{\sigma \ge \tau} k_\sigma$, and since $\sum_{\sigma \ge \tau} k_\sigma = s_\tau$, the result follows.

(iii) Apply Möbius inversion to (ii).

(iv)
$$\begin{aligned} a_\pi &= \sum_{\sigma \le \pi} \mu(0, \sigma) s_\sigma \\ &= \sum_{\sigma \le \pi} \mu(0, \sigma) \cdot \frac{1}{|\mu(0, \sigma)|} \sum_{\tau \le \sigma} \mu(\tau, \sigma) h_\tau \\ &= \sum_{\sigma \le \pi} (\text{sign } \sigma) \sum_{\tau \le \sigma} \mu(\tau, \sigma) h_\tau \\ &= \sum_{\tau \le \pi} \left(\sum_{\sigma \in [\tau, \pi]} (\text{sign } \sigma) \mu(\tau, \sigma) \right) h_\tau \\ &= \sum_{\tau \le \pi} (\text{sign } \tau) \left(\sum_{\sigma \in [\tau, \pi]} |\mu(\tau, \sigma)| \right) h_\tau \\ &= \sum_{\tau \le \pi} (\text{sign } \tau) \lambda(\tau, \pi)! h_\tau. \end{aligned}$$

(v) Proved in the same way as (iv).

(vi) Follows from (ii) and the corollary to theorem 4 in [6].

COROLLARY 1. *$\{h_\lambda\}$ is a basis for $\mathscr{S}_n$.*

COROLLARY 2. *The mapping $\theta: \mathscr{S}_n \to \mathscr{S}_n$ given by $\theta(a_\lambda) = h_\lambda$ satisfies*

(i) *θ is an involution, i.e., $\theta^2 = I$.*

(ii) *The s_λ's are the eigenvectors of θ, with $\theta(s_\lambda) = (\text{sign } \lambda) \cdot s_\lambda$.*

Proof:

(i) This is equivalent to showing that if $a_\lambda = \sum_\mu c_{\lambda\mu} h_\mu$ then $h_\lambda = \sum_\mu c_{\lambda\mu} a_\mu$. But this follows from (iv) and (v) of Theorem 7.

(ii) Let $\pi \in \lambda$. Then

$$\begin{aligned}\theta(s_\lambda) = \theta(s_\pi) &= \theta\left(\frac{1}{\mu(0,\pi)}\sum_{\sigma\le\pi}\mu(\sigma,\pi)a_\sigma\right)\\ &= \frac{1}{\mu(0,\pi)}\sum_{\sigma\le\pi}\mu(\sigma,\pi)h_\sigma\\ &= \operatorname{sign}\pi\frac{1}{|\mu(0,\pi)|}\sum_{\sigma\le\pi}\mu(\sigma,\pi)h_\sigma\\ &= (\operatorname{sign}\pi)s_\pi\\ &= (\operatorname{sign}\lambda)s_\lambda.\end{aligned}$$

It is now time to determine what $f_\pi = \sigma(\mathscr{F}_\pi)$ is. If $p \in \mathscr{F}_\pi$, then clearly $\ker p \ge \pi$. For $\tau \ge \pi$ and $f: D \to X$ with kernel τ, the number of $p \in \mathscr{F}_\pi$ with underlying function f is the number of ways of placing the blocks of π in the boxes as prescribed by f (which makes sense since $\ker f \ge \pi$) and then linearly ordering the blocks in each box. But the number of blocks of π in the various boxes is $\lambda_1, \lambda_2, \ldots, \lambda_p$, where $\lambda(\pi,\tau) = (\lambda_1, \lambda_2, \ldots, \lambda_p)$. Hence the number of $p \in \mathscr{F}_\pi$ with underlying function f is $\lambda_1!\lambda_2!\ldots\lambda_p! = \lambda(\pi,\tau)!$. Thus

$$f_\pi = \sum_{\tau\ge\pi}\lambda(\pi,\tau)!k_\tau. \tag{27}$$

Hence,

$$\begin{aligned}f_\pi &= \sum_{\tau\ge\pi}\lambda(\pi,\tau)!k_\tau\\ &= \sum_{\tau\ge\pi}\left(\sum_{\sigma\in[\pi,\tau]}|\mu(\pi,\sigma)|\right)k_\tau\\ &= \sum_{\sigma\ge\pi}|\mu(\pi,\sigma)|\sum_{\tau\ge\sigma}k_\tau\\ &= \sum_{\sigma\ge\pi}|\mu(\pi,\sigma)|s_\sigma.\end{aligned} \tag{28}$$

Thus

$$\begin{aligned}\theta(k_\pi) &= \theta\left(\sum_{\sigma\ge\pi}\mu(\pi,\sigma)s_\sigma\right)\\ &= \sum_{\sigma\ge\pi}\mu(\pi,\sigma)(\operatorname{sign}\sigma)s_\sigma\\ &= (\operatorname{sign}\pi)\sum_{\sigma\ge\pi}|\mu(\pi,\sigma)|s_\sigma\\ &= (\operatorname{sign}\pi)f_\pi.\end{aligned}$$

Thus we have proved that the f_π's are the images under θ of the k's. Defining $f_\lambda = 1/|\lambda| f_\pi$ (where $\pi \in \lambda$), which makes sense since $\gamma(f_\pi)$ depends only on the type of π, we have proved the following theorem.

THEOREM 8.

(i) $\theta(k_\lambda) = (\text{sign } \lambda) f_\lambda$

(ii) $f_\pi = \sum_{\tau \geq \pi} \lambda(\pi, \tau)! k_\tau$

(iii) $f_\pi = \sum_{\sigma \geq \pi} |\mu(\pi, \sigma)| s_\sigma$.

COROLLARY. *In the image of k_λ under θ, all monomials appearing have coefficients of the same sign, namely the sign of λ.*

Most of the relationships among the k_π's, s_π's, a_π's, h_π's, and f_π's have been stated by this point. All others (for example the h's in terms of the f's) can easily be obtained from these, and are listed in Appendix 1.

5. The vector space $\tilde{\mathscr{S}}$

It is interesting to note that for most of the results proved so far, no use is made of the fact that k_π, a_π, etc. are really symmetric functions and hence that $k_\pi = k_\sigma$ for π, σ of the same type. For example, not only is the matrix $(c_{\lambda\mu})$ given by $a_\lambda = \sum_\mu c_{\lambda\mu} k_\mu$ symmetric, but so is the matrix $c_{\pi\sigma}$ given by $a_\pi = \sum_\sigma c_{\pi\sigma} k_\sigma$, namely $c_{\pi\sigma} = \delta(0, \pi \wedge \sigma)$. Another example is the fact that $\theta(s_\pi) = (\text{sign } \pi) s_\pi$, which implies that $\theta(s_\lambda) = (\text{sign } \lambda) s_\lambda$, but which is a stronger result (for example, if we showed that $\theta(s_\pi) = \frac{1}{2}((\text{sign } \sigma) s_\sigma + (\text{sign } \tau) s_\tau)$ with π, σ, and τ all of type λ, it would follow that $\theta(s_\lambda) = (\text{sign } \lambda) s_\lambda$).

These considerations lead us to define a vector space $\tilde{\mathscr{S}}$, the vector space over Q freely generated by the symbols $\{\tilde{k}_\pi | \pi \in \prod_n\}$. We then define elements $\tilde{a}_\pi$, $\tilde{s}_\pi$, $\tilde{h}_\pi$, $\tilde{f}_\pi$, in $\tilde{\mathscr{S}}$ by

$$\tilde{a}_\pi = \sum_{\sigma : \sigma \wedge \pi = 0} \tilde{k}_\sigma \tag{29}$$

$$\tilde{s}_\pi = \sum_{\sigma \geq \pi} \tilde{k}_\sigma \tag{30}$$

$$\tilde{h}_\pi = \sum_{\sigma} \lambda(\sigma \wedge \pi)! \tilde{k}_\sigma \tag{31}$$

$$\tilde{f}_\pi = \sum_{\sigma \geq \pi} \lambda(\pi, \sigma)! \tilde{k}_\sigma. \tag{32}$$

Since all other formulas obtained in the previous two sections and in Appendix 1 can be obtained from these by Möbius inversion and similar techniques, they all hold in $\tilde{\mathscr{S}}$. Hence $\{\tilde{a}_\pi\}$, $\{\tilde{s}_\pi\}$, $\{\tilde{h}_\pi\}$, and $\{\tilde{f}_\pi\}$ are bases of $\tilde{\mathscr{S}}$. Also, if we define the mapping $\tilde{\theta} : \tilde{\mathscr{S}} \to \tilde{\mathscr{S}}$ by $\tilde{\theta}(\tilde{a}_\pi) = \tilde{h}_\pi$, using formulas (24), (25), and the same method as in the proof of part (ii) of Corollary 2 to Theorem 7, it follows that $\tilde{\theta}$ is an involution with eigenvectors $\tilde{s}_\pi$.

Summarizing the results obtained so far, we have

THEOREM 9.

(i) *$\{\tilde{a}_\pi\}$, $\{\tilde{h}_\pi\}$, $\{\tilde{s}_\pi\}$, and $\{\tilde{f}_\pi\}$ are bases of $\tilde{\mathscr{S}}$.*

(ii) *The matrices $(c_{\pi\sigma})$ and $(d_{\pi\sigma})$ given by $\tilde{a}_\pi = \sum_\sigma c_{\pi\sigma} \tilde{k}_\sigma$ and $\tilde{h}_\pi = \sum_\sigma d_{\pi\sigma} \tilde{k}_\sigma$ are symmetric.*

(iii) *$\tilde{\theta}$ is an involution*

(iv) $\tilde{\theta}(\tilde{s}_\pi) = (\text{sign } \pi) \tilde{s}_\pi$

(v) $\tilde{\theta}(\tilde{k}_\pi) = (\text{sign } \pi) \tilde{f}_\pi$.

Now we define a mapping $\phi:\tilde{\mathscr{S}} \to \mathscr{S}$ by $\phi(\tilde{k}_\pi) = k_\pi(= |\lambda(\pi)| k_{\lambda(\pi)})$. By the results of the previous two sections, we have

THEOREM 10. $\phi(\tilde{a}_\pi) = a_\pi$, $\phi(\tilde{s}_\pi) = s_\pi$, $\phi(\tilde{h}_\pi) = h_\pi$, $\phi(\tilde{f}_\pi) = f_\pi$.

The interesting fact of this section, and one that will be useful in the next two sections is that there is a map $\phi^*:\mathscr{S} \to \tilde{\mathscr{S}}$ such that $\phi\phi^* = I$ = identity map on $\mathscr{S}$. To see this, we define elements $\tilde{K}_\lambda, \tilde{A}_\lambda, \tilde{S}_\lambda, \tilde{H}_\lambda$, and $\tilde{F}_\lambda$ in $\tilde{\mathscr{S}}$ by $\tilde{K}_\lambda = \sum_{\pi\in\lambda} \tilde{k}_\pi$, $\tilde{A}_\lambda = \sum_{\pi\in\lambda} \tilde{a}_\pi$, etc., and elements $K_\lambda, A_\lambda, S_\lambda, H_\lambda$, and F_λ in $K_\lambda = \sum_{\pi\in\lambda} k_\pi = \binom{n}{\lambda}|\lambda| k_\lambda$ (where we recall that $\binom{n}{\lambda} = \frac{n!}{\lambda!|\lambda|}$ is the number of $\pi \in \prod_n$ of type λ), $A_\lambda = \sum_{\pi\in\lambda} a_\pi = \binom{n}{\lambda}\lambda! a_\lambda$, etc. Let $\phi^*:\mathscr{S} \to \tilde{\mathscr{S}}$ be defined by $\phi^*(K_\lambda) = \tilde{K}_\lambda$, and let "$\mathscr{S}$" be the image under ϕ^* of $\mathscr{S}$ (i.e., "$\mathscr{S}$" is the subspace of $\tilde{\mathscr{S}}$ spanned by $\{\tilde{K}_\lambda\}$). We then have the following theorem.

THEOREM 11.
(i) *$\tilde{A}_\lambda, \tilde{S}_\lambda, \tilde{H}_\lambda, \tilde{F}_\lambda$ are all in "$\mathscr{S}$".*
(ii) *$\phi^*(A_\lambda) = \tilde{A}_\lambda, \phi^*(S_\lambda) = \tilde{S}_\lambda$, etc.*
(iii) *$\phi\phi^* = I$ = identity map on $\mathscr{S}$.*

Proof:

(i) $$\tilde{A}_\lambda = \sum_{\pi\in\lambda} \tilde{a}_\pi = \sum_{\pi\in\lambda}\left(\sum_{\sigma:\sigma\wedge\pi=0} \tilde{k}_\sigma\right)$$
$$= \sum_\sigma (\#\text{ of } \pi\in\lambda \text{ s.t. } \sigma \wedge \pi = 0)\tilde{k}_\sigma.$$

But clearly ($\#$ of $\pi \in \lambda$ s.t. $\sigma \wedge \pi = 0$) depends only on the type μ of σ (for fixed λ), so call this number $c_{\lambda\mu}$.
Then $\tilde{A}_\lambda = \sum_\mu c_{\lambda\mu}(\sum_{\sigma\in\mu} \tilde{k}_\sigma) = \sum_\mu c_{\lambda\mu}\tilde{K}_\mu$.
Therefore $\tilde{A}_\lambda \in$ "$\mathscr{S}$"
Similarly for $\tilde{S}_\lambda, \tilde{H}_\lambda$, and $\tilde{F}_\lambda$.

(ii) $\tilde{A}_\lambda = \sum_\mu c_{\lambda\mu}\tilde{K}_\mu$ with $c_{\lambda\mu}$ as above.
Therefore $\phi(\tilde{A}_\lambda) = \sum_\mu c_{\lambda\mu}\phi(\tilde{K}_\mu) = \sum_\mu c_{\lambda\mu}K_\mu$.
Therefore $A_\lambda = \sum_\mu c_{\lambda\mu}K_\mu$.
Applying ϕ^*, $\phi^*(A_\lambda) = \sum_\mu c_{\lambda\mu}\phi^*(K_\mu) = \sum_\mu c_{\lambda\mu}\tilde{K}_\mu = \tilde{A}_\lambda$.
Similarly for $\tilde{S}_\lambda$, $\tilde{H}_\lambda$, and $\tilde{F}_\lambda$.

(iii) $\phi\phi^*(K_\lambda) = \phi(\tilde{K}_\lambda) = K_\lambda$, so the result holds since $\{K_\lambda\}$ is a basis for $\mathscr{S}$.

6. Hall's inner product

In this section we define an inner product on $\tilde{\mathscr{S}}$ in such a way that $\phi^*:\mathscr{S} \to \tilde{\mathscr{S}}$ is an isometry with respect to the inner product on $\mathscr{S}$ defined by Philip Hall (see [10]). Again in this section, many familiar results holding for Hall's inner product hold in their stronger form for $\tilde{\mathscr{S}}$, and also certain computations of inner products in $\mathscr{S}$ can be facilitated by translating them, via ϕ^*, to computations in $\tilde{\mathscr{S}}$ which are usually simpler.

Hall's inner product on $\mathscr{S}$ is defined by taking $(h_\lambda, k_\mu) = \delta_{\lambda\mu}$ (and extending by bilinearity). We define an inner product on $\tilde{\mathscr{S}}$ by putting $(\tilde{h}_\pi, \tilde{k}_\sigma) = n!\delta_{\pi\sigma}$. The notation (,) for both inner products is the same, but this should cause no confusion.

A number of important results about this inner product are given by the following theorem.

THEOREM 12.

(i) *This inner product is symmetric, i.e.,* $(\tilde{f}, \tilde{g}) = (\tilde{g}, \tilde{f})$ *for* $\tilde{f}, \tilde{g} \in \tilde{\mathscr{S}}$.

(ii) $\phi^* : \mathscr{S} \to \tilde{\mathscr{S}}$ is an isometry, i.e., $(\phi^*(f), \phi^*(g)) = (f, g)$ *for* $f, g \in \mathscr{S}$.

(iii) $(\tilde{s}_\pi, \tilde{s}_\sigma) = \delta_{\pi\sigma} \cdot n!/|\mu(0, \pi)|$.

(iv) $\tilde{\theta} : \tilde{\mathscr{S}} \to \tilde{\mathscr{S}}$ *is an isometry.*

Proof:

(i) Follows from Theorem 9 (ii), together with the following easily proved fact from linear algebra: If V is any vector space, $\{v_i\}$ and $\{w_i\}$ two bases such that $v_i = \sum_j c_{ij} w_j$, then the inner $(\,,)$ on V defined by $(v_i, w_j) = \delta_{ij}$ is symmetric if and only if the matrix (c_{ij}) is symmetric.

(ii) It suffices to show that $(\phi^*(b_\lambda), \phi^*(b'_\mu)) = (b_\lambda, b'_\mu)$ for some pair of bases $\{b_\lambda\}, \{b'_\lambda\}$ of $\mathscr{S}$. Now,

$$\begin{aligned}(\phi^*(H_\lambda), \phi^*(K_\mu)) = (\tilde{H}_\lambda, \tilde{K}_\mu) &= \left(\sum_{\pi \in \lambda} \tilde{h}_\pi, \sum_{\sigma \in \mu} \tilde{k}_\sigma\right) \\ &= \sum_{\substack{\pi \in \lambda \\ \sigma \in \mu}} \delta_{\pi\sigma} n! \\ &= n! \delta_{\lambda\mu} \cdot (\# \text{ of } \pi \in \lambda) \\ &= n! \binom{n}{\lambda} \delta_{\lambda\mu}.\end{aligned}$$

And

$$\begin{aligned}(H_\lambda, K_\mu) &= \left(\binom{n}{\lambda} \lambda! h_\lambda, \binom{n}{\mu} |\mu| k_\mu\right) \\ &= \delta_{\lambda\mu} \binom{n}{\lambda} \cdot \binom{n}{\mu} \lambda! |\mu| \\ &= \delta_{\lambda\mu} \binom{n}{\lambda}^2 \lambda! |\lambda| \\ &= \delta_{\lambda\mu} \cdot \binom{n}{\lambda} \cdot \frac{n!}{\lambda! |\lambda|} \cdot \lambda! |\lambda| \\ &= \delta_{\lambda\mu} \binom{n}{\lambda} \cdot n!\end{aligned}$$

Therefore $(\phi^*(H_\lambda), \phi^*(K_\mu)) = (H_\lambda, K_\mu)$.

(iii)
$$\begin{aligned}(\tilde{s}_\pi, \tilde{s}_\sigma) &= \left(\frac{1}{|\mu(0, \pi)|} \sum_{\tau \le \pi} \mu(\tau, \pi) \tilde{h}_\tau, \sum_{\nu \ge \sigma} \tilde{k}_\nu\right) \\ &= \frac{1}{|\mu(0, \pi)|} \sum_{\substack{\tau \le \pi \\ \nu \ge \sigma}} \mu(\tau, \pi)(\tilde{h}_\tau, \tilde{k}_\nu)\end{aligned}$$

$$= \frac{n!}{|\mu(0, \pi)|} \sum_{\tau \in [\sigma, \pi]} \mu(\tau, \pi)$$

$$= \delta_{\sigma\pi} \cdot \frac{n!}{|\mu(0, \pi)|}.$$

(iv) $(\tilde{\theta}(\tilde{s}_\pi), \tilde{\theta}(\tilde{s}_\sigma)) = ((\text{sign } \pi)\tilde{s}_\pi, (\text{sign } \sigma)\tilde{s}_\sigma)$

$$= (\text{sign } \pi)(\text{sign } \sigma)(\tilde{s}_\pi, \tilde{s}_\sigma)$$

$$= (\tilde{s}_\pi, \tilde{s}_\sigma) \quad (\text{since } (\tilde{s}_\pi, \tilde{s}_\sigma) = 0 \text{ if } \pi \neq \sigma).$$

COROLLARY:
(i) *Hall's inner product on $\mathscr{S}$ is symmetric*
(ii) $(s_\lambda, s_\mu) = \delta_{\lambda\mu} 1^{r_1} r_1! 2^{r_2} r_2! \ldots$, where $\lambda = (1^{r_1} 2^{r_2} \cdots)$.
(iii) $\theta: \mathscr{S} \to \mathscr{S}$ is an isometry.
Proof:
(i) $(f, g) = (\phi^*(f), \phi^*(g)) = (\phi^*(g), \phi^*(f)) = (g, f)$
(ii) $(s_\lambda, s_\mu) = \dfrac{1}{\binom{n}{\lambda}\binom{n}{\mu}}(S_\lambda, S_\mu) = \dfrac{1}{\binom{n}{\lambda}\binom{n}{\mu}}(\tilde{S}_\lambda, \tilde{S}_\mu)$

$$= \frac{1}{\binom{n}{\lambda}\binom{n}{\mu}}\left(\sum_{\pi \in \lambda} \tilde{s}_\pi, \sum_{\sigma \in \mu} \tilde{s}_\sigma\right)$$

$$= \delta_{\lambda\mu} \frac{1}{\binom{n}{\lambda}^2} \sum_{\pi \in \lambda} (\tilde{s}_\pi, \tilde{s}_\pi)$$

$$= \delta_{\lambda\mu} \frac{1}{\binom{n}{\lambda}^2} \cdot \binom{n}{\lambda} \cdot \frac{n!}{|\mu(0, \pi)|} \text{ (where } \pi \in \lambda)$$

$$= \delta_{\lambda\mu} \cdot \frac{n!}{\binom{n}{\lambda} |\mu(0, \pi)|}$$

$$= \delta_{\lambda\mu} \cdot 1^{r_1} r_1! 2^{r_2} r_2! \ldots$$

(iii) $(\theta(s_\lambda), \theta(s_\mu)) = ((\text{sign } \lambda)s_\lambda, (\text{sign } \mu)s_\mu)$

$$= (\text{sign } \lambda)(\text{sign } \mu)(s_\lambda, s_\mu)$$

$$= (s_\lambda, s_\mu) \quad (\text{since } (s_\lambda, s_\mu) = 0 \text{ if } \lambda \neq \mu)$$

All inner products involving the members of the bases studied so far are easily calculated, and are listed in Appendix 2.

7. The Kronecker inner product

The Kronecker inner product on $\mathscr{S}$ is a mapping from $\mathscr{S} \times \mathscr{S} \to \mathscr{S}$ which reflects the Kronecker inner product of representations of S_n (the symmetric group of degree n) in the same way as Hall's inner product mirrors the inner product of representations. Since we are assuming no knowledge of group theory, we can define the Kronecker inner product, which we denote [,], as follows:

$$\left[\frac{1}{n!}\sum_{\lambda\vdash n}\begin{bmatrix}n\\ \lambda\end{bmatrix}c_\lambda s_\lambda, \frac{1}{n!}\sum_{\lambda\vdash n}\begin{bmatrix}n\\ \lambda\end{bmatrix}d_\lambda s_\lambda\right] = \frac{1}{n!}\sum_{\lambda\vdash n}\begin{bmatrix}n\\ \lambda\end{bmatrix}(c_\lambda d_\lambda)s_\lambda \tag{33}$$

where $c_\lambda, d_\lambda \in Q$ and

$$\begin{bmatrix}n\\ \lambda\end{bmatrix} = \frac{n!}{1^{r_1}r_1!\,2^{r_2}r_2!\cdots} = \binom{n}{\lambda}\cdot|\mu(0,\pi)| \qquad (\text{for } \pi\in\lambda).$$

This mapping is bilinear, and it is easily verified that

$$[s_\lambda, s_\mu] = \delta_{\lambda\mu}\cdot\frac{n!}{\begin{bmatrix}n\\ \lambda\end{bmatrix}}s_\lambda = (s_\lambda, s_\mu)s_\lambda. \tag{34}$$

In $\tilde{\mathscr{S}}$, we define $[\,,\,]: \tilde{\mathscr{S}} \times \tilde{\mathscr{S}} \to \tilde{\mathscr{S}}$ to be the bilinear map with

$$[\tilde{s}_\pi, \tilde{s}_\sigma] = (\tilde{s}_\pi, \tilde{s}_\sigma)\cdot\tilde{s}_\pi = \delta_{\pi\sigma}\cdot\frac{n!}{|\mu(0,\pi)|}\tilde{s}_\pi.$$

Parallelling Theorem 11, we have

THEOREM 13:
(i) $[\,,\,]$ *is symmetric on* $\tilde{\mathscr{S}}$, *i.e.,* $\tilde{\mathscr{S}}$ *with product given by* $[\,,\,]$ *is a commutative algebra.*
(ii) $\phi^*[f,g] = [\phi^*(f), \phi^*(g)]$ *for* $f, g\in\mathscr{S}$, *i.e.,* $\phi^*: \mathscr{S} \to \tilde{\mathscr{S}}$ *is an algebra homomorphism.*
(iii) $[f,g] = \phi[\phi^*(f), \phi^*(g)]$ *for* $f, g\in\mathscr{S}$.
(iv) $\tilde{\theta}: \tilde{\mathscr{S}} \to \tilde{\mathscr{S}}$ *satisfies* $[\tilde{\theta}(\tilde{f}), \tilde{\theta}(\tilde{g})] = [\tilde{f}, \tilde{g}]$ *for* $\tilde{f}, \tilde{g}\in\tilde{\mathscr{S}}$, *i.e.,* $\tilde{\theta}$ *is an algebra homomorphism.*

Proof:

$$\begin{aligned}\text{(i)}\ [\tilde{s}_\pi, \tilde{s}_\sigma] &= (\tilde{s}_\pi, \tilde{s}_\sigma)\tilde{s}_\pi\\ &= (\tilde{s}_\pi, \tilde{s}_\sigma)\tilde{s}_\sigma \quad (\text{since } (\tilde{s}_\pi, \tilde{s}_\sigma) = 0 \text{ if } \pi\neq\sigma)\\ &= (\tilde{s}_\sigma, \tilde{s}_\pi)\tilde{s}_\sigma\\ &= [\tilde{s}_\sigma, \tilde{s}_\pi]\end{aligned}$$

The result follows by bilinearity.
(ii) It suffices, by bilinearity, to show that

$$\phi^*[S_\lambda, S_\mu] = [\phi^*(S_\lambda), \phi^*(S_\mu)],$$

Now

$$\phi^*[S_\lambda, S_\mu] = \phi^*\left(\binom{n}{\lambda}\binom{n}{\mu}(s_\lambda, s_\mu)s_\lambda\right)$$

$$= \phi^*\left(\delta_{\lambda\mu}\binom{n}{\lambda}^2 \frac{n!}{\binom{n}{\lambda}|\mu(0,\pi)|} s_\lambda\right) \qquad (\text{for } \pi \in \lambda)$$

$$= \phi^*\left(\delta_{\lambda\mu}\frac{n!}{|\mu(0,\pi)|}S_\lambda\right)$$

$$= \delta_{\lambda\mu}\cdot\frac{n!}{|\mu(0,\pi)|}\tilde{S}_\lambda$$

and

$$[\phi^*(S_\lambda), \phi^*(S_\mu)] = [\tilde{S}_\lambda, \tilde{S}_\mu]$$

$$= \left[\sum_{\pi\in\lambda}\tilde{s}_\pi, \sum_{\sigma\in\mu}\tilde{s}_\sigma\right]$$

$$= \sum_{\pi\in\lambda,\sigma\in\mu}[\tilde{s}_\pi, \tilde{s}_\sigma]$$

$$= \delta_{\lambda\mu}\sum_{\pi\in\lambda}(\tilde{s}_\pi, \tilde{s}_\pi)\tilde{s}_\pi$$

$$= \delta_{\lambda\mu}\cdot\frac{n!}{|\mu(0,\pi)|}\sum_{\pi\in\lambda}\tilde{s}_\pi$$

$$= \delta_{\lambda\mu}\frac{n!}{|\mu(0,\pi)|}\tilde{S}_\lambda \qquad (\text{for } \pi\in\lambda),$$

which proves (ii).

(iii) Apply ϕ to (ii) and use the fact that $\phi\phi^* = I$.

(iv) $[\tilde{\theta}(\tilde{s}_\pi), \tilde{\theta}(\tilde{s}_\sigma)] = [(\text{sign }\pi)\tilde{s}_\pi, (\text{sign }\sigma)\tilde{s}_\sigma] = (\text{sign }\pi)(\text{sign }\sigma)[\tilde{s}_\pi, \tilde{s}_\sigma] = [\tilde{s}_\pi, \tilde{s}_\sigma]$, since $[\tilde{s}_\pi, \tilde{s}_\sigma] = 0$ if $\pi \neq \sigma$.

Part (iii) of the preceding theorem can be used to translate computations of the Kronecker inner product in $\mathscr{S}$ to a computation in $\tilde{\mathscr{S}}$, which is usually easier. As an example, let us compute $[h_\lambda, h_\mu]$:

First,

$$[\tilde{h}_\pi, \tilde{h}_\sigma] = \left[\sum_{\tau\leq\pi}|\mu(0,\tau)|\tilde{s}_\tau, \sum_{\nu\leq\sigma}|\mu(0,\nu)|\tilde{s}_\nu\right]$$

$$= \sum_{\tau\leq\pi,\nu\leq\sigma}|\mu(0,\tau)|\cdot|\mu(0,\nu)|[\tilde{s}_\tau, \tilde{s}_\nu]$$

$$= \sum_{\tau\leq\pi,\sigma}|\mu(0,\tau)|^2[\tilde{s}_\tau, \tilde{s}_\tau] \quad (\text{since } [\tilde{s}_\tau, \tilde{s}_\nu] = 0 \quad \text{if } \tau \neq \nu)$$

$$= \sum_{\tau\leq\pi\wedge\sigma}|\mu(0,\tau)|^2\frac{n!}{|\mu(0,\tau)|}\tilde{s}_\tau$$

$$= n!\sum_{\tau\leq\pi\wedge\sigma}|\mu(0,\tau)|\tilde{s}_\tau$$

$$= n!\tilde{h}_{\pi\wedge\sigma}.$$

Thus,

$$[h_\lambda, h_\mu] = \frac{1}{\lambda!\binom{n}{\lambda}\mu!\binom{n}{\mu}}[H_\lambda, H_\mu]$$

$$= \frac{1}{\lambda!\binom{n}{\lambda}\mu!\binom{n}{\mu}}\phi[\phi^*(H_\lambda), \phi^*(H_\mu)]$$

$$= \frac{|\lambda|\cdot|\mu|}{(n!)^2}\phi\left[\sum_{\pi\in\lambda}\tilde{h}_\pi, \sum_{\sigma\in\mu}\tilde{h}_\sigma\right]$$

$$= \frac{|\lambda||\mu|}{(n!)}\sum_{\pi\in\lambda,\sigma\in\mu}\phi(\tilde{h}_{\pi\wedge\sigma})$$

$$= \frac{|\lambda||\mu|}{(n!)}\sum_{\pi\in\lambda,\sigma\in\mu}\lambda(\pi\wedge\sigma)!h_{\lambda(\pi\wedge\sigma)}$$

This shows that $[h_\lambda, h_\mu]$ is a positive linear combination of h_ρ's, which can be interpreted to say something interesting about permutation representations of S_n. Also, since

$$[\tilde{a}_\pi, \tilde{a}_\sigma] = [\tilde{\theta}(\tilde{a}_\pi), \tilde{\theta}(\tilde{a}_\sigma)] = [\tilde{h}_\pi, \tilde{h}_\sigma] = n!\tilde{h}_{\pi\wedge\sigma},$$

it follows that $[a_\lambda, a_\mu]$ is a positive linear combination of h_ρ's, again a fact of interest.

A complete list of Kronecker products of the various bases is to be found in Appendix 3.

Appendix 1: Connections between bases

1. $s_\pi = \sum_{\sigma\geq\pi} k_\sigma$ $\qquad$ $k_\pi = \sum_{\sigma\geq\pi}\mu(\pi,\sigma)s_\sigma$

2. $a_\pi = \sum_{\sigma:\sigma\wedge\pi=0} k_\sigma$ $\qquad$ $k_\pi = \sum_{\tau\geq\pi}\frac{\mu(\pi,\tau)}{\mu(0,\tau)}\sum_{\sigma\leq\tau}\mu(\sigma,\tau)a_\sigma$

3. $a_\pi = \sum_{\sigma\leq\pi}\mu(0,\sigma)s_\sigma$ $\qquad$ $s_\pi = \frac{1}{\mu(0,\pi)}\sum_{\sigma\leq\pi}\mu(\sigma,\pi)a_\sigma$

4. $s_\pi = \sum_{\sigma\geq\pi}\text{sign}(\pi,\sigma)f_\sigma$ $\qquad$ $f_\pi = \sum_{\sigma\geq\pi}|\mu(\pi,\sigma)|s_\sigma$

5. $h_\pi = \sum_{\sigma:\sigma\wedge\pi=0}(\text{sign}\,\sigma)f_\sigma$ $\qquad$ $f_\pi = \sum_{\tau\geq\pi}\frac{|\mu(\pi,\tau)|}{|\mu(0,\tau)|}\sum_{\sigma\leq\tau}\mu(\sigma,\tau)h_\sigma$

6. $h_\pi = \sum_{\sigma\leq\pi}|\mu(0,\sigma)|s_\sigma$ $\qquad$ $s_\pi = \frac{1}{|\mu(0,\pi)|}\sum_{\sigma\leq\pi}\mu(\sigma,\pi)h_\sigma$

7. $h_\pi = \sum_\sigma\lambda(\pi\wedge\sigma)!k_\sigma$ $\qquad$ $k_\pi = \sum_{\tau\geq\pi}\frac{\mu(\pi,\tau)}{|\mu(0,\tau)|}\sum_{\sigma\leq\tau}\mu(\sigma,\tau)h_\sigma$

8. $a_\pi = \sum_\sigma(\text{sign}\,\sigma)\lambda(\pi\wedge\sigma)!f_\sigma$ $\qquad$ $f_\pi = \sum_{\tau\geq\pi}\frac{|\mu(\pi,\tau)|}{\mu(0,\tau)}\sum_{\sigma\leq\tau}\mu(\sigma,\tau)a_\sigma$

9. $k_\pi = \sum_{\sigma \geq \pi} \operatorname{sign}(\pi, \sigma)\lambda(\pi, \sigma)!f_\sigma \qquad f_\pi = \sum_{\sigma \geq \pi} \lambda(\pi, \sigma)!k_\sigma$

10. $a_\pi = \sum_{\sigma \leq \pi} (\operatorname{sign} \sigma)\lambda(\sigma, \pi)!h_\sigma \qquad h_\pi = \sum_{\sigma \leq \pi} (\operatorname{sign} \sigma)\lambda(\sigma, \pi)!a_\sigma$

Appendix 2: The inner product on $\tilde{\mathscr{S}}$

1. $(\tilde{h}_\pi, \tilde{k}_\sigma) = n!\delta_{\pi\sigma}$

2. $(\tilde{s}_\pi, \tilde{s}_\sigma) = \delta_{\pi\sigma}\dfrac{n!}{|\mu(0, \pi)|}$

3. $(\tilde{h}_\pi, \tilde{s}_\sigma) = n!\zeta(\sigma, \pi)$, where $\zeta(\sigma, \pi) = \begin{cases} 1 & \text{if } \sigma \leq \pi \\ 0 & \text{if not} \end{cases}$

4. $(\tilde{a}_\pi, \tilde{s}_\sigma) = n!(\operatorname{sign} \sigma)\zeta(\sigma, \pi)$

5. $(\tilde{a}_\pi, \tilde{a}_\sigma) = n!\lambda(\sigma \wedge \pi)!$

6. $(\tilde{a}_\pi, \tilde{h}_\sigma) = \begin{cases} n! & \text{if } \pi \wedge \sigma = 0 \\ 0 & \text{if } \pi \wedge \sigma \neq 0 \end{cases}$

7. $(\tilde{k}_\pi, \tilde{k}_\sigma) = n! \sum_{\tau \geq \pi \vee \sigma} \dfrac{\mu(\pi, \tau)\mu(\sigma, \tau)}{|\mu(0, \tau)|}$

8. $(\tilde{a}_\pi, \tilde{k}_\sigma) = n!(\operatorname{sign} \sigma)\lambda(\sigma, \pi)!\zeta(\sigma, \pi)$

9. $(\tilde{h}_\pi, \tilde{h}_\sigma) = n!\lambda(\sigma \wedge \pi)!$

10. $(\tilde{k}_\pi, \tilde{s}_\sigma) = n!\dfrac{\mu(\pi, \sigma)}{|\mu(0, \sigma)|}\zeta(\pi, \sigma)$

11. $(\tilde{f}_\pi, \tilde{k}_\sigma) = n! \sum_{\tau \geq \pi \vee \sigma} \dfrac{|\mu(\pi, \tau)|\mu(\sigma, \tau)}{|\mu(0, \tau)|}$

12. $(\tilde{f}_\pi, \tilde{f}_\sigma) = (\operatorname{sign} \pi)(\operatorname{sign} \sigma)n! \sum_{\tau \geq \pi \vee \sigma} \dfrac{\mu(\pi, \tau)\mu(\sigma, \tau)}{|\mu(0, \tau)|}$

13. $(\tilde{f}_\pi, \tilde{h}_\sigma) = n!\lambda(\pi, \sigma)!\zeta(\pi, \sigma)$

14. $(\tilde{f}_\pi, \tilde{s}_\sigma) = (\operatorname{sign} \pi)(\operatorname{sign} \sigma)n!\dfrac{\mu(\pi, \sigma)}{|\mu(0, \sigma)|}\zeta(\pi, \sigma)$

15. $(\tilde{f}_\pi, \tilde{a}_\sigma) = n!(\operatorname{sign} \pi)\delta_{\pi\sigma}$

Appendix 3: The Kronecker inner product on $\tilde{\mathscr{S}}$

1. $[\tilde{s}_\pi, \tilde{s}_\sigma] = \dfrac{n!}{|\mu(0, \pi)|}\delta_{\pi\sigma}\tilde{s}_\pi$

2. $[\tilde{h}_\pi, \tilde{s}_\sigma] = n!\zeta(\sigma, \pi)\tilde{s}_\sigma$

3. $[\tilde{a}_\pi, \tilde{s}_\sigma] = n!(\operatorname{sign} \sigma)\zeta(\sigma, \pi)\tilde{s}_\sigma$

4. $[\tilde{f}_\pi, \tilde{s}_\sigma] = n!\dfrac{|\mu(\pi,\sigma)|}{|\mu(0,\sigma)|}\zeta(\pi,\sigma)\tilde{s}_\sigma$

5. $[\tilde{k}_\pi, \tilde{s}_\sigma] = n!\dfrac{\mu(\pi,\sigma)}{|\mu(0,\sigma)|}\zeta(\pi,\sigma)\tilde{s}_\sigma$

6. $[\tilde{h}_\pi, \tilde{h}_\sigma] = n!\tilde{h}_{\pi\wedge\sigma}$

7. $[\tilde{a}_\pi, \tilde{a}_\sigma] = n!\tilde{h}_{\pi\wedge\sigma}$

8. $[\tilde{a}_\pi, \tilde{h}_\sigma] = n!\tilde{a}_{\pi\wedge\sigma}$

9. $[\tilde{k}_\pi, \tilde{h}_\sigma] = n!\sum_{\tau\in[\pi,\sigma]}\mu(\pi,\tau)\tilde{s}_\tau$

10. $[\tilde{f}_\pi, \tilde{h}_\sigma] = n!\sum_{\tau\in[\pi,\sigma]}|\mu(\pi,\tau)|\tilde{s}_\tau$

11. $[\tilde{k}_\pi, \tilde{a}_\sigma] = n!(\text{sign }\pi)\sum_{\tau\in[\pi,\sigma]}|\mu(\pi,\tau)|\tilde{s}_\tau$

12. $[\tilde{k}_\pi, \tilde{k}_\sigma] = n!\sum_{\tau\geq\pi\vee\sigma}\dfrac{\mu(\pi,\tau)\mu(\sigma,\tau)}{|\mu(0,\tau)|}\tilde{s}_\tau$

13. $[\tilde{f}_\pi, \tilde{f}_\sigma] = n!\sum_{\tau\geq\pi\vee\sigma}\dfrac{|\mu(\pi,\tau)|\,|\mu(\sigma,\tau)|}{|\mu(0,\tau)|}\tilde{s}_\tau$

14. $[\tilde{k}_\pi, \tilde{f}_\sigma] = n!\sum_{\tau\geq\pi\vee\sigma}\dfrac{\mu(\pi,\tau)|\mu(\sigma,\tau)|}{|\mu(0,\tau)|}\tilde{s}_\tau$

15. $[\tilde{f}_\pi, \tilde{a}_\sigma] = n!(\text{sign }\pi)\sum_{\tau\in[\pi,\sigma]}\mu(\pi,\tau)\tilde{s}_\tau$

References

1. G. ANDREWS, On the foundations of combinatorial theory. V: Eulerian differential operators, Studies in Applied Math. **50** (1971), 345–375.
2. G. BIRKHOFF, Lattice Theory, third ed., Amer. Math. Soc., Providence, R. I. (1967).
3. G. BIRKHOFF AND S. MACLANE, A Survey of Modern Algebra, third ed., Macmillan, New York (1965).
4. C. CURTIS AND I. REINER, Representation Theory of Finite Groups and Associative Algebras, Wiley, New York (1962).
5. F. DAVID, M. G. KENDALL AND D. BARTON, Symmetric Function and Allied Tables, Cambridge Univ. Press (1966).
6. P. DOUBILET, G.-C. ROTA AND R. STANLEY, On the foundations of combinatorial theory. VI: The idea of generating function, Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, Berkeley Univ. Press (1972).
7. P. DWYER, Properties of polykays of deviates, Ann. Math. Statistics **35** (1964), 1167–1173.
9. P. DWYER AND D. TRACY, A combinatorial method for products of two polykays with some general formulae, Ann. Math. Statistics **35** (1964), 1174–1185.
9. J. GOLDMAN AND G.-C. ROTA, On the foundations of combinatorial theory. IV: Finite vector spaces and Eulerian generating functions, Studies in Applied Math. **49** (1970), 239–258.
10. P. HALL, The algebra of partitions, Proc. 4th Canad. Math. Congr. Banff, 1957 (1959), 147–159.
11. M. G. KENDALL AND A. STUART, The Theory of Advanced Statistics, Vol. 1, third ed., Hafner, New York (1969).
12. D. E. LITTLEWOOD, The Theory of Group Characters, second ed., Oxford Univ. Press (1950).
13. D. E. LITTLEWOOD, A University Algebra, Heinemann, London (1950).

14. P. A. MacMahon, Combinatory Analysis, Vols. 1–2, Cambridge Univ. Press (1915, 1916); reprinted by Chelsea, New York (1960).
15. R. Mullin and G.-C. Rota, On the foundations of combinatorial theory. III: Theory of binomial enumeration, Graph Theory and its Applications, Academic Press (1970), 167–213.
16. F. D. Murnaghan, The Theory of Group Representations, Johns Hopkins, Baltimore (1938).
17. G. de B. Robinson, Representation Theory of the Symmetric Group, University of Toronto Press (1961).
18. G.-C. Rota, On the foundations of combinatorial theory. I: Theory of Mobius functions, Z. Wahrscheinlichkeitstheorie 2 (1964), 340–368.
19. R. Stanley, Theory and applications of plane partitions: Part I, Studies in Applied Math. 50 (1971), 167–188.
20. D. Tracy, Some rules for a combinatorial method for multiple products of generalized k-statistics, Ann. Math. Statistics **39** (1968), 983–998.

Massachusetts Institute of Technology

(Received May 18, 1972)

JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 42, 684–760 (1973)

On the Foundations of Combinatorial Theory. VIII. Finite Operator Calculus

GIAN-CARLO ROTA,* D. KAHANER, AND A. ODLYZKO†

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
and
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

Submitted by Richard Bellman

TO SALOMON BOCHNER, MY FIRST TEACHER OF ANALYSIS, WHO TAUGHT US THAT THEORY IS THE CAPTAIN, AND COMPUTATION THE SOLDIER.

CONTENTS

* Work partially supported by the Office of Naval Research.
† Work supported by Hertz Foundation.

1. Introduction

The so-called Heaviside calculus, invented by Boole and developed without interruption to our day, is the mainspring of much contemporary work in operator theory and harmonic analysis. The spectacular analytic developments in these fields in the last fifty years, coupled with current grandiose plans for unification, cannot, however, be said to have been matched by equal strides in the computational and algorithmic aspects. The algebraic aspects of the theory of special functions have not significantly changed since the nineteenth century. As a result, a deep cleavage is now apparent between the breadth of theory and the clumsiness of special cases.

In this work we reduce to a minimum the analytic apparatus of harmonic analysis on the line, by considering only polynomials. Our objective is to present a unified theory of special polynomials by exploiting to the hilt the duality between x and d/dx.

The main technique adopted here is a rigorous version—perhaps the first one—of the so-called "umbral calculus" or "symbolic calculus," widely used in the past century. This gives an effective technique for expressing a set of polynomials in terms of another. We have throughout emphasized operator methods at the expense of generating functions, which were almost exclusively used in the past. No doubt several results given later could be rephrased in terms of generating functions, but only at the expense of conceptual clarity. Umbral methods, we hope to show, are operators in disguise.

The three kinds of polynomial sequences studied are:

(a) sequences of *binomial type*, that is, sequences of polynomials $p_n(x)$ satisfying the identities

$$p_n(x+y) = \sum_{k\geqslant 0} \binom{n}{k} p_k(x)\, p_{n-k}(y).$$

These sequences were studied in the third part of the series (referred to as III), but we repeat the main results here, both in order to render this work self-contained and in order to give some results in greater generality.

(b) *Sheffer sets*, that is, sequences $s_n(x)$ of polynomials satisfying the identities

$$s_n(x+y) = \sum_{k\geqslant 0} \binom{n}{k} s_k(x)\, p_{n-k}(y),$$

where $p_n(x)$ is a given sequence of binomial type.

(c) *Cross-sequences*, namely doubly indexed sequences $p_n^{[\lambda]}(x)$ of polynomials, satisfying

$$p_n^{[\lambda+\mu]}(x+y) = \sum_{k\geqslant 0} \binom{n}{k} p_k^{[\lambda]}(x)\, p_{n-k}^{[\mu]}(y).$$

This last theory is only touched upon here, and remains largely undeveloped.

One of the unexpected consequences of the present algebraic approach is that the theory of eigenfunction expansions for polynomials can be rendered purely algebraic. This gives a meaning to eigenfunction expansions for Hermite polynomials of arbitrary variance and for Laguerre polynomials of arbitrary α (except a negative integer, where the gamma function is not defined).

A number of examples, each of which includes, we would like to hope, a little novelty, is given at the end, both as an illustration of the theory and to show how much of the past literature on special polynomials is the iteration of a few basic principles. We have, however, resisted the temptation of developing a theory of combinatorial identities as an application, outside of a few hints.

2. Basic Polynomials

We shall be concerned with the algebra (over a field of characteristic zero) of all polynomials $p(x)$ in one variable, to be denoted **P**.

By a *polynomial sequence* we shall denote a sequence of polynomials $p_i(x)$, $i = 0, 1, 2,\ldots$, where $p_i(x)$ is exactly of degree i for all i.

A polynomial sequence is said to be of ***binomial type*** if it satisfies the infinite sequence of identities

$$p_n(x + y) = \sum_{k\geqslant 0} \binom{n}{k} p_k(x)\, p_{n-k}(y), \qquad n = 0, 1, 2, \ldots$$

The simplest sequence of binomial type is of course x^n, but we give some nontrivial examples. Other examples are found in III.

The present theory revolves around the interplay between the algebra of polynomials and another algebra, to be presently introduced and to be denoted by Σ, namely, the *algebra of shift-invariant operators*. All operators we consider are, of course, tacitly assumed to be *linear*. We denote the action of an operator T on the polynomial $p(x)$ by $Tp(x)$. This notation is not, strictly speaking, correct; a correct version is $(Tp)(x)$. However, our notational license results in greater readability.

The most important shift-invariant operators are the shift operators, written E^a, that is, $E^a p(x) = p(x + a)$. Other examples are given later.

An operator T which commutes with all shift operators is called a *shift-invariant operator*. In symbols, $TE^a = E^a T$, for all real a in the field.

We define a *delta operator*, usually denoted by the letter Q, as a shift-invariant operator for which Qx is a nonzero constant.

Delta operators possess many of the properties of the derivative operator, as we will show. In fact our first objective is to exploit the analogy between delta operators and the ordinary derivative.

PROPOSITION 1. *If Q is a delta operator, then $Qa = 0$ for every constant a.*

Proof. Since Q is shift invariant, we have

$$QE^a x = E^a Qx.$$

By the linearity of Q,

$$QE^a x = Q(x + a) = Qx + Qa = c + Qa,$$

since Qx is equal to some nonzero constant c by definition. But also

$$E^a Qx = E^a c = c$$

and so $c + Qa = c$. Hence, $Qa = 0$. Q.E.D.

PROPOSITION 2. *If $p(x)$ is a polynomial of degree n and Q is a delta operator, then $Qp(x)$ is a polynomial of degree $n - 1$.*

Proof. It is sufficient to consider the special case $p(x) = x^n$. From the binomial theorem and the linearity of Q, we have

$$Q(x+a)^n = \sum_{k\geq 0} \binom{n}{k} a^k Q x^{n-k}.$$

Also by the shift-invariance of Q

$$Q(x+a)^n = QE^a x^n = E^a Q x^n = r(x+a)$$

say, so that

$$r(x+a) = \sum_{k\geq 0} \binom{n}{k} a^k Q x^{n-k}.$$

Setting $x = 0$, we have expressed the polynomial $r(x)$ as a polynomial in the parameter a,

$$r(a) = \sum_{k\geq 0} \binom{n}{k} a^k [Qx^{n-k}]_{x=0}\,.$$

The coefficient of a^n is

$$[Qx^{n-n}]_{x=0} = [Q1]_{x=0} = 0$$

by Proposition 1. Further, the coefficient of a^{n-1} is

$$\binom{n}{n-1} [Qx^{n-n+1}]_{x=0} = n[Qx]_{x=0} = nc \neq 0.$$

Hence r is of degree $n-1$. Q.E.D.

Let Q be a delta operator. A polynomial sequence $p_n(x)$ is called the sequence of *basic polynomials* for Q if:

(1) $p_0(x) = 1$;

(2) $p_n(0) = 0$ whenever $n > 0$;

(3) $Qp_n(x) = np_{n-1}(x)$.

PROPOSITION 3. *Every delta operator has a unique sequence of basic polynomials.*

Proof. Inducing on n, assume that $p_k(x)$ has been defined for $k < n$ to satisfy the foregoing conditions. We show that $p_n(x)$ also exists and is unique. Indeed, a generic polynomial of degree n can be written in the form

$$p(x) = ax^n + \sum_{k=0}^{n-1} c_k p_k(x), \qquad a \neq 0.$$

Now,

$$Qp(x) = aQx^n + \sum_{k=1}^{n-1} c_k \cdot kp_{k-1}(x);$$

therefore, Qx^n being exactly of degree $n - 1$, there is a unique choice of the constants $c_1, \ldots, c_{n-1}$, a for which $Qp(x) = np_{n-1}(x)$. This determines $p(x)$ except for the constant term c_0, but this is in turn uniquely determined by the condition $p(0) = 0$. Q.E.D.

The typical example of a basic polynomial sequence is x^n, basic for the derivative operator D. Others are given later, or can be looked up in III.

Several properties of the polynomial sequence x^n can be generalized to an arbitrary sequence of basic polynomials. A basic property of x^n is that it is of binomial type. This turns out to be true for every sequence of basic polynomials and is one of our basic results.

THEOREM 1. (a) *If $p_n(x)$ is a basic sequence for some delta operator Q, then it is a sequence of polynomials of binomial type.*

(b) *If $p_n(x)$ is a sequence of polynomials of binomial type, then it is a basic sequence for some delta operator.*

Proof. (a) Iterating property (3) of basic polynomials, we see that

$$Q^k p_n(x) = (n)_k \, p_{n-k}(x),$$

where

$$(n)_k = n(n-1) \cdots (n-k+1).$$

And, hence, for $k = n$,

$$[Q^n p_n(x)]_{x=0} = n!,$$

while for $k < n$,

$$[Q^k p_n(x)]_{x=0} = 0.$$

Thus, we may trivially express $p_n(x)$ in the form

$$p_n(x) = \sum_{k \geqslant 0} \frac{p_k(x)}{k!} [Q^k p_n(x)]_{x=0} \,.$$

Since any polynomial $p(x)$ is a linear combination of the basic polynomials $p_n(x)$, this expression also holds for all polynomials $p(x)$, that is,

$$p(x) = \sum_{k \geqslant 0} \frac{p_k(x)}{k!} [Q^k p(x)]_{x=0} \,.$$

Now suppose $p(x)$ is the polynomial $p_n(x+y)$ for fixed y. Then

$$p_n(x+y) = \sum_{k\geqslant 0} \frac{p_k(x)}{k!} [Q^k p_n(x+y)]_{x=0}.$$

But

$$\begin{aligned}[Q^k p_n(x+y)]_{x=0} &= [Q^k E^y p_n(x)]_{x=0} \\ &= [E^y Q^k p_n(x)]_{x=0} = [E^y (n)_k p_{n-k}(x)]_{x=0} = (n)_k p_{n-k}(y),\end{aligned}$$

and so

$$p_n(x+y) = \sum_{k\geqslant 0} \binom{n}{k} p_k(x) p_{n-k}(y);$$

that is, the sequence $p_n(x)$ is of binomial type.

(b) Suppose now $p_n(x)$ is a sequence of binomial type. Setting $y=0$ in the binomial identity, we obtain

$$\begin{aligned}p_n(x) &= \sum_{k\geqslant 0} \binom{n}{k} p_k(x) p_{n-k}(0) \\ &= p_n(x) p_0(0) + n p_{n-1}(x) p_1(0) + \cdots .\end{aligned}$$

Since each $p_i(x)$ is exactly of degree i, it follows that $p_0(0) = 1$ (and, hence, $p_0(x) = 1$) and $p_i(0) = 0$ for all other i. Thus. properties (1) and (2) of basic sequences are satisfied.

We next define a delta operator for which such a sequence $p_n(x)$ is the sequence of basic polynomials. Let Q be the operator defined by the property that $Qp_0(x) = 0$ and $Qp_n(x) = np_{n-1}(x)$ for $n \geqslant 1$. Clearly Qx must be a nonzero constant. Hence, all that remains to be shown is that Q is shift-invariant.

We may trivially write the property of being of binomial type in the form

$$p_n(x+y) = \sum_{k\geqslant 0} \frac{p_k(x)}{k!} Q^k p_n(y),$$

and, repeating the device used in (a), this may be extended to all polynomials:

$$p(x+y) = \sum_{k\geqslant 0} \frac{p_k(x)}{k!} Q^k p(y).$$

Now replace p by Qp and interchange x and y on the right—an operation which leaves the left side invariant—to get

$$(Qp)(x+y) = \sum_{k\geqslant 0} \frac{p_k(y)}{k!} Q^{k+1} p(x).$$

But

$$(Qp)(x+y) = E^y(Qp)(x) = E^yQp(x)$$

and

$$\sum_{k\geqslant 0} \frac{p_k(y)}{k!} Q^{k+1}p(x) = Q\left(\sum_{k\geqslant 0} \frac{p_k(y)}{k!} Q^k p(x)\right)$$
$$= Q(p(x+y)) = QE^y p(x).$$

Q.E.D.

3. The First Expansion Theorem

We study next the expansion of a shift-invariant operator in terms of a delta operator and its powers. The difficulties caused by convergence questions are minimal, and we refuse to discuss them in this paper (but see III).

The following theorem generalizes the Taylor expansion theorem to delta operators and their basic polynomials.

Theorem 2 (First Expansion Theorem). *Let T be a shift-invariant operator, and let Q be a delta operator with basic set $p_n(x)$. Then*

$$T = \sum_{k\geqslant 0} \frac{a_k}{k!} Q^k$$

with

$$a_k = [Tp_k(x)]_{x=0}.$$

Proof. Since the polynomials $p_n(x)$ are of binomial type (Theorem 1), we may write the binomial formula as in the preceding proof:

$$p_n(x+y) = \sum_{k\geqslant 0} \frac{p_k(x)}{k!} Q^k p_n(y).$$

Apply T to both sides (regarding x as the variable and y as a parameter) and get

$$Tp_n(x+y) = \sum_{k\geqslant 0} \frac{Tp_k(x)}{k!} Q^k p_n(y).$$

Once more, by linearity, this expression can be extended to all polynomials p. After doing this and setting x equal to zero, we can replace y by x and get

$$Tp(x) = \sum_{k\geqslant 0} \frac{[Tp_k(y)]_{y=0}}{k!} Q^k p(x).$$

Q.E.D.

The reader may apply the preceding theorem to derive several of the classical expansion formulas of numerical analysis. Our present application will be of a more theoretical nature:

THEOREM 3. *Let Q be a delta operator, and let* **F** *be the ring of formal power series in the variable t over the same field. Then there exists an isomorphism from F onto the ring Σ of shift-invariant operators, which carries*

$$f(t) = \sum_{k\geqslant 0} \frac{a_k t^k}{k!} \qquad \textit{into} \qquad \sum_{k\geqslant 0} \frac{a_k}{k!} Q^k.$$

Proof. The mapping is clearly linear, and by the first expansion theorem, it is onto. Therefore, all we have to verify is that the map preserves products. Let T be the shift-invariant operator corresponding to the formal power series $f(t)$ and let S be the shift-invariant operator corresponding to

$$g(t) = \sum_{k\geqslant 0} \frac{b_k}{k!} t^k.$$

We must verify that

$$[TSp_n(x)]_{x=0} = \sum_{k\geqslant 0} \binom{n}{k} a_k b_{n-k},$$

where $p_n(x)$ are the basic polynomials of Q. Now

$$\begin{aligned}[TSp_r(x)]_{x=0} &= \Big[\Big(\sum_{k\geqslant 0} \frac{a_k}{k!} Q^k \sum_{n\geqslant 0} \frac{b_n}{n!} Q^n\Big) p_r(x)\Big]_{x=0} \\ &= \Big[\sum_{k\geqslant 0} \sum_{n\geqslant 0} \frac{a_k b_n}{k!n!} Q^{k+n} p_r(x)\Big]_{x=0}.\end{aligned}$$

But $p_n(0) = 0$ for $n > 0$ and $p_0(x) = 1$. The only nonzero terms of the double sum occur when $n = r - k$. Thus,

$$\begin{aligned}[TSp_r(x)]_{x=0} &= \Big[\sum_{k\geqslant 0} \frac{a_k b_{r-k}}{k!(r-k)!} Q^r p_r(x)\Big]_{x=0} \\ &= \Big[\sum_{k\geqslant 0} \frac{a_k b_{r-k}}{k!(r-k)!} r! p_0(x)\Big]_{x=0} \\ &= \sum_{k\geqslant 0} \binom{r}{k} a_k b_{r-k}. \qquad \text{Q.E.D.}\end{aligned}$$

COROLLARY 1. *A shift-invariant operator T is invertible if and only if $T1 \neq 0$.*

In the following, we shall write $P = p(Q)$, where P is a shift-invariant operator and $p(t)$ is a formal power series, to indicate that the operator P corresponds to the formal power series $p(t)$ under the isomorphism of Theorem 3.

COROLLARY 2. *An operator P is a delta operator if and only if it corresponds, under the isomorphism of Theorem 3, to a formal power series $p(t)$ such that $p(0) = 0$ and $p'(0) \neq 0$.*

Recall that to every formal power series $p(t)$ such that $p(0) = 0$ and $p'(t) \neq 0$ there corresponds a unique inverse power series $p^{-1}(t)$. In symbols, if

$$p(t) = \sum_{k \geq 1} \frac{a_k}{k!} t^k,$$

then

$$p(p^{-1}(t)) = \sum_{k \geq 1} \frac{a_k}{k!} (p^{-1}(t))^k = t,$$

where the sum is well defined, since $p^{-1}(0) = 0$ and $(p^{-1})'(0) \neq 0$. Similarly we have $p^{-1}(p(t)) = t$.

Essentially, the problem we wish to solve in the present paper is the following: to what "operation" in the ring of shift-invariant operators corresponds the operation of composition $p(q(t))$ of power series with $q(0) = 0$, under the isomorphism theorem? Remarkably, this question does have an answer in the present context.

Next, we connect some of the preceding results with generating functions.

COROLLARY 3. *Let Q be a delta operator with basic polynomials $p_n(x)$, and let $q(D) = Q$. Let $q^{-1}(t)$ be the inverse formal power series. Then*

$$\sum_{n \geq 0} \frac{p_n(x)}{n!} u^n = e^{xq^{-1}(u)}.$$

Proof. Expand E^a in terms of Q by the first expansion theorem. The coefficients a_n are $p_n(a)$. Hence,

$$\sum_{n \geq 0} \frac{p_n(a)}{n!} Q^n = E^a,$$

a formula which can be considered as a generalization of Taylor's formula, and which specializes to several other classical expansions. Now use the isomorphism theorem with D as the delta operator. We get

$$\sum_{n \geqslant 0} \frac{p_n(a)}{n!} q(t)^n = e^{at},$$

whence the conclusion, upon setting $u = q(t)$ and $a = x$. Q.E.D.

This result will be interpreted more explicitly later (see Section 4). Finally, we note a fact that has already been implicitly used.

COROLLARY 4. *Any two shift-invariant operators commute.*

4. The Pincherle Derivative

For the first time we introduce operators that are not shift-invariant. The simplest is multiplication by x. Let $p(x)$ be a polynomial. Multiplying each term of $p(x)$ by the variable x, that is, replacing each occurrence of x^n by x^{n+1}, $n \geqslant 0$, we obtain a new polynomial $xp(x)$. Call this the *multiplication operator* and we denote it by $\mathbf{x}$. Thus, $\mathbf{x}: p(x) \to xp(x)$. For any operator T defined on $\mathbf{P}$, the operator

$$T' = T\mathbf{x} - \mathbf{x}T,$$

will be called the *Pincherle derivative* of the operator T.

PROPOSITION 1. *If T is a shift-invariant operator, then its Pincherle derivative,*

$$T' = T\mathbf{x} - \mathbf{x}T,$$

is also a shift-invariant operator.

The proof is a straightforward verification.

As a special case of the first expansion theorem, it follows that any shift-invariant operator T can be expressed in terms of D, that is

$$T = \sum_{k \geqslant 0} \frac{a_k}{k!} D^k,$$

where $a_k = [Tx^k]_{x=0}$. Further, by the isomorphism theorem (Theorem 3) the formal power series corresponding to T is

$$\sum_{k \geqslant 0} \frac{a_k}{k!} t^k = f(t).$$

We call $f(t)$ the *indicator* of T.

Proposition 2. *If T has indicator $f(t)$, then its Pincherle derivative T' has $f'(t)$ as its indicator.*

The proof is a direct verification. Similarly, from the isomorphism theorem and from the preceding proposition, we easily infer the following.

Proposition 3. $(TS)' = T'S + TS'$.

And just as easily from the isomorphism theorem, we can infer Proposition 4.

Proposition 4. *Q is a delta operator if and only if $Q = DP$ for some shift-invariant operator P, where the inverse operator P^{-1} exists.*

We come now to the main result of this section, which enables us to compute basic sets for a given delta operator.

Theorem 4 (Closed forms). *If $p_n(x)$ is a sequence of basic polynomials for the delta operator $Q = DP$ (see Proposition 4), then for $n > 0$:*

(1) $p_n(x) = Q'P^{-n-1}x^n$;
(2) $p_n(x) = P^{-n}x^n - (P^{-n})'\, x^{n-1}$;
(3) $p_n(x) = xP^{-n}x^{n-1}$;
(4) (*Rodrigues formula*) $p_n(x) = x(Q')^{-1}\, p_{n-1}(x)$.

Proof. We shall first show that the right sides of (1) and (2) define the same polynomial sequence. Indeed,

$$\begin{aligned} Q'P^{-n-1} &= (DP)'\, P^{-n-1} \\ &= (D'P + DP')\, P^{-n-1}. \end{aligned}$$

Now, $D' = I$. Hence,

$$\begin{aligned} Q'P^{-n-1} &= P^{-n} + P'P^{-n-1}D \\ &= P^{-n} - (1/n)\,(P^{-n})'\, D, \end{aligned}$$

whence

$$Q'P^{-n-1}x^n = P^{-n}x^n - (P^{-n})'\, x^{n-1},$$

as desired. Next, recalling the definition of the Pincherle derivative of $(P^{-n})'$, we have

$$\begin{aligned} P^{-n}x^n - (P^{-n})'\, x^{n-1} &= P^{-n}x^n - (P^{-n}\mathbf{x} - \mathbf{x}P^{-n})\, x^{n-1} \\ &= xP^{-n}x^{n-1}, \end{aligned}$$

and, thus, the right side of formula (3) equals that of formulas (2) and (1). Setting

$$q_n(x) = Q'P^{-n-1}x^n$$

and writing $Q = DP$, we get

$$Qq_n(x) = DPQ'P^{-n-1}x^n = Q'P^{-n}\,Dx^n = nq_{n-1}(x).$$

Thus, if we can show that $q_n(0) = 0$ for $n > 0$, the proof that $q_n(x)$ is the sequence of basic polynomials for Q will be complete, and it will follow that formulas (1)–(3) are equivalent. From the equivalence of Eqs. (1)–(3) we see that

$$q_n(x) = xP^{-n}x^{n-1},$$

and hence $q_n(0) = 0$ for $n \geqslant 1$. Thus, (1)–(3) have been proved, and $q_n(x) = p_n(x)$.

To prove (4), first invert formula (1),

$$x^n = (Q')^{-1}\,P^{n+1}p_n(x).$$

Note that Q' is invertible (Isomorphism Theorem and Proposition 2). Change n to $n - 1$ and insert the right side into the right side of (3):

$$\begin{aligned} p_n(x) &= xP^{-n}(Q')^{-1}\,P^n p_{n-1}(x) \\ &= x(Q')^{-1}\,p_{n-1}(x), \end{aligned}$$

which is Rodrigues' formula. Q.E.D.

The following formulas relate the basic polynomials of two different delta operators in an analogous way. Their proof is immediate.

COROLLARY. *Let $R = DS$ and $Q = DP$ be delta operators with basic polynomials $r_n(x)$ and $p_n(x)$, respectively, where S^{-1} and P^{-1} exist. Then*

(5) $p_n(x) = Q'(R')^{-1}\,P^{-n-1}S^{n+1}r_n(x), \quad n \geqslant 0;$

(6) $p_n(x) = x(SP^{-1})^n\,x^{-1}r_n(x), \quad n \geqslant 1.$

A last (and useful) characterization of basic sets is the following theorem.

THEOREM 5. *Let P be an invertible shift-invariant operator. Let $p_n(x)$ be a sequence of basic polynomials satisfying*

$$[x^{-1}p_n(x)]_{x=0} = n[P^{-1}p_{n-1}(x)]_{x=0}\,,$$

for all $n > 0$. Then $p_n(x)$ is the sequence of basic polynomials for the delta operator $Q = DP$.

Proof. Define the operator Q by setting $Q1 = 0$,

$$Qp_n(x) = np_{n-1}(x)$$

and extending by linearity. It is easily seen that Q is shift-invariant. In terms of Q, the preceding identity can be rewritten in the form

$$[x^{-1}p_n(x)]_{x=0} = [P^{-1}Qp_n(x)]_{x=0}\,.$$

By linearity, this extends to an identity for all polynomials $p(x)$ with $p\,(0) = 0$ —an argument we have often used. Thus, recalling that

$$[x^{-1}p(x)]_{x=0} = [Dp(x)]_{x=0}$$

whenever $p(0) = 0$, we have

$$[Dp(x)]_{x=0} = [P^{-1}Qp(x)]_{x=0}$$

for all polynomials $p(x)$, including those for which $p(0) \neq 0$, since the formula trivially holds for constants. Setting $p(x) = q(x + a)$ we obtain, using the shift-invariance of P and Q,

$$\begin{aligned} Dq(a) &= [P^{-1}QE^a q(x)]_{x=0} \\ &= [E^a P^{-1}Qq(x)]_{x=0} \\ &= P^{-1}Qq(a), \end{aligned}$$

for all constants a. But this means that $D = P^{-1}Q$, or $Q = DP$. Q.E.D.

COROLLARY 1. *Given any sequence of constants* $c_{n,1}$, $n = 1, 2,\ldots,$ *with* $c_{1,1} \neq 0$ *there exists a unique sequence of basic polynomials* $p_n(x)$ *such that*

$$[x^{-1}p_n(x)]_{x=0} = c_{n,1}\,,$$

that is,

$$p_n(x) = \sum_{k\geqslant 1} c_{n,k}x^k, \qquad n = 1, 2,\ldots.$$

COROLLARY 2. *Let* $g(x)$ *be the indicator of* Q *in the preceding corollary. Then* $g = f^{-1}$, *where*

$$f(t) = \sum_{k\geqslant 1} c_{k,1}\frac{t^k}{k!}\,.$$

Proof. From Corollary 1

$$D = QP^{-1} = \sum_{k\geqslant 1} c_{k,1}\frac{Q^k}{k!} = f(Q),$$

and the result follows.

The preceding corollaries show that a sequence of basic polynomials is completely determined by the coefficients of their first power x. This fact can be made the starting point for a connection between the present theory and the theory of compound Poisson processes, as we hope to do elsewhere.

Note that the preceding corollary gives an explicit interpretation to the generating function of a sequence of basic polynomials, which can now be restated as

$$\sum_{n\geqslant 0} \frac{p_n(x)}{n!} t^n = \exp\left(x \sum_{k\geqslant 1} c_{k,1} \cdot t^k/k!\right),$$

a form which makes it almost evident.

5. Sheffer Polynomials

A polynomial sequence $s_n(x)$ is called a *Sheffer set* or a set of *Sheffer polynomials for the delta operator* Q if

(1) $s_0(x) = c \neq 0$,

(2) $Qs_n(x) = ns_{n-1}(x)$.

A Sheffer set for the delta operator Q is related to the set of basic polynomials of Q by the following.

Proposition 1. *Let Q be a delta operator with basic polynomial set $q_n(x)$. Then $s_n(x)$ is a Sheffer set relative to Q if and only if there exists an invertible shift invariant operator S such that*

$$s_n(x) = S^{-1}q_n(x).$$

Proof. Suppose first that $s_n(x) = S^{-1}q_n(x)$, where S is an invertible shift invariant operator. Then $S^{-1}Q = QS^{-1}$, and

$$\begin{aligned} Qs_n(x) &= QS^{-1}q_n(x) = S^{-1}Qq_n(x) \\ &= S^{-1}nq_{n-1}(x) = nS^{-1}q_{n-1}(x) = ns_{n-1}(x). \end{aligned}$$

Further, since S^{-1} is invertible $S^{-1}1 = c \neq 0$, by the isomorphism theorem, so that

$$s_0(x) = S^{-1}q_0(x) = S^{-1}1 = c.$$

Thus, $s_n(x)$ is a Sheffer set.

Conversely, if $s_n(x)$ is a Sheffer set for the delta operator Q, define S by setting

$$S : s_n(x) \to q_n(x),$$

and extending S by linearity, so that it is well defined on all polynomials.

Since the polynomials s_n and q_n are both of degree n, and $s_0(x) \neq 0$ S is invertible. It remains to show that S is shift-invariant. To this end, note that S commutes with Q. Indeed,

$$\begin{aligned} SQs_n(x) &= nSs_{n-1}(x) = nq_{n-1}(x) \\ &= Qq_n(x) = QSs_n(x), \end{aligned}$$

and again by the linearity argument we infer that $QS = SQ$; whence $SQ^n = Q^nS$. Finally, recall that by the first expansion theorem one has

$$E^t = \sum_{n\geqslant 0} \frac{a_n}{n!} Q^n, \qquad a_n = [E^t q_n(x)]_{x=0};$$

whence $E^tS = SE^t$ for all t. We conclude that S is shift-invariant.

Q.E.D.

Some of the properties of basic sets can be extended to Sheffer sets; one of the most important is

THEOREM 6 (Second Expansion Theorem). *Let Q be a delta operator with basic polynomials $q_n(x)$, let S be an invertible shift-invariant operator with Sheffer set $s_n(x)$. If T is any shift invariant operator, and $p(x)$ is any polynomial the following identity holds for all values of the parameter y:*

$$Tp(x+y) = \sum_{n\geqslant 0} \frac{s_n(y)}{n!} Q^n STp(x).$$

Proof. By the first expansion theorem we have

$$E^y = \sum_{n\geqslant 0} \frac{a_n}{n!} Q^n$$

with

$$a_n = [E^y q_n(x)]_{x=0} = [q_n(x+y)]_{x=0} = q_n(y);$$

that is,

$$E^y = \sum_{n\geqslant 0} \frac{q_n(y)}{n!} Q^n.$$

Applying this to $p(x)$,

$$E^y p(x) = p(x+y) = \sum_{n\geqslant 0} \frac{q_n(y)}{n!} Q^n p(x).$$

We may interchange the variables x and y in the sum without affecting the left side:

$$p(x+y) = \sum_{n\geq 0} \frac{q_n(x)}{n!} Q^n p(y).$$

Applying S^{-1}, regarding x as the variable and y as a parameter, this becomes

$$\begin{aligned} S^{-1}p(x+y) &= \sum_{n\geq 0} \frac{S^{-1}q_n(x)}{n!} Q^n p(y) \\ &= \sum_{n\geq 0} \frac{s_n(x)}{n!} Q^n p(y), \end{aligned}$$

for all y. Again interchanging the variables x and y

$$S^{-1}p(x+y) = \sum_{n\geq 0} \frac{s_n(y)}{n!} Q^n p(x).$$

Now again regarding y as a constant and x as a variable, and applying S followed by T

$$Tp(x+y) = \sum_{n\geq 0} \frac{s_n(y)}{n!} Q^n STp(x). \qquad \text{Q.E.D.}$$

COROLLARY 1. *If $s_n(x)$ is a Sheffer set relative to the invertible shift invariant operator S and the delta operator Q, then*

$$S^{-1} = \sum_{n\geq 0} \frac{s_n(0)}{n!} Q^n.$$

Proof. In the preceding theorem, set $y = 0$ and $T = S^{-1}$. This gives

$$S^{-1}p(x) = \sum_{n\geq 0} \frac{s_n(0)}{n!} Q^n p(x),$$

for any polynomial $p(x)$, which by definition is the same as saying that

$$S^{-1} = \sum_{n\geq 0} \frac{s_n(0)}{n!} Q^n. \qquad \text{Q.E.D.}$$

The defining property of polynomial sequences of binomial type has the following analog for Sheffer polynomials.

PROPOSITION 2 (Binomial Theorem). *Let Q be a delta operator with basic* *polynomials $q_n(x)$, and let $s_n(x)$ be a Sheffer set relative to Q and to some invertible shift-invariant operator S. Then the following identity holds*

$$s_n(x+y) = \sum_{k\geq 0} \binom{n}{k} s_k(x)\, q_{n-k}(y).$$

Proof. Since $q_n(x)$ is of binomial type we have by definition

$$\sum_{k\geq 0} \binom{n}{k} q_k(x)\, q_{n-k}(y) = q_n(x+y).$$

Apply S^{-1} to both sides, where, of course, x is the variable, to obtain

$$\begin{aligned}\sum_{k\geq 0} \binom{n}{k} s_k(x)\, q_{n-k}(y) &= S^{-1}q_n(x+y)\\ &= S^{-1}E^y q_n(x) = E^y S^{-1} q_n(x) = E^y s_n(x)\\ &= s_n(x+y).\end{aligned}$$

Q.E.D.

We next show that $s_n(x)$ are completely determined by their constant terms:

COROLLARY 1. *Let the polynomials $q_n(x)$ and $s_n(x)$ be defined as in Proposition 2. Then*

$$s_n(x) = \sum_{k\geq 0} \binom{n}{k} s_k(0)\, q_{n-k}(x).$$

Proof. Immediate from Proposition 2 upon setting $x = 0$.
The following converse of the second expansion theorem is useful.

PROPOSITION 3. *Let T be an invertible shift-invariant operator, let Q be a delta operator, and let $s_n(x)$ be a polynomial sequence. Suppose that*

$$E^a f(x) = \sum_{n\geq 0} \frac{s_n(a)}{n!} Q^n T f(x)$$

for all polynomials $f(x)$ and all constants a. Then the set $s_n(x)$ is the Sheffer set of the operator T relative to the delta operator Q.

Proof. Operating with T^{-1} and then with T after permuting variables, as we have already repeatedly done, we can recast the previous identity in the form

$$E^a f(x) = \sum_{n\geq 0} \frac{T s_n(a)}{n!} Q^n f(x);$$

whereupon, setting $f(x) = p_i(x)$, where $p_i(x)$ is the basic set of Q, we obtain

$$p_i(x + a) = \sum_{n\geqslant 0} \binom{i}{n} Ts_n(a)\, p_{i-n}(x),$$

and setting $x = 0$, this yields $p_i(a) = Ts_i(a)$ for all a. Q.E.D.

As an application, we obtain a simpler proof of Rodrigues' formula for basic polynomials (Proposition 4):

PROPOSITION 4. *Let $p_n(x)$ be the basic set for the delta operator Q. Then*

$$p_n(x) = x(Q')^{-1} p_{n-1}(x),$$

where Q' is the Pincherle derivative of Q.

Proof. From the first expansion theorem we have

$$E^a = \sum_{n\geqslant 0} \frac{p_n(a)}{n!} Q^n,$$

and taking the Pincherle derivative of both sides,

$$aE^a = \sum_{n\geqslant 0} \frac{p_{n+1}(a)}{n!} Q^n Q'.$$

By the preceding proposition, the polynomial set $x^{-1}p_{n+1}(x)$, $n \geqslant 0$, is the Sheffer set for the invertible shift-invariant operator Q' relative to the delta operator Q, as desired.

Next, using the notion of indicator developed in Section 4, we derive the generating function for the Sheffer polynomials.

PROPOSITION 5. *Let Q be a delta operator, and let S be an invertible shift-invariant operator. Let $s(t)$ and $q(t)$ be the indicators of S and Q, and let $q^{-1}(t)$ be the formal power series inverse to $q(t)$.*

Then the generating function for the sequence $s_n(x)$ is given by

$$\frac{1}{s(q^{-1}(t))} e^{xq^{-1}(t)} = \sum_{n\geqslant 0} \frac{s_n(x)}{n!} t^n.$$

Proof. From the proof of the first expansion theorem,

$$E^x = \sum_{n\geqslant 0} \frac{q_n(x)}{n!} Q^n, \qquad \text{and} \qquad S^{-1}E^x = \sum_{n\geqslant 0} \frac{s_n(x)}{n!} Q^n.$$

Also, since x^n is the basic set for the delta operator D, we have after a change of variable

$$E^x = \sum_{n \geqslant 0} \frac{x^n}{n!} D^n,$$

and consequently the indicator of E^x relative to D is e^{xt}. By the isomorphism theorem we may pass to indicators in the expansion for $S^{-1}E^x$ thereby obtaining

$$\frac{1}{s(t)} e^{xt} = \sum_{n \geqslant 0} \frac{s_n(x)}{n!} (q(t))^n.$$

Now set $u = q(t)$ and replace u by t to obtain the conclusion.

As a further consequence of Proposition 3, we have the following characterization of Sheffer polynomials by binomial identities.

PROPOSITION 6. *A sequence $s_n(x)$ is a Sheffer set relative to a basic set $q_n(x)$ if and only if*

$$s_n(x + y) = \sum_{k \geqslant 0} \binom{n}{k} s_k(x)\, q_{n-k}(y).$$

6. RECURRENCE FORMULAS

Given a set of polynomials $p_n(x)$, with $p_0(x) = 1$, under what conditions are they Sheffer polynomials? A simple answer is given by

PROPOSITION 1. *Let $p_n(x)$ be a polynomial sequence with $p_0(x) = 1$. If $p_n(x)$ is a Sheffer set then for every delta operator A there exists a sequence of constants s_n such that*

$$Ap_n(x) = \sum_{k \geqslant 0} \binom{n}{k} p_k(x)\, s_{n-k}, \qquad n \geqslant 0. \tag{*}$$

Also, if (*) holds for some delta operator A and some sequence s_n, then $p_n(x)$ is a Sheffer set.

Note that A need not be the delta operator associated with the set $p_n(x)$.

Proof. Assume that there exists a delta operator A and a sequence of numbers s_n so that (*) holds. We wish to show that $p_n(x)$ is a Sheffer set associated with some delta operator Q.

Define the linear operator Q by

$$Qp_n(x) = np_{n-1}(x), \qquad n > 0$$
$$Qp_0(x) = 0.$$

To prove that Q is a delta operator we need only show it is shift invariant. First note that $AQ = QA$ since

$$QAp_n(x) = Q \sum_{k \geqslant 0} \binom{n}{k} p_{n-k}(x)\, s_k$$

$$= \sum_{k \geqslant 0} \binom{n}{k} (n-k)\, p_{n-k-1}(x)\, s_k$$

$$= n \sum_{k \geqslant 0} \binom{n-1}{k} p_{n-k-1}(x)\, s_k = nAp_{n-1}(x) = AQp_n(x),$$

where we have used the identity

$$(n-k)\binom{n}{k} = n\binom{n-1}{k}.$$

The next to last equality is, by definition of the operator Q, the recurrence formula (*) with $n - 1$ in the place of n. Thus, $AQp_n(x) = QAp_n(x)$ for all n; by the familiar linearity argument, this implies $AQ = QA$, whence $A^kQ = QA^k$ for all positive integers k, and finally by the First Expansion Theorem that Q is shift-invariant. Thus, $p_n(x)$ is a Sheffer set associated with the delta operator Q.

To prove the converse, let $p_n(x)$ be a Sheffer set relative to the delta operator Q with basic set $q_n(x)$, and let A be an arbitrary delta operator. By the isomorphism theorem (see also Proposition 4 of Section 4) it is easily shown that an invertible shift-invariant operator R exists with the property that $Q = AR$. From this, the proof is concluded as follows. By the binomial theorem (Proposition 6 of the preceding section) we have

$$p_n(x+y) = \sum_{k \geqslant 0} \binom{n}{k} p_k(x)\, q_{n-k}(y).$$

Apply $Q = AR$ to both sides, recalling that y is a parameter, and obtain

$$(ARp_n)(x+y) = \sum_{k \geqslant 0} \binom{n}{k} ARp_k(x)\, q_{n-k}(y).$$

Now interchange x and y, as we may since the left side is symmetric in x and y, and then operate with the operator R^{-1}. This gives

$$Ap_n(x+y) = \sum_{k\geq 0} \binom{n}{k} ARp_k(y)\, R^{-1}q_{n-k}(x).$$

Again permute x with y, and recall that $ARp_k(x) = kp_{k-1}(x)$. The right side, therefore, equals

$$\sum_{k\geq 0} \binom{n}{k} kp_{k-1}(x)\, R^{-1}q_{n-k}(y).$$

Setting $y = 0$ gives

$$Ap_n(x) = \sum_{k\geq 1} \binom{n}{k-1} p_{k-1}(x)\, [R^{-1}q_{n-k}(y)]_{y=0}\, (n-k+1).$$

Defining

$$[R^{-1}q_{k-1}(y)]_{y=0}\,(k) = s_k \qquad \text{and} \qquad s_0 = 0,$$

we find

$$Ap_n(x) = \sum_{k\geq 0} \binom{n}{k} p_k(x)\, s_{n-k}\,.$$

Q.E.D.

7. Umbral Composition

In its most primitive form, umbral notation, or symbolic notation as it was called by invariant theorists in the past century, is an algorithmic device for treating a sequence $a_1, a_2, a_3, \ldots$ as a sequence of powers $a, a^2, a^3, \ldots$. Computationally, the technique turned out to be very effective in the hands of Blissard (after whom the device is sometimes named), Bell, and above all Sylvester, to name only a few. Several authors attempted to set the "calculus," as it somewhat improperly came to be called, on a rigorous foundation; the last unsuccessful attempt is Bell's paper of 1941. The present author observed in 1964 (in "The Number of Partitions of a Set") that all the mystery of the umbral calculus disappears, if we only consider a sequence a_n as defined by a *linear functional* on the space of polynomials: $a_n = L(x^n)$. The description of the sequence is then condensed into the properties of the linear functional L; only a prejudice would prevent anyone from placing such a definition of a sequence a_n on a par with a definition by recurrence or by generating function. In fact, the success of the umbral notation shows that in many cases the definition by a linear functional is preferable.

If $a_n(x)$ is a polynomial sequence, then there is a unique linear operator L on $\mathbf{P}$ such that $L(x^n) = a_n(x)$. We say that L is the *umbral representation* of the sequence $a_n(x)$.

We develop the umbral device in a form leading to a general result which embodies some of the more recondite indentities satisfied by special polynomials.

An *umbral operator* is an operator T which maps some basic sequence $p_n(x)$ into another basic sequence $q_n(x)$, that is, $Tp_n(x) = q_n(x)$. Note that an umbral operator is in general not shift-invariant. To motivate this definition, we require another definition, the *umbral composition* of two polynomial sequences:

$$a_n(x) = \sum_{k=0}^{n} a_{nk} x^k$$

and $b_n(x)$. This is the sequence of polynomials $c_n(x)$ defined by

$$c_n(x) = \sum_{k=0}^{n} a_{nk} b_k(x).$$

We use for umbral composition the notation

$$c_n(x) = a_n(\mathbf{b}(x)).$$

When $a_n(x) = x^n$, we simply write

$$c_n(x) = \mathbf{b}(x)^n.$$

There is a simple (though, if we are to judge by historical standards, not obvious) connection between umbral operators and the umbral composition of basic polynomials. For if T maps x^n to $q_n(x)$, then

$$a_n(\mathbf{q}(x)) = Ta_n(x),$$

so that umbral composition of polynomials is simply the application of umbral operators, and conversely.

Umbral composition of polynomials has been widely used; our present objective is to study the umbral composition of Sheffer and basic polynomials, thereby "explaining" a great many formulas from the intricate literature on special polynomials and mechanizing the device for guessing and proving them.

A simple instance of the use of umbral notation is the definition of a polynomial sequence of binomial type, which can be umbrally stated as

$$\mathbf{p}(x + y)^n = [\mathbf{p}(x) + \mathbf{p}(y)]^n;$$

similarly, the binomial property of Sheffer polynomials becomes

$$\mathbf{s}(x+y)^n = [\mathbf{p}(x) + \mathbf{s}(y)]^n.$$

PROPOSITION 1. *Let T be an umbral operator. Then T^{-1} exists and*

(a) *the map $S \to TST^{-1}$ is an automorphism of the algebra Σ of shift-invariant operators;*

(b) *T maps every sequence of basic polynomials into a sequence of basic polynomials;*

(c) *if Q is a delta operator, then $P = TQT^{-1}$ is also a delta operator;*

(d) *T maps every Sheffer set into a Sheffer set;*

(e) *If $S = s(Q)$, where $s(t)$ is a formal power series, then $TST^{-1} = s(P)$, where P is as in* (c).

Proof. $Tp_n(x) = q_n(x)$ for two given basic sets. To prove (a) we have the string of identities:

$$TPp_n(x) = T(np_{n-1}(x)) = nTp_{n-1}(x) = nq_{n-1}(x) = Qq_n(x) = QTp_n(x)$$

and since every polynomial is a linear combination of the $p_n(x)$'s, we infer that $TPp(x) = QTp(x)$ for all polynomials $p(x)$; that is, $TP = QT$. It is clear that T is invertible, since it maps polynomials of degree n into polynomials of degree n, for all n. Hence, $TPT^{-1} = Q$; whence, $TP^nT^{-1} = Q^n$ for all $n > 0$. Let S be any shift-invariant operator and let the expansion of S in terms of P be (first expansion theorem)

$$S = \sum_{n \geqslant 0} \frac{a_n}{n!} P^n.$$

Then

$$TST^{-1} = T\left(\sum_{n\geqslant 0} \frac{a_n}{n!} P^n\right) T^{-1} = \sum_{n\geqslant 0} \frac{a_n}{n!} Q^n, \tag{I}$$

and, thus, TST^{-1} is a shift-invariant operator. Furthermore, the map $S \to TST^{-1}$ is onto since any shift-invariant operator can be expanded in terms of Q. Thus, the map is an automorphism, as claimed.

Part (c) follows upon remarking that for delta operators the constant coefficient a_0 vanishes while $a_1 \neq 0$. This also proves (e).

To prove (b), let $r_n(x)$ be a basic sequence with delta operator R.

Let $s_n(x) = Tr_n(x)$ and let $S = TRT^{-1}$. By (c), S is a delta operator. Now,

$$Ss_n(x) = TRT^{-1}s_n(x) = TRr_n(x) = nTr_{n-1}(x) = ns_{n-1}(x).$$

To complete the proof that $s_n(x)$ are the basic polynomials of S we need only show that $s_n(0) = 0$ for $n > 0$. Now we can write

$$r_n(x) = \sum_{k\geqslant 1} a_k p_k(x),$$

since $a_0 = 0$ because $r_n(0) = 0$. Hence,

$$Tr_n(x) = \sum_{k\geqslant 1} a_k q_k(x) = s_n(x)$$

so that $s_n(0) = 0$, $n > 0$, as desired.

To prove (d), let $s_n(x)$ be a Sheffer set relative to the delta operator Q, and set $t_n(x) = Ts_n(x)$ and $P = TQT^{-1}$. By (c), P is a delta operator, and trivially $Pt_n(x) = nt_{n-1}(x)$. Q.E.D.

In view of the preceding result, it follows that the umbral composition of two sequences of basic operators is again a basic sequence. A similar phenomenon holds for Sheffer sets.

PROPOSITION 2. *Let $Wr_n(x) = s_n(x)$, where both are Sheffer sets. Then $W = S^{-1}TR$, where R and S are the invertible operators of $r_n(x)$ and $s_n(x)$ and where T is the umbral operator mapping the basic set $p_n(x)$ of $r_n(x)$ to the basic set of $q_n(x)$ of $s_n(x)$.*

Proof. Obvious.

COROLLARY. *The umbral composition of two Sheffer sets is a Sheffer set.*

The next result determines the operators corresponding to umbral composition.

THEOREM 7 (Umbral Composition). *Let $s_n(x)$ and $t_n(x)$ be Sheffer sets relative to the delta operators Q and P, and to the invertible shift-invariant operators S and T, respectively. Let $q_n(x)$ and $p_n(x)$ be the basic sets for Q and P, and let the indicators of S, Q, and P be*

$$S = s(D), \qquad Q = q(D), \qquad P = p(D),$$

where $s(t)$, $q(t)$ and $p(t)$ are formal power series. Define $r_n(x)$ to be the umbral composition of $s_n(x)$ and $t_n(x)$, in symbols

$$r_n(x) = s_n(\mathbf{t}(x)).$$

Then $r_n(x)$ is a Sheffer set relative to the shift-invariant operator

$$Ts(P) = t(D)\, s(p(D))$$

and the delta operator

$$q(p(D)),$$

having as basic set the sequence

$$q_n(\mathbf{p}(x)).$$

Proof. We begin by establishing the special case where S and T are the identity operators, so that we wish to find the delta operator of the sequence $u_n(x) = q_n(\mathbf{p}(x))$, which we know to be a basic sequence by Proposition 1. Thus, let $V: x^n \to p_n(x)$ be an umbral operator. Then $u_n(x) = Vq_n(x)$, and by (c) of Proposition 1 the delta operator VQV^{-1} of $u_n(x)$ is of the form $q(P) = q(p(D))$ as desired. Next, suppose that T is the identity operator, but not S. We study the sequence $s_n(\mathbf{p}(x))$. But

$$s_n(\mathbf{p}(x)) = Vs_n(x) = VS^{-1}q_n(x), \qquad (*)$$

and from $Vq_n(x) = q_n(\mathbf{p}(x))$ we infer that $q_n(x) = V^{-1}q_n(\mathbf{p}(x))$, so that, substituting in (*), we obtain

$$s_n(\mathbf{p}(x)) = VS^{-1}V^{-1}q_n(\mathbf{p}(x)) = VS^{-1}V^{-1}u_n(x).$$

This proves that it is a Sheffer sequence relative to the basic set $u_n(x)$ and the shift-invariant operator VSV^{-1}; and $VQV^{-1} = q(p(D))$, $VSV^{-1} = s(p(D))$, as follows from part (e) of Proposition 1.

Now to the general case, S and T arbitrary. By definition we have

$$t_n(x) = T^{-1}p_n(x), \qquad \text{and} \qquad r_n(x) = T^{-1}s_n(\mathbf{p}(x));$$

thus, we are reduced to the previous case, and the proof is complete.

Several special cases of the preceding theorems are worth stating. A Sheffer set relative to the delta operator D, namely, ordinary differentiation, is called an *Appell set.* The theory of Appell sets is quite old, in fact classical enough to be included in Bourbaki.

COROLLARY 1. *If $p_n(x)$ and $q_n(x)$ are basic sets with delta operators $P = p(D)$ and $Q = q(D)$, then $p_n(\mathbf{q}(x))$ is a basic set with delta operator $p(q(D))$.*

COROLLARY 2. *If $s_n(x)$ and $t_n(x)$ are Appell sets, then $s_n(\mathbf{t}(x))$ is an Appell set with operator ST; in particular, $s_n(t(x)) = t_n(\mathbf{s}(x))$.*

COROLLARY 3. *If $r_n(x)$ is a Sheffer set, then there is a unique Sheffer set $s_n(x)$, called the inverse set, such that $r_n(\mathbf{s}(x)) = x^n$. If $p_n(x)$ and $q_n(x)$ are the corresponding basic sequences, then the basic sequence of $r_n(\mathbf{s}(x))$ is $p_n(\mathbf{q}(x))$.*

The following result gives the solution of the so-called "problem of the connection constants."

COROLLARY 4. *Given Sheffer sets, $u_n(x)$ relative to the delta operator $U = u(D)$ and the invertible operator $W = w(D)$, and $t_n(x)$ as in Theorem 7, the constants s_{nk} such that*

$$\sum_{k=0}^{n} s_{nk} t_k(x) = u_n(x), \qquad n = 0, 1,...$$

are uniquely determined as follows. The polynomial sequence,

$$s_n(x) = \sum_{k=0}^{n} s_{nk} x^k,$$

is the Sheffer set with delta operator $u(p^{-1}(D))$ and invertible operator $w(p^{-1}(D))/t(p^{-1}(D))$.

The following result gives one of several closed-formula expressions for the coefficients of the Sheffer polynomials.

COROLLARY 5. *Let $s_n(x)$ be Sheffer polynomials as in Theorem 7 and let V be an umbral operator such that $Vs_n(x) = u_n(x)$ and $V^{-1}s_n(x) = v_n(x)$. Then*

$$s_n(x) = \sum_{k=0}^{n} \frac{v_k(x)}{k!} [SQ^k u_n(x)]_{x=0}.$$

Proof. By the second expansion theorem we have

$$Vs_n(x + y) = \sum_{k\geqslant 0} \frac{s_k(x)}{k!} [SQ^k Vs_n(y)];$$

setting $y = 0$ and applying the operator V^{-1} to both sides the result follows.

The following special case is useful.

COROLLARY 6. *Suppose $p_n(x)$ and $q_n(x)$ are the basic sequences for the delta operators P and Q, respectively. If $q_n(x)$ is inverse to $p_n(x)$, then*

$$p_n(x) = \sum_{k\geqslant 0} \frac{x^k}{k!} [Q^k x^n]_{x=0}.$$

Conversely, if the foregoing identity holds for a given delta operator Q, then the $p_n(x)$ are the basic sets for the inverse operator.

COROLLARY 7 (Summation Formula). *Let $f(x)$ be any polynomial. Then, in the notation of the preceding corollary.*

$$f(\mathbf{p}(x)) = \sum_{k\geqslant 0} \frac{x^k}{k!} [Q^k f(x)]_{x=0} .$$

The prototype of this formula is the classical formula of Dobinsky for the exponential polynomials (see III).

PROPOSITION 3. *Let $W: p_n(x) \to x^n$ be an umbral operator, and let Q be the delta operator of $p_n(x)$. Then*

$$Wxp(x) = xWQ'p(x)$$

for all polynomials $p(x)$, or $W' = xW(Q' - I)$.

Proof. Set $r_n(x) = (Q')^{-1} p_n(x)$, so that $xr_n(x) = p_{n+1}(x)$ by Theorem 4. Now, $Wxr_n(x) = x^{n+1} = xWp_n(x)$, so that

$$Wx(Q')^{-1} p_n(x) = xWp_n(x).$$

By linearity, this holds for all polynomials $p(x)$;

$$Wx(Q')^{-1} p(x) = xWp(x),$$

replacing $p(x)$ by $Q'p(x)$ the result follows.

It would be of interest to develop a theory of operator differential equations in the Pincherle derivative strong enough to give an explicit solution to the previous "differential equation" for the umbral operator W. An example of umbral operator is $Wp_k(x) = a^k p_k(x)$, which is a Sheffer set whenever $p_k(x)$ is. If Q is the delta operator of $p_k(x)$, then $a^{-1}Q$ is the delta operator of $a^k p_k(x)$. Similarly, $p_k(ax)$ is a Sheffer set, and if $Q = f(D)$, then the delta operator for $p_k(ax)$ is $f(a^{-1}D)$. Finally, if $q_k(x)$ is a basic set, then the basic set of the delta operator QE^a is easily seen from formula (4) of Theorem 4 to be $r_n(x) = xq_n(x - na)/(x - na)$. This generalizes the idea behind the Abel polynomials. Summarizing, we have the following.

PROPOSITION 4. *If $s_n(x)$ is a Sheffer set, so is $a^n s_n(bx)$ for any a and b; if it is a basic set, so are $a^n s_n(bx)$ and $xs_n(x - na)/(x - na)$.*

The preceding result "explains" the so-called "duplication formulas" found in the literature, namely, formulas expressing $p_n(ax)$ as a linear combination of $p_k(x)$. We shall see some instances of this device later.

8. Cross-Sequences

A *cross-sequence* of polynomials, written $p_n^{[\lambda]}(x)$, where λ ranges over the field and n over the nonnegative integers, is defined by the following properties:

(a) for fixed λ, $p_n^{[\lambda]}(x)$ is a polynomial sequence;

(b) for any λ and μ in the field and any x and y, the identity,

$$p_n^{[\lambda+\mu]}(x+y) = \sum_{k=0}^{n} \binom{n}{k} p_k^{[\lambda]}(x)\, p_{n-k}^{[\mu]}(y), \tag{*}$$

holds for all n.

The theory of cross-sequences (of which several examples are unconsciously present in the literature) parallels in many ways the theory of sequences of binomial type, and we shall shorten the by now familiar devices in the proofs. It will always be assumed that the upper variable ranges over the field and the lower one over the nonnegative integers.

THEOREM 8. *A sequence $p_n^{[\lambda]}(x)$ is a cross-sequence if and only if there exists a one-parameter group $P^{-\lambda}$ of shift-invariant operators and a sequence $p_n(x)$ of binomial type such that*

$$p_n^{[\lambda]}(x) = P^{-\lambda} p_n(x). \tag{**}$$

(*Thus, for fixed λ a cross-sequence becomes a Sheffer sequence relative to the operator P^{λ}.*)

Proof. We first show that every sequence defined by the right side of (**) is a cross-sequence. Recall that the group property states that

$$P^{-(\lambda+\mu)} = P^{-\lambda} P^{-\mu}.$$

Thus, apply $P^{-\lambda}$ to the binomial identity satisfied by the $p_n(x)$, thereby obtaining

$$P^{-\lambda} p_n(x+y) = \sum_{k=0}^{n} \binom{n}{k} p_k^{[\lambda]}(x)\, p_{n-k}(y).$$

Now permute x and y, and then apply $P^{-\mu}$ to both sides, to obtain (*). Now to the converse. First, note that the sequence $p_n(x) = p_n^{[0]}(x)$ is of binomial type; setting $\mu = 0$ in (*) and applying Proposition 3 of Section 5, we infer that $p_n^{[\lambda]}(x)$ is a Sheffer set relative to a shift-invariant operator which we shall call P^λ, as in (**). From (*) we have

$$P^{-\lambda}p_n(x+y) = \sum_{k=0}^{n} \binom{n}{k} p_k(x)\, p_{n-k}^{[\lambda]}(y),$$

and applying $P^{-\mu}$ to both sides, we infer that

$$P^{-\mu}(P^{-\lambda}p_n(x+y)) = \sum_{k=0}^{n} \binom{n}{k} p_k^{[\mu]}(x)\, p_{n-k}^{[\lambda]}(y).$$

But the right side equals $P^{-\lambda-\mu}p_n(x+y)$, again by (*). This gives $P^{-\mu}P^{-\lambda} = P^{-\mu-\lambda}$ and completes this proof.

COROLLARY. *If a sequence $p_n^{[\lambda]}(x)$ is a cross-sequence, then there exists delta operators Q and $R^{[\alpha]}$ such that $p_0^{[0]} = c \neq 0$,*

$$\begin{aligned} p_n^{[0]}(0) &= 0, && n > 0 \\ Qp_n^{[\lambda]}(x) &= np_{n-1}^{[\lambda]}(x), && n \geqslant 1, \qquad (***) \\ R^{[\alpha]}p_n^{[\lambda]}(x) &= np_{n-1}^{[\lambda-\alpha]}(x). \end{aligned}$$

Proof. Let Q be the delta operator of $p_n(x)$, and let $R^{[\alpha]} = P^\alpha Q$; then (***) follows from (**).

PROPOSITION 1. *The coefficients $c(n, k, \lambda)$ of a cross-sequence,*

$$P^{-\lambda}p_n(x) = p_n^{[\lambda]}(x) = \sum_{k\geqslant 0} c(n, k, \lambda)\, x^k,$$

are polynomials of degree at most n in the variable λ.

Proof. By Corollary 7 of Theorem 7 we have

$$p_n^{[\lambda]}(x) = \sum_{k\geqslant 0} \frac{x^k}{k!} [P^{-\lambda}Q^k x^n]_{x=0},$$

where Q is the delta operator of the *inverse* of this sequence $p_n(x)$.

Writing $P^{-\lambda} = p(D)^\lambda$ and $q(D) = Q$, we have

$$k!c(n, k, \lambda) = [P^{-\lambda}Q^k x^n]_{x=0} = [D^n p(x)^\lambda\, q(x)^k]_{x=0},$$

whence the conclusion.

The proof does not provide an explicit method for the computation of the coefficients $c(n, k, \lambda)$, but see Proposition 4.

A *Steffensen sequence* $s_n^{[\lambda]}(x)$ relative to a cross-sequence $p_n^{[\lambda]}(x)$, is a sequence satisfying the identities

$$s_n^{[\lambda+\mu]}(x+y) = \sum_{k\geq 0} \binom{n}{k} s_k^{[\lambda]}(x)\, p_{n-k}^{[\mu]}(y),$$

for all n, λ, μ, x, y: Steffensen sequences are characterized by

PROPOSITION 2. *The following conditions are equivalent:*

(a) $s_n^{[\lambda]}(x)$ *is a Steffensen sequence;*

(b) *there exists a delta operator Q and a one-parameter group of shift-invariant operators $P^{-\lambda}$ such that*

$$Qs_n^{[\lambda]}(x) = ns_{n-1}^{[\lambda]}(x),$$

$$P^{-\mu}s_n^{[\lambda]}(x) = s_n^{[\lambda+\mu]}(x);$$

(c) *There exists a cross-sequence $p_n^{[\lambda]}(x)$ and an invertible shift-invariant operator T such that*

$$s_n^{[\lambda]}(x) = T^{-1}p_n^{[\lambda]}(x).$$

The proof follows well trodden paths and is omitted.

PROPOSITION 3. *Let $s_n^{[\lambda]}(x)$ be a Steffensen sequence relative to a shift-invariant operator $T = (Q')^{-1}$, as in the preceding proposition, with $s_0^{[\lambda]}(0) = 1$ for all λ. Then the sequence*

$$xs_n^{[n+1]}(x),$$

is a sequence of binomial type.

Proof. Use Theorem 4. $p_n(x) = xs_n^{[0]}(x)$ is the basic sequence for the operator Q, by the Rodrigues formula.

Writing

$$xs_n^{[n+1]}(x) = xP^{-n-1}x^{-1}p_{n+1}(x)$$

and comparing with (6) of the corollary to Theorem 4, we find that the right side is basic with delta operator $R = PQ$.

PROPOSITION 4. *Suppose that $I - P = Q$, where Q is the delta operator of $p_n(x)$. Then for fixed a and for a Steffensen sequence $p_n^{[\lambda]}(x)$ relative to Q we have that*

$$p_n^{[x-n]}(a) \tag{*}$$

is, for fixed a, a Sheffer sequence relative to the difference operator $\Delta = E - I$.

Proof. We have

$$\begin{aligned} p_n^{[\lambda+1-n]}(x) &- p_n^{[\lambda-n]}(x) \\ &= P^{-\lambda+n-1}(I - P)\, p_n(x) = nP^{-\lambda+n-1}p_{n-1}(x) \\ &= np_{n-1}^{[\lambda-n+1]}(x), \end{aligned}$$

which proves the assertion.

It follows from Corollary 1 to Proposition 2 of Section 5 that any linear combination of polynomials of the form (*) is again a Sheffer set relative to Δ. In particular, the coefficients $c(n, k, \lambda)$ (polynomials, by Proposition 1) of

$$p_n^{[\lambda]}(x) = \sum_{k\geqslant 0} \frac{c(n, k, \lambda)}{k!}\, x^k,$$

have the remarkable property that $c(n, k, x - n)$ is a Sheffer set for Δ. An explicit expression could be constructed. We shall not develop in detail here the theory of umbral composition of Steffensen sets, only a few remarks.

PROPOSITION 5. *For Appell cross sequences, namely of the form $p_n^{[\lambda]}(x) = P^{-\lambda}x^n$, we have the umbral composition*

$$p_n^{[\lambda]}(\mathbf{p}^{[\mu]}(x)) = p_n^{[\lambda+\mu]}(x).$$

Proof. Apply Corollary 2 to Theorem 7.

Every invertible shift-invariant operator P can be written in the form $P = e^F$ for some shift-invariant operator (which is never invertible). Indeed, say that $P = I + S$, where $S1 = 0$. Then $F = \log(I + S)$ is well defined, and $P = e^F$. Thus,

$$P^{-\lambda} = \exp(-\lambda F).$$

Note that F is not necessarily a delta operator, though $F1 = 0$. We call F the *generator* of the cross-sequence $p_n^{[\lambda]}(x)$. Thus, an operator F is the generator of a necessarily unique cross-sequence of polynomials, if and only if $F(1) = 0$.

PROPOSITION 6. (a) *If F and G are the generators of cross-sequences $p_n^{[\lambda]}(x)$ and $q_n^{[\lambda]}(x)$ having the same basic sequence, then $F + G$ is the generator of the cross-sequence*

$$e^{-\lambda G} p_n^{[\lambda]}(x) = e^{-\lambda F} q_n^{[\lambda]}(x).$$

(b) *If P is any invertible operator, then*

$$P^{-\lambda} p_n^{[\lambda]}(x)$$

is a cross-sequence when $p_n^{[\lambda]}(x)$ is one.

9. EIGENFUNCTION EXPANSIONS

It is reasonable to surmise that a Sheffer set of polynomials over the real or complex fields should be obtainable by eigenfunction expansion of differential, difference or other Q-operators in a suitable Hilbert space. We establish the truth of this expectation in the real case. The key step consists in singling out a "natural" inner product associated with a given Sheffer set. To this end, let $s_n(x)$ be a Sheffer set relative to the invertible operator S and the delta operator Q. Let $W : s_n(x) \to x^n$ be the umbral operator sending $s_n(x)$ to x^n. For arbitrary polynomials $f(x)$ and $g(x)$ set

$$(f(x), g(x)) = [(Wf)(Q)\, Sg(x)]_{x=0}\,, \qquad (*)$$

we have then the following.

PROPOSITION 1. *The bilinear form $(f(x), g(x))$ defined by * on the vector space of all polynomials is a positive-definite inner product.*

Proof. It suffices to show that $(s_k(x), s_n(x)) = (s_n(x), s_k(x)) = 0$ for $k \neq n$, and $(s_n(x), s_n(x)) > 0$ for all n and k. Now,

$$(s_k(x), s_n(x)) = [Q^k S s_n(x)]_{x=0} = [Q^k p_n(x)]_{x=0} = (n)_k\, p_{n-k}(0) = (n)_k\, \delta_{nk}\,,$$

where $p_n(x)$ are the basic polynomials of Q. This completes the proof.

We shall call (*) the *natural inner product associated with the Sheffer set*

$s_n(x)$. We shall now require some notions of Hilbert space theory, such as one finds in any book on functional analysis.

THEOREM 9. *For any Sheffer sequence $s_n(x)$ with delta operator Q and operator S there exists a unique operator of the form*

$$A = \sum_{k\geq 1} \frac{u_k + xv_k}{(k-1)!} Q^k$$

with the following properties:

(a) *A is essentially self adjoint (and densely defined) in the Hilbert space H obtained by completing the space $\mathbf{P}$ of polynomials in the associated inner product* (*);

(b) *The spectrum of A consists of simple eigenvalues at* 0, 1, 2,...; *the eigenfunction associated with the eigenvalue n is the polynomial $s_n(x)$*;

(c) *the constants u_k and v_k in the previous expression for A are given by*

$$u_k = -\,[(\log S)'\, x^{-1} p_k(x)]_{x=0}; \qquad v_k = p_k'(0),$$

where $p_k(x)$ are the basic polynomials for the delta operator Q.

Proof. We begin by taking the Pincherle derivative of both sides in the expression

$$S^{-1}E^a = \sum_{n\geq 0} \frac{s_n(a)}{n!} Q^n,$$

obtained from the second expansion theorem:

$$(S^{-1}E^a)' = \sum_{n\geq 1} \frac{s_n(a)}{n!} nQ^{n-1}Q';$$

multiplying by $(Q')^{-1} Q$ and simplifying,

$$(-\,S^{-1}S' + a)\, S^{-1}E^a(Q')^{-1}Q = \sum_{n\geq 1} \frac{s_n(a)}{n!} nQ^n = TS^{-1}E^a, \qquad (**)$$

where we have set

$$T = (-\,S^{-1}S' + a)\,(Q')^{-1}Q = (a - (\log S)')\, Q(Q')^{-1}.$$

Next, expand the operator T in powers of Q, that is, compute the coefficients b_k in

$$T = \sum_{k\geq 0} \frac{b_k}{k!} Q^k; \qquad b_k = [Tp_k(x)]_{x=0}\,, \qquad (***)$$

as in the first expansion theorem. Set

$$q_{n-1}(x) = x^{-1}p_n(x) \qquad \text{for } n > 0.$$

Rodrigues' formula now reads

$$(Q')^{-1}\, p_n(x) = q_n(x),$$

whence

$$(Q')^{-1}\, Qp_n(x) = nq_{n-1}(x).$$

Thus, for $k = 0$ we have $b_k = 0$, and for $k > 0$

$$\begin{aligned}[Tp_k(x)]_{x=0} &= k[(a - (\log S)')\, q_{k-1}(x)]_{x=0} \\ &= kaq_{k-1}(0) - k[(\log S)'\, q_{k-1}(x)]_{x=0} = kav_k + ku_k\,,\end{aligned}$$

where

$$\begin{aligned}u_k &= -\,[S^{-1}S'q_{k-1}(x)]_{x=0} = -\,[(\log S)'\, q_{k-1}(x)]_{x=0}\,, \\ v_k &= q_{k-1}(0), \qquad k > 0.\end{aligned}$$

Now from (***) we have for any polynomial $f(x)$,

$$TS^{-1}f(x + a) = \sum_{k\geqslant 0} \frac{b_k}{k!} Q^k[S^{-1}f(x + a)].$$

But, as remarked previously,

$$S^{-1}f(x + a) = \sum_{n\geqslant 0} \frac{s_n(x)}{n!} Q^n f(a),$$

so that placing the right side into the brackets we obtain

$$\begin{aligned}TS^{-1}f(x + a) &= \sum_{k\geqslant 0} \frac{b_k}{k!} Q^k \left[\sum_{n\geqslant 0} \frac{s_n(x)}{n!} Q^n f(a)\right] \\ &= \sum_{n\geqslant 0} \left[\sum_{k\geqslant 0} \frac{b_k}{k!} Q^k s_n(x)\right] \frac{Q^n}{n!} f(a),\end{aligned}$$

where we have interchanged the order of summation. Permuting x and a once more, we obtain

$$TS^{-1}E^a = \sum_{n\geqslant 0} \left[\sum_{k\geqslant 0} \frac{b_k}{k!} Q^k s_n(a)\right] \frac{Q^n}{n!}\,,$$

and comparing this with the right side of (**), we see that the coefficients of the two expansions must agree. Upon changing a to x, we obtain

$$\sum_{k\geqslant 0} \frac{b_k}{k!} Q^k s_n(x) = n s_n(x), \qquad n \geqslant 0,$$

with

$$b_k = k(u_k + x v_k).$$

The operator

$$A = \sum_{k\geqslant 1} \frac{u_k + x v_k}{(k-1)!} Q^k$$

is clearly well defined on the set of all polynomials. We have shown that $As_n(x) = ns_n(x)$ for all $n \geqslant 0$, so that the Sheffer set $s_n(x)$ is a set of eigenfunctions of A; since it spans that Hilbert space H we infer that A is an unbounded essentially self-adjoint operator in H having the nonnegative integers as its simple spectrum, with eigenfunctions $s_n(x)$, as we wanted to show.

COROLLARY 1. *Let R be a delta operator with basic polynomials $r_k(x)$. Then the operator A defined previously can be expressed in the form*

$$A = \sum_{k\geqslant 1} \frac{a_k + x b_k}{k!} R^k,$$

with

$$a_k = -[(\log S)' Q(Q')^{-1} r_k(x)]_{x=0},$$
$$b_k = [Q(Q')^{-1} r_k(x)]_{x=0}.$$

Proof. From the preceding proof we have

$$T = \sum_{k\geqslant 0} \frac{a_k + a b_k}{k!} R^k,$$

whence the conclusion upon interchanging the roles of the variables x and a, as in the proof of Theorem 8.

The computation of the coefficients a_k and b_k is greatly simplified by use of the corollary to Theorem 4 and by various umbral devices.

The generating functions associated with the a_k and b_k are now easily found; they are immediate consequences of the isomorphism theorem:

COROLLARY 2. *Let $Q = \phi(R)$ and $S = \psi(R)$, where ϕ and ψ are formal power series. Then*

$$\sum_{k\geqslant 0} \frac{b_k}{k!} t^k = \frac{\phi(t)}{\phi'(t)} \qquad \textit{and} \qquad \sum_{k\geqslant 0} \frac{a_k}{k!} t^k = -\frac{\psi'(t)}{\psi(t)} \frac{\phi(t)}{\phi'(t)}.$$

By changes of variables, these identities can be recast in a form suitable for computation in any specific case. One question of interest is the following. When is the operator A a *polynomial* in the operator R? The answer is easily found.

COROLLARY 3. *A is a polynomial in R if and only if*

$$\phi(t) = \exp\left(\int p(t)^{-1}\, dt\right),$$

$$\psi(t) = \exp\left(\int q(t)/p(t)\, dt\right),$$

where p and q are polynomials, and $p(0) = 0$ and $p'(0) \neq 0$, as well as $q(0) = 0$.

Proof. From the preceding corollary we find the differential equations

$$\frac{\phi'(t)}{\phi(t)} = \frac{1}{p(t)}, \qquad \frac{\psi'(t)}{\psi(t)} = \frac{q(t)}{p(t)},$$

whence, integrating

$$\phi(t) = \exp\left(\int p(t)^{-1}\, dt\right), \qquad \psi(t) = \exp\left(\int q(t)/p(t)\, dt\right).$$

Now, $\phi(0) = 0$ and $\phi'(0) \neq 0$, because Q and R are delta operators; it follows that the partial fraction expansion of $1/p(t)$ must contain the summand $1/t$, and this happens only if $p(0) = 0$ and $p'(0) \neq 0$. Similarly, $\psi(0) \neq 0$ because the operator S is invertible. This requires that the partial fraction expansion of $q(t)/p(t)$ shall not contain the summand $1/t$, and, in view of $p(0) = 0$, this requires that $q(0) = 0$. Q.E.D.

Another relevant question in the present context is the representability of the inner product (*) by integral operators, evaluations of a function and its derivatives at specific points, etc. It would take us too far afield to treat this question here; suffice it to say that it can be completely answered.

The simplest case of (*) occurs when $S = I$ and $Q = D$, the ordinary derivative. We have then

$$[p(D)\, q(x)]_{x=0} = \frac{1}{\pi} \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} \overline{p(x+iy)}\, q(x+iy)\, e^{-(x^2+y^2)}\, dx\, dy,$$

an inner product of frequent occurrence in quantum field theory. From the recurrence relations for orthogonal polynomials it is easy to determine (following Sheffer) all Sheffer sets which are orthogonal polynomials over an interval of the real line. Except for linear changes of variable, they are the following:

(a) for $Q = D$, we must have $S = E^a \exp(D^2)$, and we find a generalization of the Hermite polynomials, orthogonal over $(-\infty, \infty)$;

(b) for $Q = D/(D - I)$ we must have $S = (1 - D)^{\alpha+1}$ with $\alpha > -1$, and we find the Laguerre polynomials of order α, treated later;

(c) for $Q = \log(1 + D)$ we must have $S = E^{\alpha}(I + D)^{\rho}$, $\alpha\rho \neq 0$;

(d) for $Q = \log[b(D - c)/(c(D - b))]$, then $S = (1 - D/c)^{\alpha} (1 - D/b)^{\beta}$; $b \neq c$ and $bc \neq 0$.

These are essentially the Pollaczek polynomials. A similar study can be made in the case of discrete orthogonal polynomials. The polynomials under (c) are Sheffer polynomials relative to the exponential polynomials; they seem not to have been studied. It is interesting to speculate on the possible generalizations of the notion of classical orthogonal polynomial that are suggested by the "natural" inner product (*).

10. Hermite Polynomials

We show that classical formulas pertaining to the Hermite polynomials, as found for example in Jackson or Rainville, can be obtained by specializing the preceding results. Define the Hermite polynomials *of variance v* to be the Appell set (as we shall see, the Appell cross-sequence) whose operator is the Weierstrass operator (so dubbed by Hirschman–Widder)

$$W_v p(x) = \frac{1}{(2\pi v)^{1/2}} \int_{-\infty}^{\infty} e^{-t^2/2v} p(x + t)\, dt. \qquad (*)$$

The ordinary Hermite polynomials correspond to variance one. Thus,

$$H_n^{(v)}(x) = W_v^{-1} x^n, \qquad DH_n^{(v)}(x) = nH_{n-1}^{(v)}(x),$$
$$H_n^{(v)}(x + y) = \sum_{k\geq 0} \binom{n}{k} y^{n-k} H_k^{(v)}(x), \quad \text{etc.},$$

trivially from Section 5. The indicator of the operator W_v is computed by the first expansion theorem:

$$W_v = \sum_{n\geq 0} \frac{a_n^{(v)}}{n!} D^n,$$

with

$$a_n^{(v)} = \frac{1}{(2\pi v)^{1/2}} \int_{-\infty}^{\infty} e^{-t^2/2v} t^n \, dt = \frac{v^{n/2} n!}{2^{n/2}(n/2)!} \cdot \frac{(1 + (-1)^n)}{2}$$
$$= \begin{cases} v^{n/2} \cdot 1 \cdot 3 \cdot 5 \cdots (n-1) & \text{for } n \text{ even} \\ 0 & \text{for } n \text{ odd.} \end{cases} \qquad (**)$$

We set $a_n^{(v)} = v^{n/2} b_n$. Thus,

$$W_v = \sum_{n \geqslant 0} \frac{v^n D^{2n}}{2^n \cdot n!} = e^{vD^2/2}. \qquad (***)$$

We infer that $H_n^{(v)}(x) = H_n^{[v]}(x)$ is a *cross-sequence.* Note that the definition of the Weierstrass operator by (*) is valid only for $v > 0$, but (***) always holds. Next,

$$H_n^{[v+w]}(x + y) = \sum_{k \geqslant 0} \binom{n}{k} H_k^{[v]}(x) \, H_{n-k}^{[w]}(y)$$

$$(x + y)^n = \sum_{k \geqslant 0} \binom{n}{k} H_k^{[v]}(x) \, H_{n-k}^{[-v]}(y)$$

setting $y = 0$,

$$x^{2n} = \sum_{j \geqslant 0} \binom{2n}{2j} H_{2j}^{[v]}(x) \frac{(v)^{n-j} (2n - 2j)!}{2^{n-j}(n-j)!},$$

$$x^{2n+1} = \sum_{j \geqslant 0} \binom{2n+1}{2j+1} H_{2j+1}^{[v]}(x) \frac{(v)^{n-j} (2n - 2j)!}{2^{n-j}(n-j)!},$$

and finally

$$H_n^{[v]}(x) = \sum_{k \geqslant 0} \binom{n}{k} x^{n-k} (-v)^{k/2} \, b_k,$$

where b_n are given previously; whence we glean the simpler expressions in terms of the classical Hermite polynomials

$$H_n^{[v]}(x) = v^{n/2} H_n\left(\frac{x}{(v)^{1/2}}\right),$$

as we could also have done by umbral methods.

Proposition 5 of Section 8 gives the umbral composition formula,

$$H_n^{[v]}(\mathbf{H}^{[w]}(x)) = H_n^{[v+w]}(x), \qquad (*)$$

and in particular the classical

$$H_n(\mathbf{H}(x)) = 2^{n/2} H_n\left(\frac{x}{(2)^{1/2}}\right).$$

The generating function

$$e^{-t^2/2} e^{xt} = \sum_{n\geqslant 0} \frac{H_n(x)}{n!} t^n,$$

is also immediate from Section 5, Proposition 5.

The (classical) Rodrigues formula follows using the Pincherle derivative. Starting with

$$e^{x^2/2v}(vD)\, e^{-x^2/2v} f(x) = (vD - x) f(x)$$

and

$$\begin{aligned} e^{-vD^2/2} x f(x) &= [(e^{-vD^2/2})' + x e^{-vD^2/2}] f(x) \qquad (*) \\ &= (-1)(vD - x)\, e^{-vD^2/2} f(x), \end{aligned}$$

setting $f(x) = x^{n-1}$ and iterating,

$$H_n^{(v)}(x) = (-1)^n\, e^{x^2/2v} (vD)^n\, e^{-x^2/2v},$$

as desired.

Note that this also proves the recurrence formulas, stated for $v = 1$ for convenience,

$$H_n(x) = x H_{n-1}(x) - H'_{n-1}(x) = x H_{n-1}(x) - (n-1) H_{n-2}(x),$$

from which the differential equation can be obtained by application of $H_n'(x) = n H_{n-1}(x)$ and iteration. We prefer, however, to derive the spectral theory directly from the general results of Section 9. Operational identity (*) can also be used to give a quick proof of the formulas of *Burchnall–Feldheim–Watson.* Indeed, from

$$(D - x)^n f(x) = e^{x^2/2} D^n e^{-x^2/2} f(x)$$

we find, upon applying Leibniz's formula, that the right side equals (following Burchnall)

$$e^{x^2/2} \sum_{k=0}^{n} \binom{n}{k} (D^k e^{-x^2/2})\, D^{n-k} f(x) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} H_k(x)\, D^{n-k} f(x),$$

and setting $f(x) = H_j(x)$ we find

$$H_{n+j}(x) = (-1)^n (D - x)^n H_j(x)$$
$$= \sum_{k \geqslant 0} \binom{n}{k} (-1)^{n-k} (j)_{n-k} H_{j-n+k}(x) H_k(x),$$

as desired. Similarly we can derive a formula for expressing $H_j(x) H_n(x)$ as linear combinations of $H_k(x)$ by Theorem 6.

We find that

$$p(x) = \sum_{n \geqslant 0} \frac{H_n(x)}{n!} [D^n W_1 p(y)]_{y=0}$$

for any polynomial $p(t)$. Now

$$(D^n W_1) H_j(x) H_k(x) = W_1(D^n(H_j(x) H_k(x)))$$
$$= W_1 \left(\sum_{i=0}^{n} \binom{n}{i} (j)_i H_{j-i}(x) (k)_{n-i} H_{k-n+i}(x) \right)$$
$$= \sum_{i=0}^{n} \binom{n}{i} (j)_i (k)_{n-i} W_1(H_{j-i}(x) H_{k-n+i}(x)).$$

Now (v. below)

$$[W_1 H_r(x) H_s(x)]_{x=0} = [H_r(x), H_s(x)]_1$$
$$= r! \, \delta_{rs} .$$

Therefore, if $j \leqslant k$ say, then

$$[(D^n W_1) H_j(x) H_k(x)]_{x=0}$$
$$= \begin{cases} \binom{n}{i} j! \, (k)_{(k-j+n)/2} & \text{if} \quad n \equiv k + j \, (2), \quad i = (j + n - k)/2, \\ & \qquad 0 \leqslant i \leqslant j \\ 0 & \text{otherwise,} \end{cases}$$

and we conclude that

$$H_j(x) H_k(x) = \sum_{\substack{n \geqslant 0 \\ n \equiv (k+j) (\mathrm{mod} 2) \\ n \leqslant k+j \\ n \geqslant k-j}} H_n(x) \, j! k! \frac{1}{\left(\frac{n + j - k}{2}\right)! \left(\frac{n + k - j}{2}\right)! \left(\frac{k + j - n}{2}\right)!} .$$

Proposition 1 of Section 9 shows that the Hermite polynomials are orthogonal relative to the inner product

$$(f(x), g(x))_v = [(W_v f)(D)\, W_v g(x)]_{x=0}\,.$$

We next find out when this inner product coincides with the classical inner product

$$[f(x), g(x)]_v = \frac{1}{(2\pi v)^{1/2}} \int_{-\infty}^{\infty} e^{-x^2/2v} f(x)\, g(x)\, dx, \qquad v > 0. \qquad (\dagger)$$

By Rodrigues' formula, followed by an integration by parts, we find

$$\begin{aligned}
[H_n^{(v)}(x), g(x)]_v &= \frac{v^n}{(2\pi v)^{1/2}} \int_{\infty}^{\infty} e^{-x^2/2v} D^n g(x)\, dx \\
&= \frac{v^n}{(2\pi v)^{1/2}} \left[\int_{-\infty}^{\infty} e^{-t^2/2v} D^n g(x+t)\, dt\right]_{x=0} \\
&= v^n [D^n W_v g(x)]_{x=0} \\
&= [W_v((W_v T_v H_n)(D)\, g(x))]_{x=0}\,,
\end{aligned}$$

where $T_v: f(x) \to f(xv)$ is an umbral operator. By linearity it follows that

$$[f(x), g(x)]_v = [W_v((W_v T_v f)(D)\, g(x))]_{x=0}\,, \qquad (37)$$

for all polynomials f and g. On the other hand, we verify upon replacing f and g by Hermite polynomials that

$$(f(x), g(x))_v = [W_v((W_v f)(D)\, g(x))]_{x=0}\,, \qquad (38)$$

so that the two inner products coincide only for $v = 1$. Both inner products, however, are symmetric and nondegenerate for all values of v; for (38) this is true by definition, and for (37) it is verified as follows. Setting

$$f(x) = H_n^{(v)}(x)$$

we find

$$[H_n^{(v)}(x), g(x)]_v = [W_v((vD)^n\, g(x))]_{x=0}\,,$$

and for $g(x) = H_k(x)$ this becomes

$$\begin{aligned}
[H_n^{(v)}(x), H_k^{(v)}(x)]_v &= [W_v((vD)^n\, H_k^{(v)}(x))]_{x=0} \\
&= v^n (k)_n\, [W_v H_{k-n}^{(v)}(x)]_{x=0} \\
&= v^n (k)_n\, [x^{k-n}]_{x=0} \\
&= v^n n!\, \delta_{kn}\,,
\end{aligned}$$

as desired. For $v > 0$ this inner product is positive-definite. However, definition (37) is valid for *arbitrary* v and combined with the results of Section 9 gives a formally valid eigenfunction expansion, whose inner product is nondegenerate but not positive definite in general. On the other hand, the positive-definite inner product (38), as defined in Section 9, gives a Hilbert-space eigenfunction expansion for arbitrary v. The interaction of the two bilinear forms for nonpositive v leads to interesting analytic developments which we are forced to leave to a later publication. There are also interesting applications to Feynman's integral. There remains to be found the operator of which the Hermite polynomials are the eigenfunctions, and this is given at once by Theorem 9. We have $(\log S)' = D$, since $S = W_1$, so that the formulas given there yield $u_2 = -1$, $v_1 = 1$ and all other coefficients 0. We conclude that the Hermite polynomials are a complete sequence of eigenfunctions, with eigenvalues n, of the operator

$$A = D^2 - xD$$

in the Hilbert space which is the closure of the polynomials in (†). That such a closure is the set of all square-integrable functions follows from a (well known) limiting argument. The present treatment shows that, aside from this one fact from analysis, the entire theory of Hermite expansions can be made purely algebraic.

11. Laguerre Polynomials

One of the simplest cross-sequences is

$$M_n^{[\lambda]}(x) = (I - D)^{-\lambda} x^n,$$

or, more explicitly,

$$M_n^{[\lambda]}(x) = \sum_{k \geqslant 0} \binom{n}{k} (\lambda + k - 1)_k \, x^{n-k}.$$

These polynomials seem to have a scarce literature. For $\lambda = 1$ they were considered by Sheffer, with D replaced by $D/2$ they were studied by Peters under the name "Boole polynomials of the second kind." Note that for $\lambda = 1$ they give, after dividing by $n!$, the partial sums of the exponential function.

From the properties of cross-sequences we immediately infer that

$$M_n^{[\lambda+\mu]}(x) = \sum_{k \geqslant 0} \binom{n}{k} (\lambda + k - 1)_k \, M_{n-k}^{[\mu]}(x),$$

as well as

$$M_n^{[\lambda+\mu]}(x+y) = \sum_{k\geq 0} \binom{n}{k} M_k^{[\lambda]}(x)\, M_{n-k}^{[\mu]}(y),$$

which explains several classical binomial identities. Moreover, since the $M_n^{[\alpha]}(x)$ are an Appell set, Corollary 2 to Theorem 7 implies the composition law

$$M_n^{[\alpha]}(\mathbf{M}^{[\beta]}(x)) = M_n^{[\beta]}(\mathbf{M}^{[\alpha]}(x)) = M_n^{[\alpha+\beta]}(x).$$

The cross-sequence $M_n^{[\alpha]}(x)$ is related to polynomials of Laguerre type, which are the Sheffer sets relative to the delta operator

$$Kf(x) = -\int_0^\infty e^{-t} f'(x+t)\, dt,$$

called the *Laguerre operators*. From the first expansion theorem we have

$$K = \sum_{n\geq 1} \frac{a_n}{n!} D^n; \qquad a_n = -n \int_0^\infty e^{-t} t^{n-1}\, dt = -n!,$$

so that

$$K = -D - D^2 - \cdots = D/(D-I).$$

The basic polynomials of the Laguerre operator are easily computed from Theorem 4, formula (3):

$$L_n(x) = x(D-I)^n x^{n-1}, \qquad (*)$$

called the *basic* Laguerre polynomials. From

$$e^x D e^{-x} = D - I \qquad \text{and} \qquad e^x D^n e^{-x} = (D-I)^n,$$

we obtain the classical *Rodrigues formula*,

$$L_n(x) = x e^x D^n e^{-x} x^{n-1}.$$

From formula (*) we find by binomial expansion that

$$L_n(x) = \sum_{k=1}^{n} \frac{n!}{k!} \binom{n-1}{k-1} (-x)^k,$$

where the coefficients

$$\frac{n!}{k!} \binom{n-1}{k-1}$$

are known as the (signless) *Lah numbers*.

We shall be concerned with Laguerre type sets relative to the operators (Laguerre operators of order α):

$$K_\alpha = I/(I - D)^{\alpha+1}.$$

Let us note here that for $\alpha > -1$,

$$K_\alpha f(x) = \frac{1}{\Gamma(\alpha + 1)} \int_0^\infty t^\alpha e^{-t} f(x + t)\, dt,$$

as is easily verified by the first expansion theorem. The Sheffer sets relative to these operators are polynomial sets $L_n^{[\alpha]}(x)$, classically known as *Laguerre polynomials of order* α. (Note that our definition of Laguerre polynomials differs from that used by many authors by a factor of $n!$. It does, however, agree with Jackson's notation.)

Again, by definition of the Sheffer polynomials we have

$$L_n^{(\alpha)}(x) = (I - D)^{\alpha+1} L_n(x),$$
$$(1 - D)^\beta L_n^{(\alpha)}(x) = L_n^{(\alpha+\beta)}(x).$$

We infer from (*) the identity

$$L_n^{(\alpha)}(x) = (I - D)^{\alpha+1}\, x(D - I)^n\, x^{n-1}.$$

Using the Pincherle derivative identity

$$(D - I)^n\, x - x(D - I)^n = ((D - I)^n)' = n(D - I)^{n-1},$$

we simplify this expression to

$$\begin{aligned} L_n^{(\alpha)}(x) &= (-1)^n (I - D)^{\alpha+n}\, x^n = (-1)^n \sum_{k \geqslant 0} (-1)^k \frac{(\alpha + n)_k}{k!} D^k x^n \\ &= (-1)^n \sum_{k \geqslant 0} (-1)^k \binom{n}{k} x^{-\alpha} D^k x^{n+\alpha} \\ &= x^{-\alpha}(D - I)^n\, x^{n+\alpha} = x^{-\alpha} e^x D^n e^{-x} x^{n+\alpha}, \end{aligned}$$

which is the classical Rodrigues formula.

Expanding the third formula on the right of the string of identities gives the coefficients of the Laguerre polynomials

$$\begin{aligned} L_n^{(\alpha)}(x) &= (-1)^n \sum_{k \geqslant 0} (-1)^k \frac{(\alpha + n)_k}{k!} (n)_k\, x^{n-k} \\ &= \sum_{i \geqslant 0} \frac{n!}{i!} \binom{\alpha + n}{n - i} (-x)^i. \end{aligned}$$

The binomial theorem for Sheffer polynomials (Proposition 2 of Section 5) yields the identity

$$L_n^{(\alpha)}(x+y) = \sum_{k\geqslant 0} \binom{n}{k} L_k(x) L_{n-k}^{(\alpha)}(y);$$

whence, upon applying the operator $(1-D)^{\beta+1}$ to both sides, we obtain the *first composition law*

$$L_n^{(\alpha+\beta+1)}(x+y) = \sum_{k\geqslant 0} \binom{n}{k} L_k^{(\beta)}(x) L_{n-k}^{(\alpha)}(y).$$

Further properties follow from the fact that

$$L_n^{(\alpha)}(x) = (-1)^n M_n^{[-\alpha-n]}(x), \qquad \text{or} \qquad M_n^{[\alpha]}(x) = L_n^{[-\alpha-n]}(x)\,(-1)^n.$$

Next we apply Theorem 7 to study the umbral composition of two Laguerre polynomials. A trivial identification of the various operators at hand yields

$$\begin{aligned} L_n^{(\alpha)}(\mathbf{L}^{(\beta)}(x)) &= (I-D)^{\beta-\alpha}\, x^n = M_n^{[\alpha-\beta]}(x) \\ &= (-1)^n L_n^{(\beta-\alpha-n)}(x). \end{aligned}$$

For $\beta = \alpha$ we obtain the remarkable identity

$$L_n^{(\alpha)}(\mathbf{L}^{(\alpha)}(x)) = x^n,$$

showing that all the Laguerre polynomials are self-inverse sets. This is true even of the basic Laguerre polynomials, which correspond to the case $\alpha = -1$.

So far we have considered only the umbral composition of $L_n^{(\alpha)}(x)$ with $L_n^{(\beta)}(x)$ and of $M_n^{[\alpha]}(x)$ with $M_n^{[\beta]}(x)$. Umbral composition of $M_n^{(\alpha)}(x)$ with $L_n^{(\beta)}(x)$ gives, by an application of Theorem 7, the Sheffer set relative to—oh surprise!—the delta operator $D/(D-I)$ and the operator $I/(I-D)^{(\alpha+\beta+1)}$, that is, the Laguerre polynomials again! In symbols,

$$\begin{aligned} M_n^{(\alpha)}(\mathbf{L}^{(\beta)}(x)) &= M_n^{(\beta)}(\mathbf{L}^{(\alpha)}(x)) \\ &= L_n^{(\alpha+\beta)}(x). \end{aligned}$$

Piecing together these results on umbral composition, we are led to the following remarkable *second composition law* for Laguerre polynomials:

$$\begin{aligned} &L_n^{(\alpha_1)}(\mathbf{L}^{(\alpha_2)}(\mathbf{L}^{(\alpha_3)}(\cdots \mathbf{L}^{(\alpha_k)}(x))\cdots)) \\ &\quad = \begin{cases} L_n^{(\alpha_1-\alpha_2+\alpha_3-\cdots+\alpha_k)}(x), & k \text{ odd}, \\ M_n^{[\alpha_1-\alpha_2+\alpha_3-\cdots-\alpha_k]}(x), & k \text{ even}. \end{cases} \end{aligned}$$

When expanded in powers of x, this equation leads to several binomial identities, of which we only give a sampling:

If k is even,

$$\binom{-\alpha_1+\alpha_2-\alpha_3+\cdots+\alpha_k}{m}$$
$$=\sum_{r_1,\ldots,r_{k-1}\geqslant 0}(-1)^{r_1+r_3+\cdots+r_{k-1}}\binom{\alpha_1}{r_1}\binom{\alpha_2-r_1}{r_2}\binom{\alpha_3-r_1-r_2}{r_3}\cdots$$
$$\times\binom{\alpha_{k-1}-r_1-r_2-\cdots-r_{k-2}}{r_{k-1}}\binom{\alpha_k-r_1-\cdots-r_{k-1}}{m-r_1-\cdots-r_{k-1}},$$

and if k is odd,

$$\binom{\alpha_1-\alpha_2+\alpha_3-\cdots+\alpha_k}{m}$$
$$=\sum_{r_1,\ldots,r_{k-1}\geqslant 0}(-1)^{r_2+r_4+\cdots+r_{k-1}}\binom{\alpha_1}{r_1}\binom{\alpha_2-r_1}{r_2}\cdots\binom{\alpha_k-r_1-\cdots-r_{k-1}}{m-r_1-\cdots-r_{k-1}}.$$

The so-called "duplication formulas" for Laguerre polynomials (see, e.g. Rainville) are trivial consequences of Theorem 7; we shall only derive one of them to indicate the method. We are to express $L_n(ax)$ as a linear combination of $L_k(x)$. By Section 7, the sequence $L_n(ax)$ is basic to the operator $a^{-1}D/(a^{-1}D-I)$. We are, therefore, to find a formal power series $f(t)$ such that $a^{-1}t/(a^{-1}t-1)=f(t/(t-1))$. An easy computation gives $f(t)=t/[(1-a)\,t+a]$. Now, the basic polynomials for $f(D)$ are computed by Theorem 4; they are

$$\begin{aligned}p_n(x)&=x[(1-a)\,D+aI]^n\,x^{n-1}\\&=\sum_{k=0}^{n-1}\binom{n}{k}(1-a)^k\,a^{n-k}(n-1)_k\,x^{n-k}\\&=\sum_{k=1}^{n}\frac{n!}{k!}\binom{n-1}{k-1}(1-a)^{n-k}\,(ax)^k.\end{aligned}$$

If we now apply the umbral operator $V:x^k\to L_k(x)$, then by Proposition 1 of Section 7 the sequence $Vp_n(x)$ will be basic for the delta operator

$$\begin{aligned}Vf(D)\,V^{-1}&=f(VDV^{-1})=f(K)=f(D/(D-I))\\&=\frac{a^{-1}D}{a^{-1}D-I},\end{aligned}$$

whose basic sequence is, as we have remarked, $L_n(ax)$. Thus, we are led to Erdelyi's formula

$$L_n(ax) = \sum_{k=1}^{n} \frac{n!}{k!} \binom{n-1}{k-1} (1-a)^{n-k} a^k L_k(x).$$

For the Laguerre polynomials of order α, Proposition 5 of Section 5 gives us the generating function

$$\sum_{n\geq 0} \frac{L_n^{(\alpha)}(x)}{n!} t^n = \frac{1}{(1-t)^{\alpha+1}} e^{xt/(t-1)}.$$

Since the generating function of $M_n^{[\alpha]}$ is easily seen to be

$$\sum_{n\geq 0} \frac{M_n^{[\alpha]}(x)}{n!} t^n = \frac{1}{(1-t)^{\alpha}} e^{xt}$$

and

$$M_n^{[\alpha]}(x) = (-1)^n L_n^{(-\alpha-n)}(x),$$

we obtain the following interesting relation:

$$\sum_{n\geq 0} \frac{L_n^{(\alpha-n)}(x)}{n!} t^n = (1+t)^{\alpha} e^{-xt}.$$

We will now generalize these relations and obtain generating functions for the sequences $L_n^{(\alpha+bn)}(x)$, where b is any fixed complex number. For b an integer these were first obtained by Brown, and Carlitz later generalized them to any b.

A routine calculation shows that $L_n^{(\alpha+bn)}(x)$ is Sheffer relative to the delta operator $Q_b = -D(I-D)^{-b-1}$. Since, by formula (2) of Theorem 4, the basic polynomials for Q_b are

$$(-1)^n (I-D)^{bn+n-1} (I+bD) x^n,$$

we discover that $L_n^{(\alpha+bn)}(x)$ is Sheffer relative to the invertible shift-invariant operator

$$S_{b,\alpha} = \frac{I+bD}{(I-D)^{\alpha+1}}.$$

If we now let $Q_b = q_b(D)$, $S_{b,\alpha} = s(b, \alpha, D)$, and $q_b^{-1}(t) = A(b,t)$, then by Proposition 5 of Section 5 we obtain

$$\sum_{n\geq 0} \frac{L_n^{(\alpha+bn)}(x)}{n!} t^n = (s(b, \alpha, A(b,t)))^{-1} e^{xA(b,t)},$$

which is the desired generating function. Further, since $A(b, t)$ is the (unique) formal power series solution to

$$\frac{-A}{(1-A)^{b+1}} = t,$$

an easy calculation shows that

$$A(-b-1, t) = \frac{-A(b, -t)}{1-A(b, -t)}.$$

Similarly, we discover that

$$s(-b-1, -\alpha, A(-b-1, t)) = s(b, \alpha, A(b, -t)) \cdot (1-A(b, -t)).$$

The spectral theory of Laguerre polynomials can only be sketched here. The classical inner product,

$$[f(x), g(x)]_\alpha = \int_0^\infty x^\alpha e^{-x} f(x)\, g(x)\, dx,$$

can be redefined so as to make sense not only for $\alpha > 0$, but for all α (except when α is a negative integer). Indeed, as with the Hermite polynomials we find

$$\int_0^\infty x^\alpha e^{-x} L_n^{(\alpha)}(x)\, g(x)\, dx$$
$$= \int_0^\infty D^n(x^{\alpha+n} e^{-x})\, g(x)\, dx = \int_0^\infty (-1)^n x^{\alpha+n} e^{-x} D^n g(x)\, dx$$
$$= \left[\int_0^\infty (-1)^n t^{\alpha+n} e^{-t} D^n g(x+t)\, dt\right]_{x=0} = \Gamma(\alpha+n+1)[K^n K_\alpha g(x)]_{x=0},$$

whereas, the inner product given by Proposition 1 of Section 9 is

$$[KK_\alpha(f(K)\, g(x))]_{x=0} = (f(x), g(x))_\alpha.$$

The two inner products do not coincide. The second inner product is, however, positive definite for all α; whereas, the first is symmetric for all α and gives

$$[L_n^{(\alpha)}(x), L_n^{(\alpha)}(x)]_\alpha = n!\, \Gamma(\alpha+n+1),$$

so that it is well defined, whenever α is not a negative integer. Nevertheless, the eigenfunction expansion still makes sense, and Theorem 9 readily yields the differential equation

$$L_n^{(\alpha)''}(x) + (\alpha+1-x) L_n^{(\alpha)'}(x) + nL_n(x) = 0.$$

Again we must leave a detailed analysis of these inner products to a later publication.

We shall now generalize slightly the Laguerre operator K and consider the delta operators

$$L_{\alpha,\beta} = l_{\alpha,\beta}(D) = \frac{\alpha D}{1 - \beta D}, \qquad \alpha \neq 0.$$

The Laguerre operator corresponds, of course, to $\alpha = -\beta = -1$. From formula (2) of Theorem 4 we find that the basic polynomials $J_n^{(\alpha,\beta)}(x)$ for $L_{\alpha,\beta}$ are given by

$$\begin{aligned} J_n^{(\alpha,\beta)}(x) &= \alpha^{-n}(1 - \beta D)^{n-1} x^n \\ &= \alpha^{-n} \sum_{k=1}^{n} \frac{n!}{k!} \binom{n-1}{k-1} (-\beta)^{n-k} x^k. \end{aligned}$$

Since

$$l_{\alpha,\beta}(l_{\alpha',\beta'}(D)) = l_{\alpha\alpha',\beta\alpha'+\beta'}(D),$$

we see that the $L_{\alpha,\beta}$ form a group under convolution and that this group is in fact isomorphic to the multiplicative group of matrices

$$\begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix}, \qquad \alpha \neq 0.$$

This enables us to easily compute the umbral composition of the $J_n^{(\alpha,\beta)}(x)$. Thus, for example, we obtain

$$J_n^{(\alpha,\beta)}(\mathbf{J}^{(\gamma,\delta)}(x)) = J_n^{(\alpha\gamma,\beta\gamma+\delta)}(x),$$

which yields the binomial identity. Deeper properties can be obtained by developing the theory of Sheffer sets relative to these operators.

12. Vandermonde Convolution

The difference analogs of Abel polynomials, with delta operator $E^{-b}\Delta$, may be called the *Gould polynomials* and denoted by $G_n(x, b)$. By the corollary to Theorem 4, we readily find the explicit expressions for the $G_k(x, b)$;

$$\begin{aligned} A_k(x, b) = G_k(x, b)/k! &= \frac{x}{x + bk} (x + bk)_k/k! \\ &= \frac{x}{x + bk} \binom{x + bk}{k}. \end{aligned}$$

We refer to Gould's papers for comparison. The identity expressing that these polynomials are of binomial type is sometimes known as the *Vandermonde convolution*, though the name is also applied to other identities. Gould's (1961, 1.1) is the generating function, a special case of Corollary 3 to Theorem 2. The binomial identity can be strengthened to

$$\sum_{k=0}^{n} \binom{n}{k} (p + qk)\, G_k(x, b)\, G_{n-k}(c, b) = \frac{p(x + c) + qxn}{x + c}\, G_n(x + c, b).$$

Gould's inverse relations are straightforward applications of Theorem 2. Since

$$(E^{-b}\Delta)^n = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} E^{j-nb},$$

we find that

$$F(n) = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} f(j - nb)$$

is the inverse of

$$f(x) = \sum_{n\geqslant 0} \frac{x}{x + bn} \binom{x + bn}{n} F(n) = \sum_{n\geqslant 0} A_n(x, b) F(n),$$

which can be considered as the basic inversion formulas associated with Vandermonde convolution (a recasting of Gould (1962, 3.1 and 3.2)). Several special cases are discussed by Gould, in particular, his Theorem 2 (1960).

We next obtain the connection constants of $G_n(x, c)$ in terms of $G_n(x, c - b)$. This is done most simply by expanding the first set in terms of the second. Now,

$$\begin{aligned} E^{b-c}\Delta G_n(x, c) &= E^b(E^{-c}\Delta G_n(x, c)) \\ &= E^b n G_{n-1}(x, c) = nG_{n-1}(x + b, c), \end{aligned}$$

and, therefore,

$$(E^{b-c}\Delta)^k\, G_n(x, c) = (n)_k\, G_{n-k}(x + kb, c);$$

whence, by Theorem 2

$$G_n(x+a, c) = \sum_{k\geqslant 0} \frac{G_k(a, c-b)}{k!} (E^{b-c}\Delta)^k G_n(x, c),$$

or

$$G_n(x+a, c) = \sum_{k\geqslant 0} \binom{n}{k} G_k(a, c-b)\, G_{n-k}(x+bk, c);$$

or, in Gould's notation

$$A_n(x+a, c) = \sum_{k\geqslant 0} A_k(x, c-b)\, A_{n-k}(a+bk, c).$$

For convenience we also write the inverse formulas, obtained by a change of parameters:

$$A_n(x+a, c-b) = \sum_{k=0}^{n} A_k(x, c)\, A_{n-k}(a-bk, c-b).$$

In more classical notation, this pair yields the inversion formulas:

$$f_n(x+a) = \sum_{k=0}^{n} F_k(x)\, A_{n-k}(a-bk, c-b),$$

$$F_n(x+a) = \sum_{k=0}^{n} f_k(x)\, A_{n-k}(a+bk, c).$$

This implies Gould's main theorem (1962, 5.3 and 5.4) and has the advantage of a simpler formulation. Next, the polynomials $(x+bk)_k$ are Sheffer relative to the delta operator $E^{-b}\Delta$. Hence, the binomial theorem for Sheffer polynomials (Proposition 2 of Section 5) gives

$$\binom{x+y+bn}{n} = \sum_{k\geqslant 0} \binom{x+bk}{k} \frac{y}{y+b(n-k)} \binom{y+b(n-k)}{n-k},$$

which is slightly deeper than the identity, obtained from the fact that the

$$E^a G_k(x, b) = \frac{x+a}{x+a+bk} (x+a+bk)_k$$

are a cross-sequence, namely

$$\frac{x+a+c}{x+a+c+bn}\binom{x+a+c+bn}{n}$$
$$=\sum_{k\geqslant 0}\frac{x+a}{x+a+bk}\binom{x+a+bk}{k}\frac{x+c}{x+c+b(n-k)}\binom{x+c+b(n-k)}{n-k}.$$

A similar identity follows from the fact that $E^a(x+bk)_k$ are a Steffensen sequence. These identities also give the connection constants for expressing $G_n(x, b)$ as a linear combination of $G_k(x, c)$. In short, the previous form reads

$$A_n(x+a+c, b)=\sum_{k\geqslant 0} A_k(x+a, b)\, A_{n-k}(x+c, b),$$

and the Steffensen form is

$$\binom{x+a+c+bn}{n}=\sum_{k\geqslant 0} A_{n-k}(x+a, b)\binom{x+c+bk}{k}.$$

The inverse set of the $G_n(x, b)$, call it $J_n(x, b)$, is easily computed by Theorem 7.

Consider the umbral operator W sending x^n to $(x)_n$, and, thus, $WDW^{-1}=\Delta$. The inverse operator sends $E^{-b}\Delta$ to $D(1+D)^{-b}$, a delta operator whose basic polynomials are

$$p_n(x)=x(1+D)^{nb}\, x^{n-1}=xe^{-x}D^{nb}e^{x}x^{n-1}$$
$$=\sum_{k\geqslant 0}\binom{bn}{k}(n-1)_k\, x^{n-k},$$

which are polynomials of Laguerre type.

Gould's summation formula 5.5 and Bateman's alternating convolution can also be obtained from the expansion theorem. We have thus "explained" most identities for the polynomials $A_n(x, b)$ given in Gould's two papers.

13. Examples and Applications

Appell Polynomials.

As already remarked, these are Sheffer polynomials relative to D. It is impossible to summarize here the immense literature on these sets; a few pertinent remarks must suffice.

If $p_n(x)=T^{-1}x^n$, then an easy computation gives

$$p_n(x)=((T^{-1})'\, T+x)\, p_{n-1}(x),$$

a useful recurrence formula which yields various classical formulas (for example, the recurrence for Hermite polynomials).

Expansion of the product $p_n(ax)\,g_k(bx)$ of Appell sets in terms of a third set were considered by Carlitz (1963); his results are special cases of those of Section 5.

By far the most widely studied class of Appell polynomials are the Bernoulli polynomials (see Nörlund). They correspond to the operator J^a, where

$$Jp(x) = \int_x^{x+1} p(t)\, dt.$$

(Since $DJ = \Delta$, J^a is also defined by $J^a = (\Delta/D)^a = [(e^D - I)/D]^a$.) For $a = 1$, we have $J^{-1}x^n = B_n(x)$, the familiar Bernoulli polynomials, whose elementary property can be gleaned from Section 5. The second expansion theorem yields the Euler–MacLaurin sum formula; generalizations (Nörlund) are obtained by taking the $B_n^{[a]}(x) = J^{-a}x^n$. From (3) of Theorem 4 we easily infer that the sequence $xB_{n-1}^{[na]}(x)$ is basic for the operator DJ^a. This fact, combined with the general results given previously, yields all of Nörlund's identities. The umbral properties of these polynomials are remarkable, but require an extensive separate treatment.

Appell sets with the Bernoulli-like property,

$$p_n(-x-1) = (-1)^n\, p_n(x),$$

were studied by Nielsen; Ward considered the more general functional equation,

$$p_n(ax + b) = c_n p_n(x), \qquad (*)$$

and called such Appell sequences ***regular***. If a is not a root of unity, the only regular sequence is

$$K_n(x) = c_n[x + b/(a-1)]^n; \qquad c_n = a^n.$$

When a is a root of unity, however, we find a wealth of possibilities, as follows: let a be a primitive rth root of unity, then every Appell set satisfying (*) can be uniquely represented in the form

$$p_n(x) = s_0K_n(x) + s_rK_{n-r}(x) + \cdots + s_{tr}K_{n-tr}(x),$$

and conversely.

Another extensively studied (by Nörlund) class of Appell polynomials is

$$E_n^{[a]}(x) = [1 + (\Delta/2)]^{-a}\, x^n,$$

and again their "properties" become special cases of the previous result. Again (Steffensen) the sequence $xE_{n-1}^{[na]}(x + na/2)$ is basic for the operator D cosh $(D/2)$. These sequences are variously called "Euler polynomials," an honor which is, however, bestowed upon a great many other polynomial sequences. For $a = -1$ we obtain, apart from a constant factor, the *Genocchi polynomials* $G_n(x)$, and $G_n(0)$ are the *Genocchi numbers*. The second expansion theorem applied to the Euler polynomials yields the *Boole* summation formula.

Inverse Relations.

Given two polynomial sequences $p_n(x)$ and $q_n(x)$, suppose we can determine the connection constants

$$p_n(x) = \sum_{k=0}^{n} c_{nk} q_k(x),$$

$$q_n(x) = \sum_{k=0}^{n} d_{nk} p_k(x),$$

then we can derive a pair of inverse relations. Given any sequence a_n, set $L(q_n(x)) = a_n$; this defines a linear functional L on the space $\mathbf{P}$. If $b_n = L(p_n(x))$, we have

$$b_n = \sum_{k=0}^{n} c_{nk} a_k, \qquad a_n = \sum_{k=0}^{n} d_{nk} b_k. \tag{*}$$

By specializing to suitable sets of Sheffer polynomials, a great many of the inverse relations in the literature can be explained. In this context, Theorem 7 will help find the inverse of certain infinite matrices.

The simple inverse relations in Riordan (pp. 43–49) fall under the present scheme. Glancing at Table 2.1 (Riordan, p. 49), we recognize that 1. and 2. reduce to Theorem 2 for Δ and the backward difference ∇, and the rest result from an umbral interpretation of the foregoing identities for Laguerre polynomials. For example, 6 follows from the fact that the basic Laguerre polynomials are self-inverse.

For the sake of clarity we discuss the simplest of all inverse relations, namely

$$a_n = \sum_{k \geqslant 0} (-1)^k \binom{n}{k} b_k; \qquad b_n = \sum_{k \geqslant 0} (-1)^k \binom{n}{k} a_k. \tag{*}$$

This is immediately understood by defining the linear functional $L(x^k) = b_k$, which by the first identity gives $a_n = L((1 - x)^n)$. Hence,

$$b_n = L((1 - (1 - x))^n),$$

which is the second identity.

Klee's identity (Riordan, p. 13),

$$\sum_{k \geqslant 0} (-1)^k \binom{n}{k} \binom{n+k}{m} = (-1)^n \binom{n}{m-n},$$

is another simple example of the use of such umbral techniques. Variants of the two inversion formulas derived previously are discussed by Riordan (pp. 49–54) and summarized in his Table 2.2 (p. 52). These inverse relations can be treated by the methods developed here.

Generating Functions.

To relate a generating function identity in the literature to the present techniques, we compare with the generating function of basic and Sheffer polynomials, thereby identifying the operators involved. Take, say Example 2 of Riordan (p. 100). Changing variables,

$$\frac{1}{(1-t)^x} = e^{x \log(1-t)^{-1}} = \sum_{n \geqslant 0} \frac{t^n}{n!} p_n(x),$$

where $p_n(x)$ is basic relative to backward difference; the inversion formula

$$a_n = \sum_{k \geqslant 0}^{n} \binom{p+k}{k} b_{n-k}; \qquad b_n = \sum_{k=0}^{n} (-1)^k \binom{p+k}{k} a_{n-k}$$

is, therefore, the umbral version of the expansion formula for ∇. Again following Riordan (p. 101), taking

$$e^{x \log(1-t-t^2)^{-1}} = \sum_{n \geqslant 0} \frac{p_n(x)}{n!} t^n,$$

we find that $p_n(x)$ are basic for $Q = [(5 - 4E^{-1})^{1/2} - 1]/2$ and $p_n(1)/n!$ are the Fibonacci numbers, whence a host of identities, and so forth to include Riordan (pp. 99–106).

409/42/3-16*

The case of exponential generating functions (Riordan, pp. 106–114) is simpler; most of the examples treated there reduce to Appell sets and their inverse. The same can be said of the theory of Lagrange series (Riordan, pp. 146–151).

The solution of transcendental equations is sometimes effectively carried out by operator methods. Suppose we are to find a solution t of $y = q(t)$. Letting $Q = q(D)$ (so that we require $q(0) = 0$ and $q'(0) \neq 0$), we find from

$$E^a = \sum_{n\geqslant 0} \frac{p_n(a)}{n!} Q^n$$

that the solution t (Theorem 3) is

$$e^{at} = \sum_{n\geqslant 0} \frac{p_n(a)}{n!} y^n.$$

The Heaviside Calculus

Although the name should be Boole's, the term is usually applied to the study of shift-invariant operators which are polynomials in D (the analog for Δ, although easily derived, does not seem to appear in any treatise on finite differences). There are two main applications. Any differential equation $p(D) f(x) = g(x)$ with $p(0) \neq 0$ has a ***unique*** polynomial solution for every polynomial $g(x)$, as follows immediately from Corollary 1 to Theorem 3 (this fact has been the point of departure for generalizations to functions of exponential type), and the inverse operator can be written in closed form using the Laguerre operator K and its iterations, which are easily simplified by the Riemann–Liouville formula.

The second (and less well known) is the theory of expansions of formal power series $f(t)$ in powers of a given polynomial $p(t)$ with $p(0) = 0, p'(0) \neq 0$:

$$f(t) = \sum_{n\geqslant 0} \frac{a_n}{n!} (p(t))^n.$$

How are the coefficients a_n to be determined? There is a unique inverse power series $p^{-1}(t)$ of the polynomial $p(t)$. Suppose a delta operator Q can be found for which both $R = p^{-1}(Q)$ and $f(R)$ have a simple enough form. Then $a_n = [f(R)\, p_n(x)]_{x=0}$, where $p_n(x)$ are the basic polynomials of Q, by Theorem 2. This technique works more often than it appears; we illustrate it with an example from the literature.

It was reputedly proved by Schur that in the expansion

$$\sin \pi x = \sum_{n=1}^{\infty} \frac{a_n}{n!} (x(1-x))^n \qquad (*)$$

the coefficients a_n are positive, but no explicit expression was found. Carlitz (1966) found an explicit formula for the coefficients, but it is not clear from his result that the a_n are positive.

Now, it is obvious from (*) that the delta operator in question is $Q = D(I - D)$, whose basic polynomials are $p_n(x)$, computed by

$$p_n(x) = x(I-D)^{-n} x^{n-1},$$

that is,

$$(I-D)^{-n} = \left(\frac{1}{I-D}\right)^n = (I + D + D^2 + D^3 + \cdots)^n$$

$$= I + nD + \binom{n+1}{2} D^2 + \cdots,$$

$$p_n(x) = \sum_{i=0}^{n-1} \binom{n+i-1}{i} (n-1)_i \, x^{n-i}$$

thus (Theorem 3)

$$e^{ax} = \sum_{n \geq 0} \frac{p_n(a)}{n!} (x(1-x))^n.$$

Setting $A_n = [p_n(\pi i) - p_n(-\pi i)]/2in!$, Carlitz's explicit expression is obtained. The polynomials $p_n(x)$ and the coefficients a_n can be expressed in the closed form

$$p_n(x) = \frac{x}{(n-1)!} \int_0^{\infty} e^{-t}[t(x+t)]^{n-1} \, dt,$$

$$a_n = \frac{\pi}{(n-1)!} \int_0^{\pi/2} [y(\pi - y)]^{n-1} \sin y \, dy$$

easily derived from the integral form of $(1 - D)^{-n}$. From this, the positivity of a_n can be inferred.

The well known Bessel polynomials $y_n(x)$ of Krall and Frink are not a Sheffer set, but the related set $f_n(x) = x^n y_{n-1}(x^{-1})$ is one. Its delta operator is $Q = D - D^2/2$. This makes some of the results in Carlitz (1957) special cases of the present theory. For instance, the generating function, (Carlitz's 2.5)

$$\sum_{n \geq 0} \frac{f_n(x)}{n!} t^n = e^{x[1-(1-2t)^{1/2}]},$$

the property of being of binomial type (2.7); and Carlitz's (2.8) are obtained by computing the connection constants with x^n. The formulas expressing the derivatives of $f_n(x)$ as linear combinations of the $f_n(x)$ follow from the expansion theorems (2.10, 2.12) as do (3.1, 3.2). Burchnall's $\theta_n(x)$ are the Sheffer set relative to $Q' = 1 - D$; this gives $(-2)^n \theta_n(x/2) = L_n^{(-2n-1)}(x)$ by an easy umbral computation. Carlitz's (4.4) gives the connection constants between $L_k^{(\alpha)}(2x)$ and $f_n(x)$, which follow from Theorem 7, and (4.6) connects $\theta_n(x)$ with $f_n(x)$.

Difference Polynomials

They are the Sheffer sets associated with the difference operator $\Delta = E - I$, having the basic polynomials $(x)_n = x(x-1) \cdots (x-n+1)$. (The closely related backward difference operator, $\nabla = I - E^{-1}$, has the basic polynomials $x^{(n)} = x(x+1) \cdots (x+n-1)$. Curiously, the connection constants of $x^{(n)}$ with $(x)_n$ are, apart from sign, the coefficients of the basic Laguerre polynomials (an easy computation using Theorem 7).)

The generating function of a set of difference polynomials can be written in the suggestive form $s(t)^{-1}(1+t)^x$.

The first expansion theorem applied to Δ gives the *Newton* expansion. The expansion of the Bernoulli operator J in powers of Δ is *Gregory's formula.*

Newton's expansion, combined with the identity,

$$\Delta^n = \sum_{k \geqslant 0} \binom{n}{k} (-1)^{n-k} E^k,$$

gives a pair of inverse relations which could simplify many a calculation in the literature (e.g. Carlitz (1952)). Notable difference sets (cf. Boas and Buck) are:

(a) *Poisson-Charlier* polynomials, with $S = E$ (apart from a parameter);

(b) *Narumi* polynomials, with $S = D^k/(\log(I + D))^k$;

(c) *Boole* polynomials, with $S = I + (I + D)^k$;

(d) *Peters* polynomials, with $S = (I + (I + D)^k)^\lambda$;

(e) *Bernoulli* polynomials of the second kind $b_n(x) = J(x)_n$, extensively studied by Jordan.

(f) The *Stirling polynomials* $N_n(x)$, introduced by Nielsen (p. 72), are the basic set inverse to the upper factorial powers $x^{(n)}$. They are, therefore, easily reduced to the exponential polynomials. Nielsen's notation $\psi_n(x)$ is related to the present notation by $(x+1)\psi_n(x) = N_n(-x-1)/n!$. The central difference operator $S = [E^{1/2} - E^{-1/2})/2$ has an extensive literature (but see Riordan, pp. 212–217); it is a special case of an Abel operator. Its basic polynomials are written $x^{[n]}$; their connection constants (the central factorial coefficient) with x^n were computed by Carlitz and Riordan, and their results are derived from Theorem 7 and its corollaries. Expansions in powers of S, such as the formulas of Lubbock and Woolhouse, are heuristically derived by Steffensen; they can, of course, be verified by Theorem 2, whose application becomes particularly useful when the sign of a square root is to be chosen.

It does not seem to have been realized that Newton's expansion and its variants obtained from Theorem 6 yield a powerful technique for proving *binomial identities*. We give a sampling, taken from Riordan (pp. 1–18).

The original Vandermonde formula (3a),

$$\binom{n+p}{m} = \sum_{k\geqslant 0} \binom{n}{k}\binom{p}{m-k},$$

follows from the expansion of $(n+p)_m$ in terms of the basic polynomials $(n)_k$. Grosswald's identity (Example 7),

$$\sum_{k=0}^{2p} (-2)^{-k} \binom{n}{m+k}\binom{n+m+k}{k} = (-1)^p\, 2^{-2p} \binom{n}{p}, \qquad n-m=2p,$$

becomes clear when one replaces m by $m-n$:

$$\sum_{k=0}^{2p} (-2)^{-k} \binom{n}{m-n+k}\binom{m+k}{k} = (-1)^p\, 2^{-2p} \binom{n}{p},$$

with $2n-m=2p$. Again replacing k by $2p-k$ on the left, this reduces to

$$\sum_{k=0}^{2p} (-2)^{k-2p} \binom{n}{k}\binom{m+2p-k}{2p-k} = (-1)^p\, 2^{-2p} \binom{n}{p},$$

and this is clearly a Newton expansion relative to the basic polynomials $(n)_k$; the computation of the coefficient is routine.

The expansion of a product of two binomial coefficients (10),

$$\binom{n}{p}\binom{n}{g} = \sum_{k\geqslant g} \binom{k}{g}\binom{g}{k-p}\binom{n}{k}, \tag{*}$$

follows the same reasoning. Because of its importance, we derive it in full. Jordan's formula,

$$\Delta^k(uv) = \sum_{j=0}^{k} \binom{k}{j} \Delta^j u \Delta^{k-j} E^j v,$$

gives, when $u = (x)_p$ and $v = (x)_g$ and $g \geqslant p$,

$$\begin{aligned}[\Delta^k((x)_p\,(x)_g)]_{x=0} &= \binom{k}{p} p![\Delta^{k-p}E^p(x)_g]_{x=0} \\ &= \binom{k}{p} p!(g)_{k-p}\,[E^p(x)_{g-k+p}]_{x=0} \\ &= \binom{k}{p} p!(g)_{k-p}\,(p)_{g+p-k} = \binom{k}{g}\binom{g}{k-p} p!g!,\end{aligned}$$

as desired.

Shanks' result that

$$\binom{x}{i}^k = \sum_{j=1}^{ik-i+1} A_{kj}^i \binom{x+j-1}{ik}$$

with $A_{kj}^i > 0$, can be established in the same way, but the literature on the A_{kj}^i is scarce.

Abel polynomials

They are the basic polynomials for the delta operator $Q = E^\alpha D$, given by (3) of Theorem 4 as

$$A_n^{(\alpha)}(x) = x(x - n\alpha)^{n-1}.$$

Expansions into Abel polynomials have an extensive theory (Hurwitz, Salié, Boas, and Buck). The polynomials have notable statistical and combinatorial significance. Identities for the Abel polynomials, as well as for the related Sheffer polynomials $(x - (n + 1)\,a)^n$, follow the same pattern as those for the Gould polynomials. All identities in Riordan (pp. 18–23) can be obtained either by one of the expansion theorems or by umbral composition (sometimes by both methods). Similarly, the Abel inverse relations of Riordan (pp. 92–99) can be obtained by either of the foregoing methods or by recognizing a cross-sequence. As we have already described the techniques in deriving Gould's inversion formulas, we shall not repeat them here. As a simple example of an inverse pair, we quote the following, due to Clarke:

$$b_n = \sum_{k\geqslant 0} \binom{n}{k} kn^{n-k-1}a_k\,,$$

$$a_n = \sum_{k\geqslant 0} (-1)^{n+k} \binom{n}{k} k^{n-k}b_k\,,$$

which the reader will readily identify.

Abel's identity,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} (y + ka)^{n-k} x(x - ka)^{k-1},$$

is nothing but an instance of the first expansion theorem as is the superficially remarkable identity in Bernoulli and Abel polynomials

$$B_n(x + y) = \sum_{k=0}^{n} \binom{n}{k} B_{n-k}(y + ka) x(x - ka)^{k-1},$$

and many similar formulas stated by Nörlund, Steffensen, and others. The inverse set to the Abel polynomials does not seem to have been considered, though they have a combinatorial significance, and we shall briefly derive its properties here. Let

$$\begin{aligned} B_n^{(a)}(x) &= \sum_{k\geqslant 0} \binom{n}{k} x^k (ka)^{n-k} \\ &= \sum_{k\geqslant 0} \frac{x^k}{k!} [E^{ka} D^k x^n]_{x=0}; \end{aligned}$$

from the summation formula we recognize that these are indeed the inverses of the Abel polynomials. Their umbral recursion formula is

$$\mathbf{B}^{(a)}(x) (\mathbf{B}^{(a)}(x) - na)^{n-1} = x^n,$$

and the identity stating that the two sets are inverse is

$$x^n = \sum_{k\geqslant 0} \binom{n}{k} (ka)^{n-k} x(x - ka)^{k-1}.$$

The summation formula (Corollary 7 of Theorem 7) becomes

$$f(\mathbf{B}^{(a)}(x)) = \sum_{k\geqslant 0} \frac{x^k}{k!} f^{(k)}(ka).$$

This identity gives ample evidence of the simplicity of the umbral method.

Various authors have considered basic polynomials relative to the operator $Q = E^a(1 + D)^b D$. The connection constants with the Abel polynomials are easily found by Theorem 4:

$$p_n(x) = \sum_{k=0}^{n-1} (-1)^k \binom{nb + k - 1}{k} (n - 1)_k A_{n-1-k}^{(a)}(x).$$

For $Q = E^a e^{D^2/2} D$ we find a generalization of the Hermite polynomials considered by Steffensen. The theory of crosssequences expresses them at once in terms of the Hermite polynomials $H_n(x)$, that is,

$$n^{(n-1)/2}xH_{n-1}\left(\frac{n\alpha - x}{n^{1/2}}\right) = p_n^{(\alpha)}(x).$$

The connection constants with x^n can be computed by the summation formula, in view of the fact that the inverse polynomials can be expressed in terms of the inverses of the Abel polynomials. This gives

$$p_n^{(\alpha)}(x) = \sum_{k=0}^{n-1} \binom{n-1}{k} n^{k/2} H_k[\alpha(n^{1/2})]\, x^{n-k}.$$

The inverse connection constants can also be computed by Theorem 7; for $\alpha = 0$ we have

$$x^{2n} = \sum_{k=1}^{n} \frac{(2n)_{2n-2k}}{(n-k)!} k^{n-k} p_{2k}^{(0)}(x),$$

$$x^{2n+1} = \sum_{k=0}^{n} \frac{(2n+1)_{2n-2k}}{(n-k)!} \left(\frac{2k+1}{2}\right)^{n-k} p_{2k+1}^{(0)}(x),$$

two Hermite-reminding identities.

Cotlar Polynomials

An interesting class of Sheffer operators associated with the difference operator Δ has been studied by Cotlar. It is easy to see that a polynomial sequence $p_n(x)$ has the property that $p_n(k) = p_k(n)$ for all nonnegative integers k and n, if and only if it can be written in the form

$$p_n(x) = \sum_{i=0}^{n} \binom{n}{i} \lambda_i (x)_i \,.$$

for some sequence $\lambda_i \neq 0$. Such sequences of polynomials are said to be ***permutable***. There is one and only one permutable Sheffer set—except for a parameter; it must be a Sheffer set for the delta operator Δ and the invertible operator $(I - a\Delta)^{-1}$; it has the explicit expression

$$a^n(x)_n + \binom{n}{1} a^{n-1}(x)_{n-1} + \cdots + 1 = p_n(x).$$

Again, all Sheffer sets $p_n(x)$ such that the sequence $g_n(x) = p_n(x)/n!$ is permutable can be classified (Cotlar). The delta operator is $\log(1 + aD/(D - I))$

and the invertible operator is $(1 - D)^{-1}$. In particular for $a = -2$ one obtains a sequence of Sheffer polynomials $M_n(x)$ enjoying the remarkable properties

$$\begin{aligned} M_n(-x-1) &= (-1)^n M_n(x), \\ M_n(k) &= M_k(n), \qquad & k, n \geqslant 0, \\ (-1)^k M_{k-1}(-n) &= (-1)^n M_{n-1}(-k); \qquad & k, n \geqslant 1. \end{aligned}$$

It can be shown that the three foregoing properties uniquely determine the sequence $M_n(x)$, which is in fact explicitly given by

$$M_n(x) = \frac{2^n}{n!}(x)_n + \frac{2^{n-1}}{(n-1)!}(x)_{n-1} + \cdots + 1.$$

The inverse set of the $M_n(x)$ can be expressed in terms of Bernoulli polynomials.

Exponential polynomials

Also of statistical origin are the ***exponential polynomials*** $\phi_n(x)$, introduced by Steffensen and studied further by Touchard and others. Some of their properties were developed in III. We recall that they are the basic polynomials for the delta operator $\log(I + D)$, and that they are inverse to $(x)_n$, so that

$$\phi(\phi - 1) \cdots (\phi - n + 1) = x^n,$$

and

$$\begin{aligned} \phi_n(x) &= \sum_{k \geqslant 0} \frac{x^k}{k!} [\Delta^k x^n]_{x=0} \\ &= \sum_{k \geqslant 0} S(n, k)\, x^k, \end{aligned}$$

where, following Riordan's notation, the $S(n, k)$ denote the Stirling numbers of the second kind (and $s(n, k)$ those of the first). Also, the Rodrigues formula ((4) of Theorem 4) says that

$$\phi_n(x) = x(\phi_{n-1}(x) + \phi'_{n-1}(x)).$$

The generalized Dobinsky formula follows most easily by umbral methods. Let $p_n(x) = (x)_n$. Then

$$p_n(\phi(x)) = x^n = e^{-x} \sum_{k \geqslant 0} \frac{p_n(k)}{k!} x^k,$$

and, hence, by linearity

$$p(\phi(x)) = e^{-x} \sum_{k \geqslant 0} \frac{p(k)}{k!} x^k,$$

for every polynomial $p(x)$. Setting $p(x) = x^n$ we obtain finally

$$\phi_n(x) = e^{-x} \sum_{k \geqslant 0} \frac{k^n x^k}{k!}.$$

Similarly one establishes the recursion

$$\phi_{n+1}(x) = x(\phi(x) + 1)^n.$$

We shall add to the properties developed in III the generating function,

$$\sum_{n \geqslant 0} \frac{\phi_n(x)}{n!} t^n = e^{x(e^t - 1)},$$

and Rodrigues' formula, implicitly established in III, that

$$\phi_n(x) = e^{-x}(\mathbf{x}D)^n e^x,$$

which shows the roots of these polynomials to be real. Also, recall that the connection constants with x^n are the Stirling numbers of the second kind. The connection constants between x^n and $\phi_n(x)$ are the Stirling numbers of the first kind, since the $\phi_n(x)$ are the inverse set of the $(x)_n$.

As an example of computation of a "new" set of connection constants, we shall connect the Laguerre polynomials with the polynomials $\phi_n(-x)$. It is easy to see that the $\phi_n(-x)$ are basic for the delta operator $\log(I - D)$. Thus, we must find a formal power series $f(t)$ such that $f(\log(1 - t)) = t/(t - 1)$. Clearly $f(t) = 1 - e^{-t}$ is the desired series. The connection constants are therefore given by the coefficients of the basic sequence for the backward difference operators $\nabla = I - E^{-1}$, namely the polynomials $x(x + 1) \cdots (x + n - 1)$. In symbols,

$$L_n(t) = \phi(x)(\phi(x) + 1) \cdots (\phi(x) + n - 1)$$
$$= \sum_{k=0}^{n} |s(n, k)| \phi_k(x).$$

Riordan's treatment of operators (pp. 200–205) furnishes a further batch of examples of the present theory. We shall now briefly develop some of the properties of the polynomials

$$\psi_n(x) = \sum_{k \geqslant 0} s(n, k)(x)_k,$$

which are the difference analogs of the exponential polynomials. The umbral theory of these two sets of polynomials can be used to systematically develop identities for the Stirling numbers.

If V is the umbral operator defined by $V(x)_n = \psi_n(x)$, then by Proposition 1 of Section 7 $\psi_n(x)$ is basic for $V\Delta V^{-1}$. But $Vx^k = (x)_k$, since

$$(x)_n = \sum_{k\geqslant 0} s(n, k)\, x^k,$$

and so $VDV^{-1} = \Delta$. Therefore, $\psi_n(x)$ is basic for

$$Q = V\Delta V^{-1} = V(e^D - I)\, V^{-1} = e^\Delta - I.$$

But then, by Theorem 7,

$$\psi_n(\phi(x)) = \phi_n(\psi(x)) = (x)_n\,,$$

which give orthogonality relations for the Stirling numbers. The reader should convince himself that Stirling number identities can be inferred from identities relating the $\phi_n(x)$ and the $\psi_n(x)$. We give a sampling, leaving the umbral proofs as exercises.

(1) $\psi_n(x) = e^\Delta \left(\dfrac{\Delta}{e^\Delta - I}\right)^{n+1} (x)_n\,.$

(2) $\phi_{n+1}(x) = x(\phi(x) + 1)^n$ gives

$$S(n + 1, k) = \sum_{i\geqslant 0} \binom{n}{i} S(i, k - 1).$$

(3) $\phi_n(\psi(x)) = (x)_n$ gives

$$\sum_{k\geqslant 0} S(n, k)\, s(k, i) = \delta_{ni}\,.$$

$\psi_n(\phi(x)) = (x)_n$ gives

$$s(n, k) = \sum_{k,i\geqslant 0} s(n, k)\, s(k, i)\, S(i, j).$$

(4) $\phi_n(x) = e^{-x} \sum_{k\geqslant 0} \dfrac{x^k k^n}{k!}\,.$

Taylor's expansion gives

$$e^x\phi_n(x) = \sum_{k\geqslant 0} \frac{x^k}{k!} [D^k e^t \phi_n(t)]_{t=0},$$

which implies

$$k^n = \sum_{i\geqslant 0} \binom{k}{i} i!\, S(n, i).$$

Also by Taylor,

$$e^{-x} \sum_{k\geqslant 0} \frac{x^k k^n}{k!} = \sum_{k\geqslant 0} \frac{x^k}{k!} \left(D^k \sum_{i\geqslant 0} \frac{e^{-t} t^i i^n}{i!}\right)_{n=0}$$

$$= \sum_{k\geqslant 0} \frac{x_k}{k!} \sum_{i\geqslant 0} \binom{k}{i} (-1)^{k-i} i^n,$$

which implies

$$S(n, k) = \frac{1}{k!} \sum_{i\geqslant 0} \binom{k}{i} (-1)^{k-i} i^n.$$

(5) $\phi_n(x)$ of binomial type gives

$$\binom{i+j}{i} S(n, i+j) = \sum_{k\geqslant 0} \binom{n}{k} S(k, i)\, S(n-k, j).$$

(6) $\psi_n(x)$ of binomial type gives

$$\binom{i+j}{i} s(n, i+j) = \sum_{k\geqslant 0} \binom{n}{k} s(k, i)\, s(n-k, j).$$

14. Problems and History

We have assembled in random order some open questions suggested by the preceding theory. Other problems are mentioned in the text.

(1) The present work unifies and extends the identities given by Riordan (pp. 1–23, 43–54, 92–116, 128–131, 141–152, 200–205, 212–217), that is, 82 out of 146 pages of text or 56%. We have excluded the exercises for reasons of time. Notable exceptions are Riordan's theory of Chebychev and Legendre inversions, the Bell polynomials, and differential operators of the type xD. Each of these topics calls for a development along a similar line but with a different invariance property than shift-invariance.

(2) Expansions of products of polynomials of one set in terms of those of another can be carried out by the foregoing methods but with difficulty. Indications from special identities (e.g. Hermite, Laguerre) are that there should be a general technique, which could apply more successfully to summing multiple binomial coefficients.

(3) Let $Qx = 1$ for the delta operator Q. Then Q can be embedded in a one-parameter group of operators $Q^{(t)}$ whose indicators satisfy the functional equation

$$q^{(t)}(q^{(s)}(x)) = q^{(s+t)}(x).$$

The corresponding basic sets satisfy

$$q_n^{(t)}(\mathbf{q}^{(s)}(x)) = q_n^{(s+t)}(x).$$

Develop the theory of such sets. How can the "infinitesimal generator" be computed? The simplest example of this is the basic Laguerre set.

(4) It has been suggested by Gould that some of the identities in Vandermode convolution are analogous to Kapteyn series. Several other analogies with classical eigenfunction expansions can be noted, which suggest an extension of the theory to classes of special functions. Truesdell's theory is helpful in this connection. Another possible extension is to exponential polynomials.

(5) Statistical, probabilistic and combinatorial interpretations of the identities are worthwhile. Several special sets, e.g. Abel, are connected with particular distributions of statistics (see e.g. Dwass, Pyke). There are at least three possibilities; interpretation as compound Poisson processes; interpretation through stationary stochastic processes, as in the relation of Hermite polynomials to Brownian motion of the Poisson–Charlier polynomials by the Poisson process, and, finally, the combinatorial interpretation through counting binomial type structures such as reluctant functions (see III). Very little is known about combinatorial interpretation of Sheffer polynomials; occasionally (Laguerre) they arise in counting permutations with restricted position. A major step forward would be a combinatorial or probabilistic interpretations of Bernoulli numbers; we surmise that the fact that these are, apart from a factor, the cumulants of the uniform distribution is relevant.

(6) One of the most difficult open problems is that of estimating the remainder after n terms in the expansion formulas. Little is known except in the Appell case. For p-adic convergence, the results are comparatively simple (see LeVeque, p. 55ff.), but undeveloped.

(7) Which Sheffer sets are orthogonal relative to some weight function in some region of the complex plane? Such a region is probably related to the convergence region of Boas and Buck.

(8) Another approach to the present theory is through the techniques of Hopf algebras. The algebra of polynomials in the variable x is a Hopf algebra, with diagonal map

$$\Delta: x^n \to \sum_{k=0}^{n} \binom{n}{k} x^k \otimes x^{n-k}.$$

The dual Hopf algebra is the algebra of differential operators with constant coefficients, the pairing between the two being given by

$$\langle p(D), q(x) \rangle = [p(D)\, q(x)]_{x=0} .$$

An umbral operator can be defined as one that commutes with the diagonal map, for example. The greater elegance of this approach is evident, as are some of its advantages: one can consider differential operators acting or polynomials or polynomials $p(x)$ as operators on operators. In addition, this point of view should point the way to a generalization to several variables, to the exterior algebra (in infinite dimensions) and to more general Hopf algebras. The theory of spherical harmonics should fit in one such generalization.

(9) There is a curious relationship between the coefficients of the expansion of a probability distribution into Hermite polynomials, and the cumulants. If the mean is zero and the variance one, the two coincide up to $n = 5$; this led Jordan (1972) to mistakenly conclude (p. 150) that they all coincide, but see Kendall and Stuart (p. 158). At any rate, the relationship between the two sets of coefficients seems fairly simple and should be worked out, especially in view of the mystery underlying the cumulants. Note that one can define cumulants relative to any sequence of binomial type, e.g. the factorial cumulants (Kendall and Stuart). Do these lend themselves to easier interpretations?

(10) There is no special reason for choosing polynomials instead of trigonometric polynomials; various identities relating Fourier and Dirichlet expansions might become clearer, for example the relationship between Bernoulli numbers and the values of the zeta function.

(11) Work out formulas for $p_n(Q)$, when $p_n(x)$ is a Sheffer set relative to the delta operator Q.

(12) There are several relationships between the factorization of differential operators with polynomial coefficients (of which no general theory exists) and Sheffer sets, see e.g. the last chapter of Riordan and various papers of Klamkin and Newman. One should begin by developing the theory of xD; for example, $L_n(xD)$ has a simple expression (why?). See also Rainville (1941), Carlitz (1930), and Carlitz (1932).

(13) The Laguerre polynomials are formally related to the gamma distributions as the Hermite to the normal, the Poisson–Charlier to the Poisson; nevertheless, a specific construction of the corresponding stochastic process or a group of transformations relative to which they are the "spherical harmonics" seems to still be missing.

(14) Various representations of the inner product making the Sheffer polynomials orthogonal are possible, and they should be investigated. The classical theory of orthogonal polynomials may have extensions to inner products "involving derivatives." In what sense is the inner product of Section 9 "natural"? The inner product for the Hermite polynomials with negative or imaginary variance is particularly interesting, in view of possible connection with the Feynman integral.

(15) The explicit representation of umbral operators leads to operator-differential equations in the Pincherle derivative, and is an untouched subject of great interest.

(16) The theory of factorial series (see e.g. Nörlund or Nielsen) indicates that expansions in series of the form $\sum_{n \geqslant 0} a_n/p_n(x)$ are at least possible in some cases. Is it possible to extend the present theory in this direction?

(17) In the same vein, the divided difference operation,

$$\Delta : f(x) \to \frac{f(x) - f(y)}{x - y},$$

is easily checked to be coassociative. This suggests that the theory be best developed in the context of coalgebras (Sweedler) and that a suitable notion of shift-invariance may be at hand. The same may be said of Thiele's inverse differences (Nörlund).

(18) An operational calculus, as understood in the last fifty years, is an isomorphism of a function algebra into an algebra of operators. In this respect, the isomorphism in the present calculus possesses one extra feature: it preserves functional composition, in fact, it gives meaning to it in terms of an operation on operators. Can this feature be carried over to other operational calculi?

(19) Work out representations of shift-invariant operators analogous to Post's inversion formula for the Laplace transform.

(20) Under what conditions are the zeros of a Sheffer set real?

(21) Evidently the kind of umbral composition we have considered is not as general as it should be, as it does not explain why $H_{2n}(x)$ is a constant multiple of $L_n^{(-1/2)}(x^2)$.

(22) The analogy between the functions e^{ax} and $(a - x)^{-1}$ suggests that there should be a theory of operators where shift-invariance is replaced by the functional equation

$$(a - x)^{-1} - (a - y)^{-1} = (x - y)(a - x)^{-1}(a - y)^{-1}.$$

This suggests parametrized families T_x of operators such that

$$T_x T_y = (T_x - T_y)/(x - y).$$

Some work of Redheffer supports this feeling.

(23) It is easy to see that a polynomial $p(x)$ is positive for all integer values of x if and only if its expansion in a Newton series has nonnegative coefficients. We conjecture that analogous results exist for Laguerre and Hermite polynomials and relate to the position of the zeros of these polynomials.

History

It is impossible to account for the detailed development of the Heaviside calculus from its beginnings; we shall only mention the works that relate to the present approach. Perhaps the most striking feature of this subject is that each author in the past would develop one approach to the exclusion of others. Thus, Carlitz, Riordan, and Steffensen, while feeling at home with generating functions, are somewhat ill-at-ease when handling operators, called by Steffensen "symbols." Pincherle, on the other hand, is fully aware of the abstract possibilities of the concept of operator, but ignorant of the nitty-gritty of numerical analysis, where he would have found a fertile ground for his ideas. Sheffer also uses power series in preference to operators, with a resulting lack of completeness.

The characterizations of basic polynomials, Sheffer polynomials and cross-sequences in terms of a binomial property (Theorems 1 and 8, and Proposition 6 of Section 5) are new. Other authors have used characterizations in terms of operators, thereby missing one of the main techniques. The two expansion theorems may also be said to be new, although various partial versions may be found in the literature from Pincherle on. The notions of a delta operator and basic sets are due to Steffensen (who, however, did not give them a name and did not realize that they were one and the same as sequences of binomial type) as is that of a cross-sequence (again unnamed and uncharacterized). The isomorphism theorem was at least intuited by Pincherle, and has been tacitly—and often unrigorously—used by several authors.

The idea of applying the Pincherle derivative (the name is ours) in the present context is new; it greatly simplifies the proof of Theorem 4 (first guessed by Steffensen) as well as the theory of Laguerre and Hermite polynomial, to name only a few instances. Theorem 5 is new (first stated in III). The recurrence formulas are due to Sheffer, as are the eigenfunction expansion formulas, with the exception of the explicit inner products; his proofs, however, use power series. Section 7 is new, as are most of the results in Section 8. In the examples, detailed references are given.

An extended bibliography has been appended as a hunting ground for further applications and extensions of the present methods. Items cited in the bibliography of Mullin–Rota will not be repeated here.

Acknowledgments

Writing this paper has been like assembling a dinosaur from a few charred bones in the desert. Although some of the results are completely new, a great many were suggested, inspired, and partially proved by previous authors. We owe a special debt of gratitude to Askey, Carlitz, Gould, Pincherle, Riordan, Sheffer, Stein, Steffensen, Touchard, and Truesdell.

David Kahaner wrote the early version of Sections 5, 7, 10, and 11, based upon Rota's lectures at Los Alamos, and Andrew Odlyzko thoroughly checked the manuscript and contributed parts of Sections 10, 11 and 13.

References

N. Abdul-Hakim and W. A. Al-Salam, A characterization of the Laguerre polynomials, *Rend. Sem. Mat. Univ. Padova* **XXXIV** (1964), 176–179.

W. A. Al-Salam and L. Carlitz, On certain generating functions for the Hermite polynomials, *Portugal. Math.* **25** (1966), 35–45.

W. A. Al-Salam and A. Verma, Generalized Sheffer polynomials, *Duke Math. J.* **37** (1970), 361–365.

W. A. Al-Salam, Some characterizations of the Laguerre and Hermite polynomials, *Michigan Math. J.* **10** (1963), 381–383.

W. A. Al-Salam, Operational derivation of some formulas for the Hermite and Laguerre polynomials, *Bull. UMI* **18** (1963), 358–363.

W. A. Al-Salam, Operational representations for the Laguerre and other polynomials, *Duke Math. J.* **31** (1964), 127–142.

W. A. Al-Salam, Remarks on some operational formulas, *Rend. Sem. Mat. Univ. Padova* **35** (1965), 128–131.

C. A. Anderson, Some properties of Appell-like polynomials, *J. Math. Anal. Appl.* **19** (1967), 475–491.

P. Appell, Sur une classe de polynômes, *Ann. Sci. École Norm. Sup.* **9** (1880), 119–144.

E. T. Bell, Exponential polynomials, *Ann. Math.* **35** (1934), 258–277.

E. T. Bell, Exponential numbers, *Amer. Math. Monthly* **XLI** (1934), 411–419.

E. T. Bell, The iterated exponential function, *Ann. Math.* **39** (1938), 539–557.

E. T. Bell, A history of Blissard's symbolic method, with a sketch of its inventor's life, *Amer. Math. Monthly* **XLV** (1938), 414–421.

E. T. BELL, Postulational bases for the umbral calculus, *Amer. J. Math.* **62** (1940), 717–724.

L. BERNSTEIN, An explicit summation formula and its application, *Proc. Amer. Math. Soc.* **25** (1970), 323–334.

M. T. BIRD, On generalizations of sum formulas of Euler–MacLaurin type, *Amer. J. Math.* **LVIII** (1938), 487–503.

Z. W. BIRNBAUM AND R. PYKE, On some distributions related to the statistic D_n^+, *Amer. Math. Stat.* **29** (1958), 179–187.

J. BLISSARD, Theory of generic equations, *Quart. Pure Appl. Math.* eleven paper in vols. 4–9, starting in 1861.

R. P. BOAS AND R. C. BUCK, "Polynomial Expansions of Analytic Functions," Springer, New York, 1964.

G. BOOLE, "A Treatise on Differential Equations," Chelsea, New York (reprinted from 1859).

G. BOOLE, A Treatise on the Calculus of Finite Differences, Dover, New York, 1960 (reprinted from 1872).

N. BOURBAKI, "Éléments de Mathématique, Première Partie," Livre IV, Chapitre IV–VII, Hermann, Paris, 1951.

F. BRAFMAN, Some generating functions for Laguerre and Hermite polynomials, *Canad. J. Math.* **9** (1957), 180–187.

J. W. BROWN, On zero type sets of Laguerre polynomials, *Duke Math. J.* **35** (1968), 821–823.

H. BUCHHOLTZ, "The Confluent Hypergeometric Functions," Springer, New York, 1969.

R. C. BUCK, "Studies in Modern Analysis," Prentice-Hall, Englewood Cliffs, NJ, 1962.

J. L. BURCHNALL, A note on the polynomials of Hermite, *Quart. J. Math.* **12** (1941), 9–11.

L. CARLITZ, On a class of finite sums, *Amer. Math. Monthly* **XXXVII** (1930), 472–479.

L. CARLITZ, On arrays of numbers, *Amer. J. Math.* **LIV** (1932), 749–752.

L. CARLITZ, A characterization of the Laguerre polynomials, *Monatsh. Math.* **66** (1952), 389–392.

L. CARLITZ, Note on a paper of Shanks, *Amer. Math. Monthly* **LIX** (1952), 239–242.

L. CARLITZ, Congruence properties of the polynomials of Hermite, Laguerre and Legendre, *Math. Z.* **59** (1954), 474–483.

L. CARLITZ, Congruences for generalized Bell and Stirling numbers, *Duke Math. J.* **22** (1955), 193–206.

L. CARLITZ, The product of certain polynomials analogous to the Hermite polynomials, *Amer. Math. Monthly* **LXIV** (1957), 723–725.

L. CARLITZ, The bilinear generating function for Hermite polynomials in several variables, *Math. Z.* **68** (1957), 284–289.

L. CARLITZ, A note on the Bessel polynomials, *Duke Math. J.* **24** (1957), 151–161.

L. CARLITZ, Note on some special polynomials, *Rev. Mat. Hisp. Amer.* **XVIII** (1958), 2–7.

L. CARLITZ, Some formulas of Jensen and Gould, *Duke Math. J.* **27** (1960), 319–322.

L. CARLITZ, A note on the Laguerre polynomials, *Michigan Math. J.* **7** (1960), 219–223.

L. CARLITZ, A bilinear generating function for the Hermite polynomials, *Duke Math. J.* **28** (1961), 531–536.

L. CARLITZ, On the product of two Laguerre polynomials, *J. London Math. Soc.* **36** (1961), 399–402.

L. Carlitz, Note on bilinear generating functions for the Laguerre polynomials, *Bull. UMI* **16** (1961), 24–30.
L. Carlitz, Characterizations of the Krawtchouk polynomials, *Rev. Mat. Hisp.-Amer.* **XXI** (1961), 77–84.
L. Carlitz, Some formulas of E. Feldheim, *Acta Math. Acad. Sci. Hungar.* **XIV** (1963), 21–29.
L. Carlitz, Products of Appell polynomials, *Collect. Math.* **XV** (1963), 246–258.
L. Carlitz, A binomial identity arising from a sorting problem, *SIAM Rev.* **6** (1964), 20–30.
L. Carlitz, Some multiple sums and binomial identities, *SIAM J.* **13** (1965), 469–486.
L. Carlitz, The coefficients of cosh $x/\cos x$, *Monatsh. Math.* **69** (1965), 129–135.
L. Carlitz, Multiple sums and generating functions, *Collect. Math.* **XVII** (1965), 281–296.
L. Carlitz, Note on Schur's expansion of sin πx, *Bull. UMI* **XXI** (1966), 353–357.
L. Carlitz, Some generating functions for Laguerre polynomials, *Duke Math. J.* **35** (1968), 825–828.
L. Carlitz, Some extensions of Mehler's formula, *Collect. Math.* **XXI** (1970), 117–130.
L. Carlitz, A conjecture concerning Genocchi numbers, *Kong. Norske Videnskabers Selskab, Skiffer* **9** (1971), 1–4.
B. C. Carlson, Polynomials satisfying a binomial theorem, *J. Math. Ann. Appl.* **32** (1970), 543–558.
A. M. Chak, A class of polynomials and a generalization of Stirling numbers, *Duke Math. J.* **23** (1956), 45–56.
C. A. Charalambides, A new kind of number appearing in the n-fold convolution of truncated binomial and negative binomial distributions, preprint.
M. Cotlar, Estudio de una clase de polynomios de Bernoulli, *Math. Notae* **VI** (1946), 69–95.
H. B. Curry, Abstract differential operators and interpolation formulas, *Portugal. Math.* **10** (1951), 135–162.
S. C. Dhar, On the product of parabolic cylinder functions, *Bull. Calcutta Math. Soc.* **26** (1934), 57–64.
M. Dwass, The distribution of a generalized D_n^+ statistic, *Amer. Math. Stat.* **30** (1959), 1024–1028.
A. Erdelyi (Ed.), "Higher Transcendental Functions," Vols. II and III, McGraw-Hill, New York, 1953 and 1955.
L. Euler, Institutiones calculi differentialis, St. Petersburg, 1755.
E. Feldheim, Quelques nouvelles relations pour les polynômes d'Hermite, *J. London Math. Soc.* **13** (1938), 22–29.
E. Feldheim, Une propriété charactéristique des polynômes de Laguerre, *Comm. Mat. Helv.* **13** (1940), 6–10.
E. Feldheim, Developpements en série de polynômes d'Hermite et de Laguerre à l'aide des transformations de Gauss et de Hankel, I, II, III, Kon. Nederlandsche Akad. van Wet. Proceedings **43** (1940), 224–239, 240–248, 379–386.
E. Feldheim, Relations entre les polynômes de Jacobi, Laguerre et Hermite, *Acta Math.* **74** (1941), 117–138.
J. L. Goldberg, A note on polynomials generated by $A(t)\,\psi[xH(t)]$, *Duke Math. J.* **32** (1965), 643–651.
H. W. Gould, Note on a paper of Sparre-Anderson, *Math. Scand.* **6** (1958), 226–230.
H. W. Gould, Stirling number representation problems, *Proc. Amer. Math. Soc.* **11** (1960), 447–451.

H. W. GOULD, A binomial identity of Greenwood and Gleason, *The Mathematics Student* **XXIX** (1961), 53–57.
H. W. GOULD, A series transformation for finding convolution identities, *Duke Math. J.* **28** (1961), 193–202.
H. W. GOULD, Note on a paper of Klamkin concerning Stirling numbers, *Amer. Math. Monthly* **68** (1961), 477–479.
H. W. GOULD, A new convolution formula and some new orthogonal relations for the inversion of series, *Duke Math. J.* **29** (1962), 393–404.
H. W. GOULD, Congruences involving sums of binomial coefficients and a formula of Jensen, *Amer. Math. Monthly* **69** (1962), 400–402.
H. W. GOULD, Note on two binomial coefficient sums found by Riordan, *Ann. Math. Stat.* **34** (1963), 333–335.
H. W. GOULD, On the recurrence relations for Stirling numbers, *Publ. Inst. Math. Biograd* **6** (1966), 115–119.
H. W. GOULD, "Combinatorial Identities," Morgantown, WV, 1972 (privately printed).
D. T. HAIMO, The Weierstrass-Laguerre transform, *J. Math. Anal. Appl.* **23** (1968), 41–57.
A. P. HILLMAN, P. L. MARIA, AND C. T. MCABEE, A symmetric substitute for the Stirling numbers, *Fibonacci Quart.* **9** (1971), 51–73.
H. HOCHSTADT, "Special Functions of Mathematical Physics," Holt, New York, 1961.
E. JABOTINSKI, Representation of functions by matrice, application to Faber polynomials, *Proc. Amer. Math. Soc.* **4** (1953), 546–553.
D. JACKSON, "Fourier Series and Orthogonal Polynomials," Mathematical Association of America, Buffalo, 1941.
K. JORDAN, "Chapters on the Classical Calculus of Probability," Akademiai Kiadó, Budapest, 1972.
B. KAUFMAN, Special functions of mathematical physics from the viewpoint of Lie algebras, *J. Math. Phys.* **7** (1966), 447–457.
G. S. KAZANDZIDIS, On the Bernoulli polynomials, *Bull. Soc. Math. Grèce* **10** (1969), 151–182.
M. G. KENDALL AND A. STUART, "The Advanced Theory of Statistics," Vol. I, Griffin, London, 1963.
M. KLAMKIN, On a generalization of the geometric series, *Amer. Math. Monthly* **LXIV** (1951), 91–93.
M. S. KLAMKIN AND D. J. NEWMAN, On the reducibility of some linear differential operators, *Amer. Math. Monthly* **66** (1959), 293–295.
D. KNUTH, "The Art of Computer Programming," Vols. 1 and 2, Addison-Wesley, Reading, PA, 1967.
D. KNUTH AND T. BUCHHOLTZ, Computation of tangent, Euler and Bernoulli numbers, *Math. Comp.* **21** (1967), 663–688.
R. LAGRANGE, Mémoire sur les Suites de Polynômes, *Acta Math.* **51** (1928), 201–309.
J. W. LAYMAN, Expansion of analytic functions in exponential polynomials, *Proc. Amer. Math. Soc.* **22** (1969), 519–522.
W. J. LEVEQUE (Ed.), "Studies in Number Theory," Prentice Hall, Englewood Cliffs, NJ, 1969.
Y. L. LUKE, "The Special Functions and their Approximations," two volumes, Academic Press, New York, 1969.
E. MCBRIDE, "Obtaining Generating Functions," Springer, New York, 1971.
W. MILLER, "Lie Theory and Special Functions," Academic Press, New York, 1968.

S. G. MOHANTY AND B. R. HANDA, Extensions of Vandermande type convolutions with several summations and applications I, *Canad. Math. Bull.* **12** (1969), 45–62.

N. MORRISON, Smoothing and extrapolation of continuous time series using Laguerre polynomials, *SIAM J. Appl. Math.* **16** (1968), 1280–1304.

R. MULLIN AND G.-C. ROTA, On the Foundations of Combinatorial Theory, III: Theory of Binomial Enumeration, in "Graph Theory and Its Applications" (B. Harris, Ed.), pp. 167–213, Academic Press, New York, 1970.

L. I. NEIKIRK, Some symbolic identities, *Bull. Amer. Math. Soc.* **43** (1937), 848–850.

D. J. NEWMAN AND H. S. SHAPIRO, Certain Hilbert spaces of entire functions, *Bull. Amer. Math. Soc.* **72** (1966), 971–977.

N. NIELSEN, Recherches sur les polynômes d'Hermite, Kgl. Dauske Vidensk. Selskab, Mat.-fys. Meddelelser (*J*) **6** (1918), 1–78.

N. NIELSEN, "Traité élémentaire des nombres de Bernoulli," Gauthier-Villars, Paris, 1923.

N. NIELSEN, "Die Gammafunktion," Chelsea, New York, 1965 (reprinted from 1906).

N. E. NÖRLUND, Leçons sur les séries d'interpolation, Gauthier-Villars, Paris, 1926.

N. E. NÖRLUND, "Vorlesungen ueber Differenzenrechnung," Chelsea, New York, 1954 (reprinted from 1923).

A. M. OSTROWSKI, Ueber das Restglied der Euler–MacLaurinschen Formul, *ISNM* **10** (1968), 358–364.

A. OSTROWSKI, Note on Poisson's treatment of the Euler–MacLaurin formulas, *Comm. Math. Helv.* **44** (1969), 202–206.

A. OSTROWSKI, On the remainder term of the Euler–MacLaurin formula, *Crelle's J.* **239/240** (1970), 268–286.

M. PARODI, Formules sommatoires pour les polynômes d'Hermite, *C. R. Acad. Sci. Paris* **271** (1970), 995–999.

S. PINCHERLE, Operatori lineari e coefficienti di fattoriali, *Alti Accad. Naz. Lincei, Rend. Cl. Fis. Mat. Nat.* (6) **XVIII** (1933), 417–519.

E. D. RAINVILLE, A discrete group arising in the study of differential operators, *Amer. J. Math.* **LXIII** (1941), 136–140.

E. D. RAINVILLE, Symbolic relations among classical polynomials, *Amer. Math. Monthly* **53** (1946), 299–305.

E. D. RAINVILLE, "Special Functions," MacMillan, New York, 1960.

R. REDHEFFER, Algebraic properties of certain integral transforms, *Amer. Math. Monthly* **73** (1966), 91–95.

J. RIORDAN, Inverse relations and combinatorial identities, *Amer. Math. Monthly* **71** (1964), 485–498.

H. SALIÈ, Ueber Abels Verallgemeinerung der binomischen Formul, *Ber. Ver. Sachs, Aka. Wiss. Leipzig, Math.-Nat. Kl.* **98** (1951), 19–22.

A. SALMERI, Introduzione alla teoria dei coefficienti fattoriali, *Giom. di Mat. Battaglini* **90** (1962), 44–54.

C. SCARAVELLI, Sui polinomi di Appell, *Riv. Mat. Univ. Parma* **6** (1965), 103–116.

I. J. SCHWATT, "Introduction to Operations with Series," Chelsea, New York, 1961 (reprinted from 1924).

E. B. SHANKS, Iterated sums of powers of the binomial coefficients, *Amer. Math. Monthly* **58** (1951), 404–407.

I. M. SHEFFER, A differential equation for Appell polynomials, *Bull. Amer. Math. Soc.* **41** (1935), 914–923.

409/42/3-17

I. M. SHEFFER, Note on Appell polynomials, *Bull. Amer. Math. Soc.* **51** (1969), 739–744.

J. SHOHAT, The relation of classical orthogonal polynomials to the polynomials of Appell, *Amer. J. Math.* **58** (1936), 453–464.

I. N. SNEDDON, "Special Functions ...," Oliver and Boyd, London, 1956.

J. F. STEFFENSEN, "Interpolation," Chelsea, New York, 1950 (reprinted from 1927).

M. SWEEDLER, "Hopf Algebras," Benjamin, New York, 1969.

J. P. TALMAN, "Special Functions," Benjamin, New York, 1968.

L. TOSCANO, Operatori differenziali e polinomi di Laguerre, *Le Matematiche* **XXIII** (1968), 197–223.

L. TOSCANO, Numeri di Sterling generalizzati e operatori permutabili di secondo ordine, *Le Matematiche* **XXIV** (1969), 492–518.

C. A. TRUESDELL, "An Essay Towards a Unified Theory of Speical Functions," Princeton Univ. Press, Princeton, NJ, 1948.

C. TWEEDIE, The Stirling numbers and polynomials, *Proc. Edinburgh Math. Soc.* (1) **37** (1919), 2–25.

M. WARD, A certain class of polynomials, *Ann. Math.* **31** (1930), 43–51.

G. N. WATSON, Notes on the generating functions of polynomials: (2) Hermite polynomials, *J. London Math. Soc.* **8** (1933), 194–199.

G. N. WATSON, A note on the polynomials of Hermite and Laguerre, *J. London Math. Soc.* **13** (1938), 29–32.

L. WEISNER, Group-theoretic origins of certain generating functions, *Pacific J. Math.* **4**(1955), 1033–1039.

L. WEISNER, Generating functions for Hermite functions, *Canad. J. Math.* **11** (1959), 141–147.

J. E. WHITTAKER, "Interpolatory Function Theory," Cambridge Univ. Press, London, 1935.

E. T. WHITTAKER AND G. N. WATSON, "A Course in Modern Analysis," Cambridge Univ. Press, 1952 (reprinted from 1902).

D. ZEITLIN, On convoluted numbers and sums, *Amer. Math. Monthly* **74** (1967), 235–246.

Studies in Applied Mathematics, Vol. LIII, No. 3, September 1974. Copyright © 1974 by The Massachusetts Institute of Technology

On the Foundations of Combinatorial Theory: IX Combinatorial Methods in Invariant Theory

By Peter Doubilet, Gian-Carlo Rota, and Joel Stein

Dedicated
to
Beniamino Segre

Contents

1. Introduction

We develop an algebraic system designed for computation with subspaces of a finite-dimensional vector space over an arbitrary field, based upon two operations, which we call join and meet. The join is the same as the wedge product in exterior algebra, and the meet roughly corresponds to Grassmann's regressive product, with one important difference. Whereas Grassmann and all other authors up to and including Bourbaki defined the regressive product by means of the duality of vector spaces, we introduce a special device which enables us to define the meet directly. This device is the notion of *Cayley space*, namely, a vector space endowed with a non-degenerate alternating multilinear form, called the *bracket*. It seems astonishing that this notion should not have been previously singled out, as it is the basic tool—recognized or not—of classical invariant theory. A Cayley space

* Supported by NSF Contract #P36739

should be thought of as a natural companion to Hilbert space and symplectic space.

The present definition leads to the derivation of a complete set of identities holding between join and meet, an undertaking that in the past would have been notationally impossible to carry out. We call these identities the *alternative laws.* The body of this work consists in various applications of the alternative laws. We show that these laws easily yield the classical identities holding among minors of a matrix, as well as a systematic procedure for translation of universal theorems of synthetic projective geometry into identities. The main application we derive of the alternative laws is the *straightening formula*; this can be considered to be the end product and the definitive version of a train of thought which began with Clebsch, was developed by Gordan and Capelli, and later by Young and Turnbull. The straightening formula can be interpreted as giving the solution of a word problem. It is a central result in the characteristic-free theory of the projective group; in fact it holds over commutative rings.

As an application of the straightening formula we obtain a characteristic-free version of the classical theory of representations of the symmetric group, as well as two elementary proofs of the First Fundamental Theorem of invariant theory over arbitrary fields. The only previous work on this subject is Igusa's.

Various other applications, which we hope to further develop elsewhere, are sketched throughout the paper. These will include a thorough treatment of classical invariant theory over arbitrary fields, as well as of the symmetric group.

2. Cayley spaces

Throughout this work V will denote a vector space over an arbitrary field. A *bracket*, written

$$[x_1, \ldots, x_n], \quad \text{where } x_i \in V,$$

is a non-degenerate (that is, not identically zero) multilinear alternating form, taking values in the field.

A *Cayley space* is the pair consisting of the vector space V, together with a bracket.

A *standard Cayley space* is a Cayley space over a vector space V of dimension n, whose bracket has the additional property that for every vector x in V, there exist vectors $x_2, \ldots, x_n$ such that

$$[x, x_2, \ldots, x_n]$$

is not equal to zero. In a standard Cayley space the length of the bracket equals the dimension of the space, and conversely. Unless otherwise stated, all Cayley spaces occurring in this work will be standard.

The *exterior algebra* of a standard Cayley space is constructed by imposing an equivalence relation on sequences of vectors. Given two sequences of vectors of length k, we shall write

$$a_1 \ldots a_k \sim b_1 \ldots b_k$$

when for every choice of the vectors $x_{k+1}, \ldots, x_n$ we have

$$[a_1, \ldots, a_k, x_{k+1}, \ldots, x_n] = [b_1, \ldots, b_k, x_{k+1}, \ldots, x_n].$$

An equivalence class under this relation will be called an *extensor*, or *decomposable k-vector*, and will be written as

$$a_1 \vee a_2 \vee \cdots \vee a_k.$$

The operation $\vee$ is called the *join* (and is elsewhere written $\wedge$; our departure from customary notation is well motivated). Note that the join is non-zero if and only if $\{a_1, \ldots, a_k\}$ is a linearly independent set.

A non-zero extensor is of *step k* if it is the join of k linearly independent vectors. If it is of step zero it is called a *scalar*.

The extensors of V span a vector space of dimension 2^n, called $\bar{V}$, whose elements are called *antisymmetric tensors*. The algebra of $\bar{V}$ together with join is the *exterior algebra* of V. It is an antisymmetric associative algebra with identity (the scalar, one) with the usual properties which will not be recalled here.

The extensors of step n form a one dimensional sub-space of $\bar{V}$. Choosing a basis $\{a_1, \ldots, a_n\}$ of V, whose bracket $[a_1, \ldots, a_n]$ equals one, or a *unimodular basis*, we may construct a basis for this subspace, the element

$$E = a_1 \vee \cdots \vee a_n.$$

E is called the *integral*.

We shall frequently indicate the join of extensors by simple juxtaposition of symbols,

$$ab = a \vee b.$$

Also, if A and B denote two extensors the sum of whose steps is n, we shall write $[A, B] = [AB]$ for their bracket.

Every extensor A defines a unique subspace of the vector space V, namely

$$\bar{A} = \operatorname{span}\{a_1, \ldots, a_k\}$$

where $\{a_1, \ldots, a_k\}$ is any set of vectors such that

$$a_1 \vee \cdots \vee a_k = A.$$

The subspace $\bar{A}$ is called the *support* of A. If A and B are extensors, then $A \vee B$ is non-zero if and only if $\bar{A} \cap \bar{B} = 0$, in which case the support of $A \vee B$ is the sub-space $\bar{A} \cup \bar{B}$ spanned by $\bar{A}$ and $\bar{B}$.

A linear transformation T of V into itself is said to be *unimodular* if it preserves the bracket.

Given an extensor A of step n–k in a standard Cayley space, we define the *bracket relative to A* by

$$[x_1 \ldots x_k]_A = [x_1 \ldots x_k A].$$

A relative bracket is an alternating k-linear form on a vector space of dimension n. Conversely, any alternating k-linear form on an n-dimensional vector space defines a unique relative bracket. The pair consisting of a vector space V with a relative bracket is a non-standard Cayley Space, called the *contraction* of the standard Cayley Space by the extensor A.

In a non-standard Cayley space on V, a vector in V is said to be of *rank zero* when for all choices of the vectors $x_1, \ldots, x_{k-1}$ in V

$$[x, x_1, \ldots, x_{k-1}] = 0.$$

Otherwise it is said to be of *rank one*.

3. Splits and shuffles

A *split* of the linearly ordered set, or sequence $A = a \ldots bc \ldots de \ldots f$ is a partition of A into blocks which are intervals of A, namely

$$B_1 = (a, \ldots, b), \qquad B_2 = (c, \ldots, d), \ldots, \qquad B_k = (e, \ldots, f).$$

If B_j contains i_j elements for each j we call the split the $(i_1, \ldots, i_k)$-*split* of A.

A *shuffle* of the $(i_1, \ldots, i_k)$-split of A is a permutation $\sigma : A \to \sigma(A)$ of the elements of A with the property that each block of the $(i_1, \ldots, i_k)$-split of $\sigma(A)$ is a subsequence of A. That is, the linear order of A is preserved in each block of $\sigma(A)$.

A *bracket product* is an expression of the form

$$[a_1 \ldots a_n][b_1 \ldots b_n] \ldots [c_1 \ldots c_n] d_1 \vee \ldots \vee d_p$$

for some arbitrary number of brackets.

Let

$$a_1 \ldots a_i \quad b_1 \ldots b_j \ldots c_1 \ldots c_k \quad d_1 \ldots d_m$$

denote a subsequence of the vectors in a bracket product. We define the *split-sum* of their $(i, j, \ldots, k, m)$-split as the expression

$$\sum_\sigma \operatorname{sgn}(\sigma)[\sigma(a_1) \ldots \sigma(a_i) a_{i+1} \ldots a_n][\sigma(b_1) \ldots \sigma(b_j) b_{j+1} \ldots b_n] \ldots$$
$$\times [\sigma(c_1) \ldots \sigma(c_k) c_{k+1} \ldots c_n] \sigma(d_1) \ldots \sigma(d_m) d_{m+1} \ldots d_p$$

where the sum ranges over all shuffles of the above split. Alternatively, we write this as

$$[a_1^\sigma \ldots a_i^\sigma \quad a_{i+1} \ldots a_n][b_1^\sigma \ldots b_j^\sigma \quad b_{j+1} \ldots b_n] \ldots$$
$$[c_1^\sigma \ldots c_k^\sigma \quad c_{k+1} \ldots c_n] d_1^\sigma \ldots d_m^\sigma \quad d_{m+1} \ldots d_p.$$

The split-sum is thus formed by applying to the sequence of variables marked by the superscript σ in a bracket product, the shuffles of the split whose blocks are determined by the brackets.

One can iterate a split-sum. When the sets are disjoint, iteration reduces to an interchangeable double summation. In the general case, split-sums are not commutative.

As an example,

$$[a^\theta b^\theta c^\sigma d^\sigma ef][g^\theta h^\sigma i^\sigma jkl]$$

denotes the split-sum of the (2, 1)-split of a, b, g either followed or preceded by the split-sum of the (2, 2)-split of c, d, h, i. However,

$$[a^{\theta\sigma} b^{\theta\sigma} c^\theta def][g^{\theta\sigma} h^\theta i^\theta jkl]$$

denotes the (non-commuting) split-sum of the (3, 3)-split of a, b, c, g, h, i followed by the split-sum of the (2, 1)-split of the sequence $\theta(a)$, $\theta(b)$, $\theta(g)$.

In a single split-sum, we often replace the superscripts by dots. Thus,

$$[a^\sigma b^\sigma cd][e^\sigma fgh] = [\dot{a}\dot{b}cd][\dot{e}fgh].$$

The use of dots to indicate split-sums will be called the *Scottish Convention* after H. W. Turnbull who used it informally.

4. Cayley algebras

We now define a second operation on a Cayley space, called the *meet*. Let $A = a_1 \dots a_k$ and $B = b_1 \dots b_p$ be extensors of indicated steps satisfying $k + p \geq n$. We define their *meet*

$$A \wedge B = a_1 \dots a_k \wedge b_1 \dots b_p$$

by the expression

$$A \wedge B = \sum_{\sigma} \operatorname{sgn}(\sigma)[a_{\sigma(1)}a_{\sigma(2)} \dots a_{\sigma(n-p)}b_1 \dots b_p]a_{\sigma(n-p+1)} \dots a_{\sigma(k)},$$

where σ ranges over all shuffles of the $(n - p, k - n + p)$-split of $a_1 \dots a_k$. Alternatively, we may write this as

$$A \wedge B = [\dot{a}_1 \dots \dot{a}_{(n-p)}b_1 \dots b_p]\dot{a}_{(n-p)+1} \dots \dot{a}_k,$$

where the dots indicate the split-sum of the $(n - p, k - (n - p))$-split of $a_1 \dots a_k$. If $k + p < n$ the meet is defined to be zero and in either case it is extended by linearity to all linear combinations of extensors.

PROPOSITION. *The meet satisfies the identity*

$$[\dot{a}_1 \dots \dot{a}_{(n-p)}b_1 \dots b_p]\dot{a}_{(n-p)+1} \dots \dot{a}_k = \dot{b}_1 \dots \dot{b}_{p-(n-k)}[a_1 \dots a_k\dot{b}_{p-(n-k)+1} \dots \dot{b}_p].$$

The verification is a simple consequence of the alternating property of the bracket.

THEOREM 1. *The meet is associative and anticommutative following the rule*

$$B \wedge A = (-1)^{(n-p)(n-k)}A \wedge B,$$

where A is an extensor of step k, and B of step p.

The verification is a straightforward computation.

The *Cayley algebra* of a Cayley space is the algebraic structure obtained by endowing the exterior algebra with the additional operations of bracket and meet. Thus, a Cayley algebra is the vector space $\bar{V}$ endowed with three operations in the sense of universal algebra: meet, join and bracket.

COROLLARY. *The integral E is an identity for meet in the Cayley algebra, that is,*

$$E \wedge A = A \wedge E = A$$

for all A.

The meet of two extensors has an important geometric interpretation:

PROPOSITION. *If A and B are extensors of step k and p, supporting subspaces $\bar{A}$ and $\bar{B}$ of a standard Cayley space over V, and the span $\bar{A} \cup \bar{B}$ equals V, then the meet $A \wedge B$ supports the intersection $\bar{A} \cap \bar{B}$.*

Proof: Take a basis $e_1, \dots, e_n$ of V such that $e_1, \dots, e_r$ is a basis of $\bar{A} \cap \bar{B}$, $e_1, \dots, e_k$ a basis of $\bar{A}$ and $e_{k+1}, \dots, e_n, e_1, \dots, e_r$ a basis for $\bar{B}$. We may therefore write, for some scalars c and d,

$$A = c\, e_1 \dots e_k$$

$$B = d\, e_1 \dots e_r e_{k+1} \dots e_n.$$

Expanding $A \wedge B$, we get

$$A \wedge B = cd[e_1 \dots e_n]e_1 \dots e_r, \qquad \text{q.e.d.}$$

COROLLARY. *The meet of two or more extensors is an extensor.*

5. Duality

Let $A_1, \dots, A_n$ be extensors of step $n - 1$ in V, which we call *covectors*. We define a new alternating multilinear form on covectors in V, called the *double bracket*, by setting

$$[[A_1, \dots, A_n]] = A_1 \wedge \cdots \wedge A_n.$$

We infer from the properties of meet that the double bracket is non-degenerate and of step zero (that is, a scalar). Thus, since the vector space spanned by covectors is of dimension n, the double bracket defines a Cayley space on covectors. The associated Cayley algebra is called the *dual Cayley algebra.* A Cayley algebra and its dual are isomorphic. The role of join and meet are interchanged under the canonical isomorphism.

A set of covectors $A_1, \dots, A_n$ with non-zero double bracket constitutes a basis of covectors. In such a case, a corresponding basis of vectors $a_1, \dots, a_n$ can always be found satisfying

$$A_i = a_1 \dots \hat{a}_i \dots a_n,$$

where $\hat{a}_i$ indicates that a_i is deleted. It is verified in Section 7 that

$$[[A_1, \dots, A_n]] = [a_1, \dots, a_n]^{n-1},$$

an identity known as Cauchy's theorem on the adjugate. By duality and Cauchy's theorem, we may construct from every identity between joins and meets, another identity where the roles of join and meet are interchanged, step k is replaced by step $n - k$, and suitable powers of the bracket appear as multipliers to restore homogeneity.

For example, if A is an extensor of step k and the b_j are covectors, the identity

$$A \vee (b_1 \wedge \cdots \wedge b_{p+k}) = (\dot{b}_1 \wedge \cdots \wedge \dot{b}_p) \vee (A \wedge \dot{b}_{p+1} \wedge \cdots \wedge \dot{b}_{p+k})$$

is immediate, as it is the dual of the identity

$$B \wedge (a_1 \vee \cdots \vee a_{p+k}) = (\dot{a}_1 \vee \cdots \vee \dot{a}_p) \wedge (B \vee \dot{a}_{p+1} \vee \cdots \vee \dot{a}_{p+k}),$$

where B is an extensor of step $n - k$ and the a_i are vectors.

The *principle of complementary minors* which associates with every identity holding among the minors of a matrix another identity holding among the complementary minors of the adjugate matrix, is a special case of the duality between joins and meets.

By introducing the analogue of the contraction of a bracket by an extensor A, in the dual Cayley algebra, we may construct in the given Cayley algebra the dual operation, called the *co-contraction* or *reduction* by A. Thus, if A is of step k and the x_i are covectors, write A as the meet of $n - k$ covectors and define the reduction by A as

$$\begin{aligned}[x_1, \dots, x_k]^A &= [[x_1, \dots, x_k]]_A \\ &= [[x_1, \dots, x_k, A]].\end{aligned}$$

The notions of contraction and reduction in the Cayley algebra correspond roughly to the meanings these terms have in combinatorial geometry.

6. Identities in the Cayley algebra

We present a sampling of identities which describe how joins and meets are distributed through each other, or *alternative laws.*

We begin with some notation. Juxtaposition of vectors denotes join and juxtaposition of covectors denotes meet. The *inner product* of a vector a and a covector x is defined as

$$\langle a|x\rangle = a \wedge x.$$

Similarly, if extensors $A = a_1 \dots a_k$ and $X = x_1 \wedge \cdots \wedge x_k$ are given, where the a_i are vectors and the x_i are covectors, we define their inner product of length k as

$$\begin{aligned}\langle A|X\rangle &= \langle a_1 \dots a_k|x_1 \dots x_k\rangle \\ &= (a_1 \dots a_k) \wedge (x_1 \dots x_k).\end{aligned}$$

THEOREM 6.1. *Let $a_1, \dots, a_k$ be vectors and $x_1, \dots, x_s$ be covectors.*
If $k \geq s$, then

$$(a_1 \dots a_k) \wedge (x_1 \dots x_s) = \langle \dot{a}_1|x_1\rangle \dots \langle \dot{a}_s|x_s\rangle \dot{a}_{s+1} \dots \dot{a}_k;$$

If $k < s$, then

$$(a_k \dots a_1) \vee (x_s \dots x_1) = \dot{x}_s \dots \dot{x}_{k+1}\langle a_k|\dot{x}_k\rangle \dots \langle a_1|\dot{x}_1\rangle.$$

Proof: We verify the first identity. From the definition of meet,

$$\begin{aligned}(a_1 \dots a_k) \wedge (x_1 \dots x_s) &= \langle a_1^\sigma|x_1\rangle(a_2^\sigma \dots a_k^\sigma) \wedge (x_2 \dots x_s) \\ &= \langle a_1^\sigma|x_1\rangle\langle a_2^{\sigma\theta}|x_2\rangle(a_3^{\sigma\theta} \dots a_k^{\sigma\theta}) \wedge (x_3 \dots x_s) \\ &\;\;\vdots\end{aligned}$$

Here σ ranges over the split-sum of the $(1, k-1)$-split of $a_1 \dots a_k$, θ ranges over the split-sum of the $(1, k-2)$-split of $a_{\sigma(2)} \dots a_{\sigma(k)}$, and so forth. But by an elementary coset argument this is equal to the split-sum of the $(1, \dots, 1, k-s)$-split of $a_1 \dots a_k$.

THEOREM 6.2. *Let $a_1, \dots, a_k$ be vectors and $x_1, \dots, x_s$ be covectors.*
If $k \geq s$, then

$$(a_1 \dots a_k) \wedge (x_1 \dots x_s) = \langle \dot{a}_1 \dots \dot{a}_i|x_1 \dots x_i\rangle \dots \times \langle \dot{a}_{i+\dots+j+1} \dots \dot{a}_s|x_{i+\dots+j+1} \dots x_s\rangle \dot{a}_{s+1} \dots \dot{a}_k.$$

If $k < s$, then

$$(a_k \dots a_1) \vee (x_s \dots x_1) = \dot{x}_s \dots \dot{x}_{k+1}\langle a_k \dots a_{i+\dots+j+1}|\dot{x}_k \dots \dot{x}_{i+\dots+j+1}\rangle \dots \langle a_i \dots a_1|\dot{x}_i \dots \dot{x}_1\rangle.$$

Proof: By the associative law,

$$(a_1 \dots a_k) \wedge (x_1 \dots x_s) = (a_1 \dots a_k) \wedge (x_1 \dots x_i) \wedge \cdots \wedge (x_{i+\dots+j+1} \cdots x_s)$$

whence proceed as in Theorem 6.1. The second expression is derived similarly.

COROLLARY 1. *Let $C_1, \ldots, C_s$ be extensors of step $n - i, \ldots, n - j, n - l$ and let $k = i + \cdots + j + l$. Then*

$$a_1 \ldots a_k \wedge (C_1 \wedge \cdots \wedge C_s) = [\dot{a}_1 \ldots \dot{a}_i C_1] \ldots [\dot{a}_{i+\cdots+j+1} \ldots \dot{a}_k C_s]$$

If $A = a_1 \ldots a_k$ and $X = x_1 \ldots x_s$ are vector and covector decompositions of flats we shall sometimes employ the notation

$$\dot{A} = \dot{a}_1 \ldots \dot{a}_k \qquad \text{and} \qquad \dot{X} = \dot{x}_1 \ldots \dot{x}_s.$$

COROLLARY 2. *Let A_j and X_j be extensors of complementary step for each j. Then*

$$(A_1 \vee \cdots \vee A_{k+1}) \wedge (X_1 \wedge \cdots \wedge X_k) = \langle \dot{A}_1 | X_1 \rangle \ldots \langle \dot{A}_k | X_k \rangle \dot{A}_{k+1}$$

$$(A_k \vee \cdots \vee A_1) \vee (X_{k+1} \wedge \cdots \wedge X_1) = \dot{X}_{k+1} \langle A_k | \dot{X}_k \rangle \ldots \langle A_1 | \dot{X}_1 \rangle$$

THEOREM 6.3. *Let $A_k, B_l, C_p, D_q, X_{n-(k+l)}$, and $Y_{n-(p+q)}$ be extensors of indicated steps. Then*

$$(A \vee X) \wedge BC \wedge (D \vee Y) = \pm((A \vee \dot{B}) \wedge X) \vee ((\dot{C} \vee D) \wedge Y).$$

Proof:

$$\begin{aligned}(A \vee X) \wedge BC \wedge (D \vee Y) &= \pm[A\dot{B}X]\dot{C} \wedge (D \vee Y)\\ &= \pm((A \vee \dot{B}) \wedge X)\dot{C} \wedge (D \vee Y)\\ &= \pm((A \vee \dot{B}) \wedge X)[\dot{C}DY]\\ &= \pm((A \vee \dot{B}) \wedge X) \vee ((\dot{C} \vee D) \wedge Y).\end{aligned}$$

THEOREM 6.4. *If $A \vee B$ is of step n, then*

$$A \vee B = (A \wedge B) \vee E$$

where E is the integral.

The proof is a simple verification.

We now present the main result of this section.

THEOREM 6.5. *Let $C^{(1)}, \ldots, C^{(r)}$ be extensors of step $n - q_1, \ldots, n - q_r$ and let $k + s = q_1 + \cdots + q_r$. Then*

$$(a_1 \ldots a_k b_1 \ldots b_s) \wedge C^{(1)} \wedge \cdots \wedge C^{(r)} = (b_1 \ldots b_s) \vee \sum_{i_1 + \cdots + i_r = s} (-)^{(i_1, \ldots, i_r)}$$

$$\times \{\dot{a}_1 \ldots \dot{a}_{q_1 - i_1} C^{(1)}\} \wedge \cdots \wedge \{\dot{a}_{[q_1 - i_1 + \cdots + q_{r-1} - i_{r-1}] + 1} \ldots \dot{a}_k C^{(r)}\}$$

where the integer $(i_1, \ldots, i_r)$ is specified below.

Proof: For simplicity of notation take $s < q_r$. By Theorem 6.2, we have, calling the left side I,

$$I = [\dot{a}_1 \ldots \dot{a}_{q_1} C^{(1)}][\dot{a}_{q_1+1} \ldots \dot{a}_{q_1+q_2} C^{(2)}] \ldots [\dot{a}_{[q_1+\cdots+q_{r-1}]+1} \ldots \dot{a}_k \dot{b}_1 \ldots \dot{b}_s C^{(r)}].$$

The permutations acting in this equation may be separated into classes according to their effect on the b's. Thus, a given permutation positions, say, i_1 of the b's in the bracket containing $C^{(1)}, \ldots, i_r$ of the b's in the bracket containing $C^{(r)}$. Affixing the appropriate sign,

$$I = \sum_{i_1+\cdots+i_r=s} (-)^{(i_1,\ldots,i_r)}[b_1^\theta \ldots b_{i_1}^\theta a_1^\sigma \ldots a_{q_1-i_1}^\sigma C^{(1)}]$$
$$\times [b_{i_1+1}^\theta \ldots b_{[i_1+i_2]}^\theta a_{[q_1-i_1]+1}^\sigma \ldots a_{[q_1-i_1+q_2-i_2]}^\sigma C^{(2)}] \ldots$$
$$\times [b_{[i_1+\cdots+i_{r-1}]+1}^\theta \ldots b_s^\theta a_{[q_1-i_1+\cdots+q_{r-1}-i_{r-1}]+1}^\sigma \ldots a_k C^{(r)}].$$

Here θ ranges over the split-sum of the $(i_1, \ldots, i_r)$-split of $b_1 \ldots b_s$, and σ ranges over the split-sum of the $(q_1 - i_1, \ldots, q_r - i_r)$-split of $a_1 \ldots a_k$.

We first evaluate $(i_1, \ldots, i_r)$:

$$\begin{aligned}(i_1, \ldots, i_r) &= i_1(q_1 + \cdots + q_r - s) + i_2(q_2 + \cdots + q_r - s + i_1) + \cdots + i_r(q_r - s + i_1 \ldots + i_{r-1})\\ &= q_1(i_1) + q_2(i_1 + i_2) + \cdots + q_r(i_1 + \cdots + i_r) - s(i_1 + \cdots + i_r)\\ &\quad + i_2(i_1) + i_3(i_2 + i_1) + \cdots + i_r(i_{r-1} + \cdots + i_1)\\ &= q_1(i_1) + q_2(i_1 + i_2) + \cdots + q_r(i_1 + \cdots + i_r) + h_2(i_1 \ldots i_r)\end{aligned}$$

where $h_2(i_1 \ldots i_r)$ is the homogeneous symmetric function of degree two on $i_1, \ldots, i_r$.

We now factor out the b's using Theorem 6.2. This gives the desired identity.

We conclude this Section with two examples which illustrate the correspondence of theorems of projective geometry with identities in Cayley algebras.

DESARGUES'S THEOREM. *The corresponding sides of two collinear triangles intersect in collinear points if and only if the joins of corresponding vertices are concurrent.*

Proof: Let a, b, c be vectors and x, y, z be covectors in a Cayley space of three dimensions. Juxtaposition of vectors denotes join and juxtaposition of covectors denotes meet. The identity

$$abc \wedge [(a \vee yz) \wedge (b \vee zx) \wedge (c \vee xy)] = xyz \wedge [(bc \wedge x) \vee (ca \wedge y) \vee (ab \wedge z)]$$

is easily verified. Now let $x = b'c'$, $y = c'a'$, $z = a'b'$ so that $xyz = [a'b'c']^2$. This gives

$$[(bc \wedge b'c') \vee (ca \wedge c'a') \vee (ab \wedge a'b')] = [(aa') \wedge (bb') \wedge (cc')][abc][a'b'c'].$$

Desargues's theorem for triangles whose vertices are a, b, c and a', b', c' is then the statement that one side of this identity is zero if and only if the other side is zero.

PAPPUS' THEOREM. *If a, b, c are collinear, and a', b', c' are collinear and if all six points are distinct, then $ab' \wedge a'b$, $bc' \wedge b'c$, and $ca' \wedge c'a$ are also collinear.*

Proof: The theorem is a restatement of the identity

$$\begin{aligned}&(bc' \wedge b'c) \vee (ca' \wedge c'a) \vee (ab' \wedge a'b)\\ &\quad = [aa'b'][bb'c'][cc'a'][abc] - [abb'][bcc'][caa'][a'b'c'].\end{aligned}$$

Note that the algebraic version of each of these theorems is the stronger one, as it includes the geometric result as well as degeneracies.

7. Determinant identities

Identities between minors of matrices find elegant verification in the language of Cayley algebras. We illustrate with some examples.

Let $\{e_1, \ldots, e_n\}$ be a unimodular basis of vectors. With it we associate a basis of covectors $\{1, \ldots, n\}$ by setting $j = x_j = e_1 \ldots \hat{e}_j \ldots e_n$.* Thus any extensor A of step k may be uniquely expressed as a linear combination of monomials of the form $i_1 \ldots i_{n-k}$, where $i_1 < \cdots < i_{n-k} \in I$ and juxtaposition indicates meet. It is easily verified that $\{1, \ldots, n\}$ is also unimodular, that is, that $1 \ldots n$ is equal to unity.

Given an extensor $A = a_1 \ldots a_n$ of step n we may re-express its determinant $[a_1 \ldots a_n]$ in coordinate form by applying the alternative laws to $A \wedge 1 \ldots n$:

$$A \wedge 1 \ldots n = [a_1 \ldots a_n] = \langle a_1 | \dot{i} \rangle \ldots \langle a_n | \dot{n} \rangle$$

where $\langle a_i | j \rangle = a_i \wedge j$ is the j-th coordinate of a_i relative to $e_1, e_2, \ldots, e_n$.

A similar procedure may be used to coordinatize a flat of any step. Thus, if A is of step k we may write

$$A = A \wedge e_1 \ldots e_n = \dot{e}_1 \ldots \dot{e}_k (A \wedge \dot{e}_{k+1} \ldots \dot{e}_n)$$

or

$$A = A \vee 1 \ldots n = \dot{x}_1 \dot{x}_2 \ldots \dot{x}_{n-k} (A \wedge \dot{x}_{n-k+1} \ldots \dot{x}_n).$$

The first expansion represents a covariant coordinatization while the second represents the associated contravariant coordinatization. The numerical coefficients occurring in these expansions are the well known Plücker coordinates of the flat relative to the indicated basis.

Given a determinant $\Delta = [a_1, a_2, \ldots, a_n]$, the *adjugate* of Δ is the determinant

$$\Delta^* = \bar{a}_n \wedge \bar{a}_{n-1} \wedge \cdots \wedge \bar{a}_1$$

where $\bar{a}_i = a_1 \ldots \hat{a}_i \ldots a_n$. The adjugate is thus the determinant of $(n-1) \times (n-1)$ minors of Δ. Many determinant identities describe the relationships between these two determinants.

We begin with the expansion of Δ due to Laplace.

(1) *The Laplace expansion*: This describes how to expand Δ in terms of the set of minors of Δ in a given subset of $\{a_1, \ldots, a_n\}$. Thus by Theorem 6.2,

$$\begin{aligned} \Delta &= a_1 \ldots a_n \wedge 1 \ldots n \\ &= (a_1 \ldots a_n) \wedge (1 \ldots k) \wedge (k+1 \ldots n) \\ &= \langle a_1 \ldots a_k | \dot{1} \ldots \dot{k} \rangle \langle a_{k+1} \ldots a_n | k \dot{+} 1 \ldots \dot{n} \rangle. \end{aligned}$$

The Laplace expansion is thus a consequence of one of the alternative laws.

(2) *Cauchy's Theorem on the adjugate*: The adjugate is the $(n-1)$th power of the original determinant. By the associative law for meet,

$$\begin{aligned} \Delta^* &= (-)^{n(n-1)/2} \bar{a}_1 \wedge \cdots \wedge \bar{a}_n \\ &= (-)^{n(n-1)/2} (a_2 \ldots a_n) \wedge (a_1 a_3 \ldots a_n) \wedge \cdots \wedge (a_1 \ldots a_{n-1}) \\ &= -(-)^{n(n-1)/2} [a_1 \ldots a_n] (a_3 \ldots a_n) \wedge (a_1 a_2 a_4 \ldots a_n) \wedge \cdots \wedge (a_1 \ldots a_{n-1}) \\ &\;\;\vdots \\ &= \Delta^{n-1}. \end{aligned}$$

* Note our unconventional usage of integers as variables.

(3) *Jacobi's Theorem on the adjugate*: A minor of order r of the adjugate is equal to the complementary minor in the original determinant multiplied by the $(n-r-1)$th power of Δ.

We illustrate with the case $r = 2$. Consider the identity

$$\begin{aligned}\bar{a}_3 \wedge \cdots \wedge \bar{a}_n &= a_1 a_2 \hat{a}_3 \ldots a_n \wedge a_1 a_2 a_3 \hat{a}_4 \ldots a_n \wedge \cdots \wedge a_1 a_2 \ldots \hat{a}_n \\ &= (-)^{(n-2)(n-3)/2} a_1 a_2 [a_1 a_2 \dot{a}_3 \ldots a_n] \ldots [a_1 \ldots a_{n-1} \dot{a}_n] \\ &= (-)^{(n-2)(n-3)/2} a_1 a_2 \Delta^{n-3}.\end{aligned}$$

Now meet both sides with

$$ij = e_1 \ldots \hat{e}_i \ldots \hat{e}_j \ldots e_n,$$

This gives

$$\langle \bar{a}_n \ldots \bar{a}_3 | e_1 \ldots \hat{e}_i \ldots \hat{e}_j \ldots e_n \rangle = \Delta^{n-3} \langle a_1 a_2 | ij \rangle,$$

which is the desired result.

(4) *The Bazin–Reiss–Picquet Identity*: Starting with Cauchy's theorem on the adjugate, meet both sides with $a \quad b \ldots c \quad a_{k+1} \ldots a_n$. This gives

$$\begin{aligned}&[a \quad b \ldots c \quad a_{k+1} \ldots a_n][a_1 \ldots a_n]^{n-1} \\ &= [\dot{a} \quad a_2 \ldots a_n][a_1 \dot{b} \quad a_3 \ldots a_n][a_1 a_2 \ldots \dot{c} \quad a_{k+1} \ldots a_n][a_1 \ldots \dot{a}_{k+1} \ldots a_n] \ldots [a_1 \ldots \dot{a}_n]\end{aligned}$$

so that

$$[a \quad b \ldots c \quad a_{k+1} \ldots a_n]\, \Delta^{k-1} = [\dot{a} a_2 \ldots a_n][a_1 \dot{b} \ldots a_n] \ldots [a_1 \ldots \dot{c} \ldots a_n],$$

as desired.

(5) *Sylvester's Theorem on Compound Determinants*: Form the set of monomials $a_{i_1} \ldots a_{i_k}$ where $i_1 < \cdots < i_k$ from the sequence $\{a_1, \ldots, a_n\}$ and order them lexicographically as $\{A_1, \ldots, A_{\binom{n}{k}}\}$. Also, let the set $\{X_1, \ldots, X_{\binom{n}{k}}\}$ be formed from the set $\{1, \ldots, n\}$ of covectors, in the same way. The determinant

$$\Delta_k = \langle A_1 | \dot{X}_1 \rangle \ldots \langle A_{\binom{n}{k}} | \dot{X}_{\binom{n}{k}} \rangle$$

is called the k-th compound of Δ. Sylvester's theorem states that $\Delta_k = \Delta^{\binom{n-1}{k-1}}$.

We illustrate the method for the case $n = 4$ and $k = 2$, so that $\binom{n-1}{k-1} = 3$.

By Cauchy's theorem,

$$\begin{aligned}[abcd]^3 &= (abc) \wedge (abd) \wedge (acd) \wedge (bcd) \\ &= (ab \quad [acbd]) \wedge ([adbc] \quad cd) \\ &= (ab \vee (ac \wedge bd)) \wedge ((ad \wedge bc) \vee cd).\end{aligned}$$

Similarly,

$$[1234]^3 = (12 \vee (13 \wedge 24)) \wedge ((14 \wedge 23) \vee 34)$$

Now substitute for $[abcd]^3$ and $[1234]^3$ on the left hand side of

$$[abcd]^3 \vee [1234]^3 = [abcd]^3$$

and expand the resulting expression by the alternative laws. This gives the result.

Sylvester's identity shows how to construct a Cayley space on the extensors of step k.

8. The Straightening Formula

We now derive the basic result of the theory of Cayley algebras. In its simplest form, it can be viewed as stating that a set of vectors is a basis of a certain vector space. It can also be interpreted as the solution to a word problem in the Cayley algebra, (see Section 12).

Our main application of the Straightening Formula is a characteristic-free proof of the First Fundamental Theorem of invariant theory. We also sketch applications to the classification of identities in associative algebras and to the theory of symmetric functions.

Some of the results below can be extended to spaces of arbitrary dimensions, but we have preferred to preserve the more elegant approach by Cayley algebras. The finite-dimensional case proved here is actually the stronger.

Let K be a field of arbitrary characteristic and let R_K be the polynomial ring over K obtained by adjoining mn transcendentals $(a_i|x_j)$ where $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.

Let $\alpha = (\alpha_1, \ldots, \alpha_m)$ and $\beta = (\beta_1, \ldots, \beta_n)$ be sequences of non-negative integers. We define

$$V_{\alpha,\beta}$$

to be the vector space over K spanned by all monomials in the $(a_i|x_j)$ which contain α_i occurences of a_i and β_j occurrences of x_j, or all monomials of *content* (α, β) for short.

A *double tableau* of content (α, β) is denoted by the double matrix

$$T = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1\lambda_1} & x_{11} & \cdots & x_{1\lambda_1} \\ \vdots & & & \vdots & & \\ a_{s1} \ldots a_{s\lambda_s} & & & x_{s1} \ldots x_{s\lambda_s} & & \end{array}\right)$$

where $n \geq \lambda_1 \geq \cdots \geq \lambda_s$ and where the elements a_{ij} of the left tableau are chosen from $\{a_1, \ldots, a_m\}$ and the elements x_{ij} of the right tableau are chosen from $\{x_1, \ldots, x_n\}$, such that each a_i occurs with multiplicity α_i and each x_j occurs with multiplicity β_j.

The tableau T is defined to be the expression

$$T = (a_{11} \ldots a_{1\lambda_1}|x_{11} \ldots x_{1\lambda_1}) \ldots (a_{s1} \ldots a_{s\lambda_s}|x_{s1} \ldots x_{s\lambda_s}),$$

where we set

$$(a_{j1} \ldots a_{j\lambda_j}|x_{j1} \ldots x_{j\lambda_j}) = \sum_\sigma \operatorname{sgn}(\sigma)(a_{j1}|x_{j\sigma(1)}) \ldots (a_{j\lambda_j}|x_{j\sigma(\lambda_j)}),$$

the above sum extending over all permutations σ of the sequence $1, \ldots, \lambda_j$.

Assign to the a_i and x_j the linear orderings

$$a_1 < \cdots < a_m \quad \text{and} \quad x_1 < \cdots < x_n.$$

Relative to these orderings, a double tableau is said to be *standard* when in each tableau the entries in each row are increasing from left to right and the entries in each column are non-decreasing downward.

The *shape* of a double tableau T is the row length vector

$$\lambda[T] = (\lambda_1, \ldots, \lambda_s).$$

Shapes of tableaux are ordered lexicographically by $\lambda > \mu$ when $\lambda_i > \mu_i$ and $\lambda_j = \mu_j$ for $j < i$.

Using this ordering on shapes we now linearly order all tableaux. Associate with T the sequence

$$\pi[T] = a_{11} \ldots a_{1\lambda_1} a_{21} \ldots a_{s\lambda_s} x_{11} \ldots x_{s\lambda_s},$$

and order the set of these sequences lexicographically.

If S denotes another double tableau then set $T > S$ if $\lambda[T] > \lambda[S]$, or if $\lambda[T] = \lambda[S]$ and $\pi[T] < \pi[S]$.

Remark: Identities in a Cayley algebra between inner products may be interpreted in R_K. To do this, substitute for each inner product $\langle a_{i_1} \ldots a_{i_k} | x_{j_1} \ldots x_{j_k} \rangle$ the double tableau $(a_{i_1} \ldots a_{i_k} | x_{j_1} \ldots x_{j_k})$. Conversely, *any* identity in R_K may be interpreted in a Cayley algebra over the integral domain R_K, and we shall use the two notations interchangeably.

LEMMA 1. *Let* $k \geq l$ *and*

$$B = b_1 \ldots b_{j-1} \qquad Y = y_1 \ldots y_k$$
$$C = c_{j+1} \ldots c_l \qquad Z = z_1 \ldots z_l$$

where the b_i *and* c_i *are vectors taken from the set* $\{a_1 \ldots a_m\}$ *and the* y_i *and* z_i *are covectors from* $\{x_1 \ldots x_n\}$. *Then the expression*

$$I = \langle B\dot{b}_j \ldots \dot{b}_k | Y \rangle \langle \dot{c}_1 \ldots \dot{c}_j C | Z \rangle$$

is equal to a sum of products of pairs of inner products, each pair containing one inner product of length at least $k + 1$.

Proof: By Theorem 6.3 we have

$$I = (\pm)(B \vee Y) \wedge (b_j \ldots b_k \quad c_1 \ldots c_j) \wedge (C \vee Z).$$

Setting $b_j \ldots b_k \quad c_1 \ldots c_j = D$, we now use Theorem 6.5 to distribute B through the other factors. This gives

$$I = (\pm) Y \wedge \sum_{s=0}^{j-1} (-)^{(s)} \{(b_1^\sigma \ldots b_s^\sigma) \vee D\} \wedge \{(b_{s+1}^\sigma \ldots b_{j-1}^\sigma) \vee (C \vee Z)\}$$

Distributing Z by the dual of Corollary 1 to Theorem 6.2, this becomes

$$I = (\pm) \sum_{s=0}^{j-1} (-)^{(s)}$$
$$\{(b_1^\sigma \ldots b_s^\sigma D) \wedge Y \wedge (\dot{z}_1 \ldots \dot{z}_{s+1})\} \vee \{(b_{s+1}^\sigma \ldots b_{j-1}^\sigma C) \wedge (\dot{z}_{s+2} \ldots \dot{z}_l)\},$$

or

$$I = (\pm) \sum_{s=0}^{j-1} (-)^{(s)} \begin{pmatrix} b_1^\sigma \ldots b_s^\sigma D & Y \dot{z}_1 \ldots \dot{z}_{s+1} \\ b_{s+1}^\sigma \ldots b_{j-1}^\sigma C & \dot{z}_{s+2} \ldots \dot{z}_l \end{pmatrix},$$

which concludes the proof.

THEOREM 1. (*Straightening Formula*) *The double standard tableaux of content* (α, β) *span* $V_{\alpha\beta}$.

Proof: Any monomial of step zero equals a linear combination of monomials of the form

$$\langle a|x\rangle\langle b|y\rangle\ldots\langle c|z\rangle = \begin{pmatrix} a & x \\ b & y \\ \vdots & \vdots \\ c & z \end{pmatrix}$$

We show that any double tableau equals a linear combination of double standard tableux. We proceed by induction on the linear ordering of tableaux, and show that every non-standard double tableau T of content (α, β) equals a linear combination of greater tableaux of content (α, β). Since there are only finitely many double tableaux of content (α, β) iteration of this argument must then eventually express T as a linear combination of double standard tableaux.

If two entries in T satisfy $t_{ij} \geq t_{i,j+1}$ or $t_{ij} > t_{i+1,j}$ call this a *violation* of standard form in T.

Assume a violation occurs in the left tableau. If it is a row violation, $a_{ij} > a_{i,j+1}$ then set $T = -S$ where S is obtained by reversing the positions of a_{ij} and $a_{i,j+1}$ in T. Note that $\pi[T] > \pi[S]$ so that $S > T$.

Now assume a column violation $a_{ij} > a_{i+1,j}$ occurs.

Let T_1 denote the first $i - 1$ rows of T, T_2 denote the next two rows of T, and T_3 denote the remaining rows. We are primarily concerned with T_2, which we display as

$$T_2 = \begin{pmatrix} B & b_j \ldots b_k & Y \\ c_1 \ldots c_j & C & Z \end{pmatrix},$$

where

$$B = b_1 \ldots b_{j-1} \quad Y = y_1 \ldots y_k$$

$$C = c_{j+1} \ldots c_l \quad Z = z_1 \ldots z_l.$$

Consider the expression

$$I = \begin{pmatrix} B & \dot{b}_j \ldots \dot{b}_k & Y \\ \dot{c}_1 \ldots \dot{c}_j & C & Z \end{pmatrix}.$$

Since any indicated permutation σ, except the identity, exchanges elements from the first row of I with elements from the second row, and since

$$c_1 < \cdots < c_j < b_j < \cdots < b_k,$$

it must be true that $\sigma(c_j) > \sigma(b_j)$. Thus we have that

$$I = T_2 + \sum_{S > T_2} c(S)S$$

where $c(S)$ are integers. By Lemma 1 we also have

$$I = \sum_{Q > T_2} c(Q)Q$$

Combining these results gives

$$T_2 = \sum_{Q > T_2} c(Q)Q + \sum_{S > T_2} c(S)S$$

which expresses T_2 as a linear combination of greater tableaux. Appending this expression for T_2 to T yields an expression for T as a linear combination of greater tableaux. Similarly, if violations occur in the right tableau of T, they may be straightened by an analogous procedure.

This completes the proof.

In the course of the proof the following result has been implicitly established:

COROLLARY. *Let P and Q be elements of $V_{\alpha\beta}$, and let*

$$P = (a_{i_1} \cdots a_{i_s} | x_{j_1} \ldots x_{j_s})Q.$$

Then P equals a linear combination with integer coefficients of double standard tableaux, whose first rows are of length s or greater.

Theorem 1 has an interpretation in a Cayley algebra over K.

THEOREM 2 (*Straightening Formula for Cayley Algebras*). *Any monomial of content (α, β) of step zero in the vectors a_i and the covectors x_j, built out of joins and meets in the Cayley algebra of a vector space of dimension d equals a linear combination with integer coefficients of double standard tableaux of content (α, β), whose rows are of length at most d.*

We next establish the linear independence of the double standard tableaux, using a new kind of polarization. We begin with some definitions.

The *set-polarization operator*

$$D^k(b, a) = D^k_{ba}$$

acts on a monomial in $V_{\alpha\beta}$ by replacing it by the sum of the monomials obtained by replacing in turn every subset of k entries equal to a by a subset of k entries equal to b. If the given monomial has p occurrences of the symbol a, then the result of applying the operator D^k_{ba} is the sum of $\binom{p}{k}$ terms. If the monomial has fewer than k occurrences of the symbol a, the result is 0. For $k = 1$ the operator $D^1(b, a)$ is the classical polarization operator.

The *substitution operator*

$$S(b, a) = S_{ba}$$

acts on monomials in $V_{\alpha\beta}$ by replacing each occurrence of the symbol a by an occurrence of the symbol b.

Now extend set-polarization and substitution to all of $V_{\alpha,\beta}$ by linearity.

The following combinatorial lemma is easily proven by the pigeonhole principle:

LEMMA 1. *Let S and T be single tableaux of the same content with $\lambda[S] \leq \lambda[T]$. If in each tableau the entries in each row are strictly increasing, then one of two alternatives occurs:*

(1) *S and T are of the same shape, and the entries in each column of T are obtained by permuting the entries in the corresponding column of S, or*

(2) *Some row of T contains at least two entries which appear in the same column of S.*

We are now ready to prove the linear independence of the double standard tableaux in R_K.

THEOREM 3. *The double standard tableaux of content* (α, β) *form a basis for* $V_{\alpha\beta}$.

Proof: It suffices to produce for any double standard tableau $\{T_1|T_2\}$ a linear transformation $P(T_1|T_2)$ from $V_{\alpha\beta}$ to some vector space satisfying

$$(*)\qquad \begin{aligned} P(T_1|T_2)\{T_1|T_2\} &= w \\ P(T_1|T_2)\{D_1|D_2\} &= 0 \end{aligned}$$

for $w \neq 0$ and where $\{D_1|D_2\}$ is any other double standard tableau $\{D_1|D_2\}$ of shape $\geq \lambda$, where λ = shape of $\{T_1|T_2\}$. For then, if the double standard tableaux were not independent, there would be a non trivial linear combination $\mathscr{L}$ of double standard tableaux equalling zero, and if we were to take a tableau $\{T_1|T_2\}$ of least shape with non zero coefficient in $\mathscr{L}$ (say the coefficient of $\{T_1|T_2\}$ is d), then applying $P(T_1|T_2)$ to $\mathscr{L}$ would yield $d \cdot w = 0$ which is impossible since $d \neq 0$ and $w \neq 0$. Hence the double standard tableaux would have to be independent.

Let M_K be the polynomial ring over K obtained by adjoining transcendentals $(s_{ij}|t_{kl})$ and $(b_p|y_q)$ where indices range over finite sets of sufficient size to perform the following constructions. Let W denote the vector space with the $(s_{ij}|t_{kl})$ and $(b_p|y_q)$ as a basis.

In the double tableau $\{T_1|T_2\}$ let α_{ij} be the number of entries equal to a_i in column j of T_1 and let β_{ij} be the number of entries equal to x_i in column j of T_2. Set

$$D(T_1|T_2) = \prod_{i,j} D^{\alpha_{ij}}(s_{ij}, a_i) \prod_{i,j} D^{\beta_{ij}}(t_{ij}, x_i)$$

Now let

$$S(T_1|T_2) = \prod_{ij} S(b_j, s_{ij}) \prod_{ij} S(y_j, t_{ij}).$$

By the above definitions, the operator

$$P(T_1|T_2) = S(T_1|T_2)D(T_1|T_2)$$

is a linear operator which maps $V_{\alpha,\beta}$ into W.

To see that $P(T_1|T_2)$ satisfies (*), we begin by computing $D(T_1|T_2)\{T_1|T_2\}$. This is a sum of the form

$$D(T_1|T_2)\{T_1|T_2\} = \{T'_1|T'_2\} + \sum \{V_1|V_2\}$$

where $\{T'_1|T'_2\}$ is obtained by replacing the α_{ij} entries in the j-th column of T_1 which are equal to a_i by s_{ij} and simultaneously replacing the β_{ij} entries in the j-th column of T_2 which are equal to x_i by t_{ij}. Each term $\{V_1|V_2\}$ has the property that it may not be obtained from $\{T'_1|T'_2\}$ by permuting the elements within a column. We claim that

$$P(T_1|T_2)\{T_1|T_2\} = \{T''_1|T''_2\} \neq 0,$$

where all entries in the jth column of T''_1 or T''_2 equals b_j or y_j, respectively. Clearly $\{T''_1|T''_2\}$ is one term in $P(T_1|T_2)\{T_1|T_2\}$ since it is the image of $\{T'_1|T'_2\}$ under $S(T_1|T_2)$. But by the above property of the other terms $\{V_1|V_2\}$, and since $D(T_1|T_2)$ preserves the shape of a double tableau, we have by the preceding lemma that $S(T_1|T_2)\{V_1|V_2\} = 0$.

Now consider any other double standard tableau $\{G_1|G_2\}$ of shape $\geq \lambda$. $D(T_1|T_2)\{G_1|G_2\}$ is a sum of terms $\{Y_1|Y_2\}$ of shape $\geq \lambda$ which are not equal to $\{T'_1|T'_2\}$ or obtained from it by rearranging elements within a column. Hence by the Lemma $P(T_1|T_2)\{G_1|G_2\} = 0$. This completes the proof.

The Straightening Formula for R_K states that $V_{\alpha\beta}$ has two bases, the monomials of content (α, β) and the double standard tableaux of content (α, β). This result can be related to an identity in the theory of the symmetric group.

Let $M(\alpha, \beta)$ be the dimension of $V_{\alpha,\beta}$, and note that this number equals the number of matrices with non-negative integer entries and with row sums $(\alpha_1, \alpha_2, \ldots)$ and column sums $(\beta_1, \beta_2, \ldots)$. Let $K(\alpha, \lambda)$ be the number of *single* standard tableaux of content α and shape λ. Then the above yields the identity

$$M(\alpha, \beta) = \sum_{\lambda} K(\alpha, \lambda) K(\beta, \lambda),$$

as λ ranges over all partitions of the integer n.

We now extend the linear independence of the standard tableaux to a more general ring. We begin by motivating our construction with imprecise but, we hope, suggestive language. In a vector space of dimension d, monomials in the inner products $\langle a_i | x_j \rangle$ are not always linearly independent. This leads to constructing a homomorphic image of R_K which is isomorphic with the ring of inner products of vectors and covectors in dimension d.

Consider the ideal J_d in R_K generated by the elements

$$\det_{\substack{i \in I \\ k \in K}} (a_i | x_k)$$

as I and K range over all subsets of $d + 1$ elements of $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$, where d is a given integer.

The ideal J_d is invariant under permutation of the variables a_i and x_j. Furthermore, every double tableau having one row longer than d belongs to J_d. By Theorem 3, these double standard tableaux are independent, and by the Corollary to Theorem 2, every element of J_d equals a linear combination of double standard tableaux each of which has a row longer than d. Concluding, we have proved the

LEMMA. *The ideal J_d has a basis consisting of all double standard tableaux in the entries a_i and x_j having at least one row of length greater than d.*

We can now state the main result of this Section:

THEOREM 4. *In the quotient ring $G_d(K)$ the double standard tableaux whose rows are of length at most d form an integral basis.*

Proof: By the preceding lemma, taking the quotient by the ideal J_d amounts to setting to zero all double standard tableaux having one row longer than d, and only these. Hence, the conclusion follows from Theorem 3.

Finally we note the remarkable fact that by Theorem 4, even though monomials in the $(a_i | x_j)$ are not independent, nevertheless the double standard tableaux are.

9. The First Fundamental Theorem

We now apply the Straightening Formula to derive the main results on vector invariants over arbitrary fields. The technique is simpler than the ones classically used, which apply only to fields of characteristic zero.

Let V be an n-dimensional vector space over a field K, and let

$$F(x_1, \ldots, x_N) = F(x_1, \ldots, x_N; e_1, \ldots, e_n)$$

be a polynomial function of the coordinates of the vectors $x_1, \ldots, x_N$ relative to the basis of covectors $e_1, \ldots, e_n$. Since the jth coordinate of the vector x_i may be written as

$$\pm x_{ij} = x_i \wedge e_j = \langle x_i | e_j \rangle = \langle x_i | j \rangle,$$

the function $F(x_1, \ldots, x_N)$ equals a linear combination of double tableaux in the vectors x_i and the covectors e_j.

A polynomial is *invariant* when for every non-singular linear transformation T on V,

$$F(Tx_1, \ldots, Tx_N; e_i, \ldots, e_n) = \lambda(T)F(x_1, \ldots, x_N; e_1, \ldots, e_n)$$

where $\lambda(T)$ is some scalar function.

Since T induces through its adjoint T^*, a non-singular linear transformation on covectors satisfying

$$\langle Tx_i | e_j \rangle = \langle x_i | T^* e_j \rangle,$$

and since F depends only on the $\langle x_i | e_j \rangle$, we may alternately define an invariant as a polynomial which satisfies

$$F(x_1, \ldots, x_N; T^*e_1, \ldots, T^*e_n) = \mu(T^*)F(x_1, \ldots, x_N; e_1, \ldots, e_n)$$

for all non-singular linear transformations T^* acting on covectors.

We also define a *formal invariant* as a polynomial $F(x_1, \ldots, x_N)$ which is an invariant when considered over the extension field $K(x_{11}, \ldots, x_{Nn})$, where K is the ground field of V and the coordinates x_{ij} are transcendentals.

We shall prove the following result over an arbitrary field.

THEOREM 1. *Every invariant (or formal invariant when the field K is finite) in the vectors $x_1, \ldots, x_N$ is expressible as a linear combination of products of brackets in the x_i, where each summand has the same number of bracket factors. In other words, every invariant is a word in the Cayley algebra, built out of joins and meets of $x_1, \ldots, x_N$ alone with no explicit reference to $e_1, \ldots, e_n$, in which every summand is of the same total degree.*

Proof: As noted F may be written as a linear combination of double tableaux, and thus, by the Straightening Formula, as a linear combination $\mathscr{L} = \sum_i \lambda_i \{C_i | D_i\}$ of double standard tableaux. We must therefore show that the right tableau of each summand in $\mathscr{L}$ is given by—writing j in place of e_j—

$$D = \begin{pmatrix} 12 & \ldots & n \\ & \vdots & \\ 12 & \ldots & n \end{pmatrix}$$

where D has (say) g rows.

We begin by showing that each right tableau in $\mathscr{L}$ contains each variable $e_1, e_2, \ldots, e_n$ the same number of times. From the definition of an invariant, by considering the linear transformation

$$T^*e_i = ce_i$$
$$T^*e_j = e_j \qquad j \neq i,$$

for some scalar c, we may conclude that each integer i occurs the same number of times, say g_i, in each right tableau in $\mathscr{L}$. Now by considering the linear transformation

$$\begin{aligned} T^*e_i &= e_j \\ T^*e_j &= e_i \\ T^*e_k &= e_k \qquad k \neq i, j, \end{aligned}$$

we conclude that $g_i = g_j$ for all i and j, and call the common value g.

Let us now analyze the possible order of the entries in a right tableaux D_i, more particularly in the rows. If in each row of every D_i, every integer j is immediately followed by $j + 1$, then the proof is concluded. We may therefore assume that there is a *smallest* integer j and a *first* row, say the $(k + 1)$th, such that j is not followed by $j + 1$ in this row. The rows with this property will be adjacent and below the kth. Say there are Q such rows, $R_{k+1}, \ldots, R_{k+Q}$. Then there are Q entries equal to $j + 1$ out of position. They cannot be in any of the rows preceding R_{k+1}, because these rows already contain an entry equal to $j + 1$. Hence they must lie in the rows following R_{k+Q}. Let R be one such row containing an entry equal to $j + 1$. Then $j + 1$ must be at the left of this row. For it cannot be to the right of the jth place, otherwise the tableau would not be standard in the corresponding column, and it cannot be between the first and the jth place, otherwise the minimality of j would fail.

Hence, following row R_{k+Q} there are Q further rows $R_{k+Q+1}, \ldots, R_{k+2Q}$ for each of which the left entry is $j + 1$.

Thus, the tableau must be of the following form:

$$\begin{array}{r l}
k \text{ rows} & \left\{\begin{array}{llllllll} 1 & 2 & \cdots & j & j+1 & * & \cdots & . \\ 1 & 2 & \cdots & j & j+1 & * & \cdots & . \\ \vdots & & & & & & & \\ 1 & 2 & \cdots & j & j+1 & * & \cdots & \end{array}\right. \\
Q \text{ rows} & \left\{\begin{array}{llllll} 1 & 2 & \cdots & j & * & . \ \cdots \\ \vdots & & & & & \\ 1 & 2 & \cdots & j & * & . \ . . \end{array}\right. \\
Q \text{ rows} & \left\{\begin{array}{llllll} j+1 & * & \cdots & . & . & . \ . \\ \vdots & & & & & \\ j+1 & * & \cdots & . & . & . \end{array}\right. \\
 & \begin{array}{llllll} * & . & \cdots & . & . & . \\ * & . & \cdots & . & . & \\ * & . & \cdots & . & . & \end{array}
\end{array}$$

where the stars stand for entries greater than $j + 1$.

Since this analysis accounts for all $(j + 1)$'s out of position, we must have $Q = g - k$. Thus, since k was chosen to be minimal Q is the maximal number of j's not followed by $j + 1$ in any D_i. Say this occurs in the tableaux $\{C_1|D_1\}, \ldots,$

$\{C_\alpha|D_\alpha\}$ of $\mathscr{L}$, so that

$$F = \sum_{i=1}^{\alpha} \lambda_i\{C_i|D_i\} + \text{other terms.}$$

Consider the linear transformation

$$T^*e_j = e_j + e_{j+1}$$
$$T^*e_i = e_i \qquad i \neq j$$

Under T^*, each tableaux $\{C_i|D_i\}$ is sent into the sum of the tableaux obtained by replacing in turn every subset of Q or fewer entries equal to j by $j+1$. Of course the resulting tableaux may not be standard or may even equal zero.

Let us see what happens to the first α tableaux by this substitution. Replacing the Q entries equal to j in rows $R_{k+1},\ldots,R_{k+Q}$ by $j+1$ we obtain standard tableaux with Q fewer entries equal to j. These standard tableaux have fewer j's than necessary, and must be cancelled out by tableaux obtained from other substitutions. By the maximality of Q and the linear independence of the standard tableaux this is impossible. We have thus reached a contradiction which concludes the proof.

We now give an alternative version of the First Fundamental Theorem valid for all fields.

The following lemma is a simple consequence of the multinomial expansion:

LEMMA 1. *Let $F(x,\ldots,z)$ be a homogeneous polynomial function, of degree g, of the coordinates of the vectors $x,\ldots,z$. Then for any scalars $\lambda_i,\ldots,\mu_i$ and vectors $x_i,\ldots,z_i$ we have*

$$F\left(\sum_i \lambda_i x_i,\ldots,\sum_i \mu_i z_i\right) = \sum_{i_1,i_2\ldots}\cdots\sum_{k_1,k_2\ldots} \lambda_1^{i_1}\lambda_2^{i_2}\ldots\mu_1^{k_1}\mu_2^{k_2}$$
$$F_{i_1i_2\ldots,k_1k_2\ldots}(x_1,x_2,\ldots,z_1,z_2,\ldots)$$

where the sum ranges over all $i_1,\ldots,k_1,\ldots$ such that

$$\sum_j i_j + \cdots + k_j = g$$

and the $F_{i_1i_2\ldots,k_1k_2\ldots}$ are homogeneous of degree g.

The proof is omitted, as the result is well-known.

LEMMA 2. *In a Cayley space of dimension n, let $F(x_1,x_2,\ldots,x_n)$ be a scalar valued function of vectors $x_1,\ldots,x_n$ which is invariant under all non-singular linear transformations T, that is, such that for some scalar function $\lambda(T)$,*

$$F(Tx_1,Tx_2,\ldots,Tx_n) = \lambda(T)F(x_1,\ldots,x_n)$$

Then

$$F(x_1,x_2,\ldots,x_n) = c[x_1,x_2,\ldots,x_n]^g$$

for some constant c and integer g.

The proof is omitted, as the result is well known to hold over an arbitrary field, and an easy consequence of the fact that the determinant is an irreducible polynomial.

LEMMA 3. *Let $F(x_1, \ldots, x_N)$ be a homogeneous invariant of degree g. Then the polynomial*

$$[x_1, \ldots, x_n]^g F(x_1, \ldots, x_N)$$

equals a polynomial in the brackets $[x_{i_1}, x_{i_2}, \ldots, x_{i_n}]$.

Proof: Since the function F is homogeneous of degree g,

$$[x_1, \ldots, x_n]^g F(x_1, \ldots, x_N)$$
$$= F([x_1, \ldots, x_n]x_1, \ldots, [x_1, \ldots, x_n]x_n, [x_1, \ldots, x_n]x_{n+1}, \ldots, [x_1, \ldots, x_n]x_N).$$

Using the identity

$$[x_1, \ldots, x_n]x_j = \sum_{k=1}^{n} [x_1, \ldots, x_{k-1}, x_j, x_{k+1}, \ldots, x_n]x_k$$

and expanding as in Lemma 1, we find that

$$(*) \qquad [x_1, \ldots, x_n]^g F(x_1, \ldots, x_N) = \sum_m c_m F_m(x_1, x_2, \ldots, x_n)$$

where the subscript m ranges over a set of multi-indices, and the coefficients c_m are products of brackets of the form

$$b_j = [x_1, x_2, \ldots, x_{k-1}, x_j, x_{k+1}, \ldots, x_n].$$

Note that for $j > n$ the b_j are algebraically independent (in the case of finite fields of p elements, after making the reduction $x^p = x$). This follows from the algebraic independence of the $\langle x_i | e_j \rangle$.

Because of Lemma 2, the proof will be concluded if we can show that each of the $F_m(x_1, \ldots, x_n)$ is an invariant. Since multiplying an invariant by a product of brackets preserves invariance, we may conclude that

$$[x_1 \ldots x_n]^g F(x_1 \ldots x_N)$$

is an invariant. Thus

$$[Tx_1 \ldots Tx_n]^g F(Tx_1 \ldots Tx_N) = \lambda(T)[x_1 \ldots x_n]^g F(x_1 \ldots x_N)$$

Substituting in (*) we get, since the c_m are also invariants,

$$\sum_m c_m(Tx_1 \ldots Tx_N)F_m(Tx_1 \ldots Tx_n) = \mu(T) \sum_m c_m(x_1, \ldots, x_N)F_m(x_1 \ldots x_n)$$

Since both sides are polynomials in the b_j, and since the b_j are algebraically independent, their coefficients must coincide. This gives

$$F_m(Tx_1 \ldots Tx_n) = \nu(T)F_m(x_1 \ldots x_n)$$

which concludes the proof.

THEOREM 2. (*First Fundamental Theorem of Invariant Theory*). *Every homogeneous invariant in the vectors $x_1, \ldots, x_N$ is expressible as a word in the Cayley algebra, built out of joins and meets alone.*

Proof: By Lemma 3, there is an integer g such that

$$[x_1, \ldots, x_n]^g F(x_1, \ldots, x_N)$$

is a polynomial in the brackets, that is, a linear combination of double tableaux of the form

$$\sum_i \{T_i|D\}$$

where

$$D = \begin{pmatrix} 12 & \dots & n \\ & \vdots & \\ 12 & \dots & n \end{pmatrix}.$$

We wish to show that it is possible to cancel $[x_1, \dots, x_n]^g$ while retaining the rectangular form of the right tableaux. By the Straightening Formula, F may be written as

$$F = \sum \{U_i|V_i\},$$

a linear combination of double standard tableaux. Let

$$U_i' = \begin{pmatrix} x_1 & \dots & x_n \\ & \vdots & \\ x_1 & \dots & x_n \\ & U_i & \end{pmatrix} \qquad V_i' = \begin{pmatrix} 1 & \dots & n \\ & \vdots & \\ 1 & \dots & n \\ & V_i & \end{pmatrix},$$

where vertical dots indicate that a total of g rows have been placed above each of U_i and V_i as shown. U_i' and V_i' are clearly standard. Now note that

$$[x_1, \dots, x_n]^g F(x_1, \dots, x_n) = \sum_i \{U_i'|V_i'\}.$$

We have thus written $[x_1, \dots, x_n]^g F(x_1, \dots, x_n)$ as a linear combination of double standard tableaux in two different ways. By the linear independence of the double standard tableaux these must agree, giving

$$V_i' = D.$$

It follows from this that V_i is also rectangular with rows equal to $1 \dots n$, which concludes the proof.

10. Time-ordering (*sketch*)

We consider here the space $V_{\alpha,\beta}$ introduced in the statement of the Straightening Formula, and now assume that the entries of the vector β are all equal to zero or one; that is, that there are no repeated covectors in any monomial in $V_{\alpha,\beta}$. We now treat $V_{\alpha,\beta}$ as a module over the group-ring of the symmetric group acting on the set of covectors. The proof of the Straightening Formula, considered in this context, says that every submodule of $V_{\alpha,\beta}$ which is invariant under permutations of vectors is spanned by linear combinations of double standard tableaux.

We shall begin by determining the structure of minimal submodules. In characteristic zero, these give an irreducible representation of the symmetric group; but these representations make sense over any field, although they may not be irreducible.

A submodule M of $V_{\alpha\beta}$ which is spanned by inner products of the form $\langle x_i|X_j\rangle$ is minimal if and only if the set of double standard tableaux in M is the set of all possible right tableaux of some fixed shape λ, adjoined to one left tableau L of shape λ with the property that the vectors in row $i+1$ of L are a subset of the vectors in row i for all i.

Proof: We need to show (a) that a submodule of $V_{\alpha,\beta}$ which has as a basis any proper subset of S is no longer invariant under the given permutation group, and (b) that if the covectors in the right tableau of any double standard tableau in S are permuted, then the resulting double tableau may be written as a linear combination of tableaux in S.

Part (a) is true since the set M is transitive under the given permutation group. Part (b) is a consequence of the straightening algorithm, since upon straightening, any tableaux of higher shape which occur will have repeated elements in some row of the left tableau.

An example of minimal invariant module is associated with shape λ as follows. One takes the set S to be the set of all tableaux whose first column on the left side has all entries equal to x_1, whose second column has all entries equal to x_2, etc. These tableaux give explicitly the matrix units of a representation of the symmetric group which in characteristic zero is always irreducible; it can be shown that one obtains in this way all the irreducible representations of the symmetric group.

By extending the above reasoning one can classify all submodules of $V_{\alpha\beta}$ which are spanned by double standard tableaux. A submodule A of $V_{\alpha,\beta}$ spanned by double standard tableaux is spanned by the set of all standard tableaux obtained from a given set S of standard tableaux by iterating the straightening algorithm until no further standard tableaux may be obtained.

In characteristic zero, one obtains in this way the complete reducibility of invariant submodules. However, the algorithm gives an analog of complete reducibility for arbitrary fields.

The preceding idea can be applied to the study of submodules of free associative algebras which are invariant under arbitrary permutations of the variables, by the device of entangling and disentangling, which we now describe.

Let π be a partition of the integer n which we write as $n = \pi_1 + \cdots + \pi_k$ where $\pi_1 \leq \cdots \leq \pi_k$, and let W_π be the submodule of the free associative algebra in the variables $x_1, \ldots, x_n$ spanned by all monomials whose content is the vector $\alpha_\sigma = (\pi_{\sigma_1}, \ldots, \pi_{\sigma_k})$ for some permutation σ of $\{1, 2, \ldots, k\}$.

Such a monomial is of the form

$$x_{i_1} \ldots x_{i_n}$$

where the multiplicities of the x_{i_j} are the integers $\pi_1, \ldots, \pi_k$ in some order.

Associate with this monomial the product

$$\langle x_{i_1}|1\rangle \ldots \langle x_{i_n}|n\rangle$$

in the *commutative* variables $\langle x_{i_1}|1\rangle, \ldots, \langle x_{i_n}|n\rangle$. This association extends to a linear operator F, the *entangling operator*, from W_π to the vector space

$$V_\pi = \sum_{\alpha_\sigma} \sum_{\beta} V_{\alpha\beta}.$$

where we sum over all β such that β has n ones and all other entries zero.

Conversely, given an element of V_π, we can recover an element of W_π by applying the *disentangling operator* F^{-1}. For example, from

$$\langle x_1|1\rangle\langle x_2|2\rangle - \langle x_1|2\rangle\langle x_2|1\rangle$$

we obtain, by disentangling, the element

$$x_1x_2 - x_2x_1$$

of W_α. In other words, the Roman numerals in the brackets of V_π indicate the positions of the variables x_i in W_π.

Now, any set of commutative symbols $\langle x_i|j\rangle$ can be interpreted as inner products of vectors x_i and covectors j. We can therefore apply the Straightening Formula, and by the entangling and disentangling operators express every element of W_π in a canonical way as a linear combination of the polynomials obtained in this way from the double standard tableaux.

In this way, the classification of identities in associative algebras is reduced under suitable homogeneity assumptions to the classification of the identities defined by double standard tableaux. Consider an associative algebra A in the variables $x_1, \ldots, x_N$. An identity holding in A is an expression of the form

$$\sum a_{i_1 \ldots i_n} x_{i_1} \ldots x_{i_n} = 0,$$

where the $a_{i_1 \cdots i_n}$ are elements of the field F which are invariant under any permutation of the variables $x_1, \ldots, x_N$. This identity is associated with the submodule generated by the monomials

$$\sum_{i_1, \ldots, i_n \in \{1 \ldots N\}} a_{i_1 \ldots i_n} x_{\sigma(i_1)} \ldots x_{\sigma(i_n)}$$

as σ ranges over all permutations. Upon applying the entangling operator, this submodule is mapped into a subspace V_π. The Straightening Formula now yields a basis of double standard tableaux. The image of this basis under the disentangling operator F^{-1} yields a canonical set of monomials in A which generate the submodule. For example, the tableau

$$\langle x_1 \ldots x_n|12 \ldots n\rangle$$

gives after disentangling the standard identity

$$\sum_\sigma (\text{sign } \sigma) x_{\sigma 1} x_{\sigma 2} \ldots x_{\sigma n}.$$

An interpretation of the First Fundamental Theorem in this context gives some pertinent information.

11. Symmetric functions (*sketch*)

The classical identities between symmetric functions can be obtained from identities in a Cayley algebra.

Let the field F be obtained from a base field K by adjoining as many transcendentals (variables) as will be needed in the sequel. Choose a doubly infinite sequence of vectors

$$x^{(1)}, x^{(2)}, x^{(3)}, \ldots$$

and covectors

$$U^{(1)}, U^{(2)}, U^{(3)}, \ldots$$

in an n-dimensional vector space V over F, and assume that all coordinates, relative to a coordinate system which will remain fixed from now on, are independent transcendentals $x_j^{(i)}$ and $U_j^{(i)}$.

Let K_λ be the field obtained from K by adjoining n transcendentals λ_j and let L be the linear map of the field F into K_λ defined as follows

$$L(x_i^{(k)} U_j^{(k)}) = \delta_{ij}\lambda_j,$$

$$L(x_{i_1}^{(1)} U_{i_1}^{(1)} x_{i_2}^{(2)} U_{i_2}^{(2)} \ldots x_{i_k}^{(k)} U_{i_k}^{(k)}) = L(x_{i_1}^{(1)} U_{i_1}^{(1)}) L(x_{i_2}^{(2)} U_{i_2}^{(2)}) \ldots L(x_{i_k}^{(k)} U_{i_k}^{(k)})$$

and so forth, where the indices are not necessarily distinct. Other values of L on monomials are set equal to zero. Note that

$$L(x^{(i)} \wedge U^{(j)}) = \sum_k \lambda_k \delta_{ij}$$

The polynomial

$$L(\langle x^{(1)} \ldots x^{(k)} | U^{(1)} \ldots U^{(k)} \rangle)$$

equals $k!a_k$, the kth elementary symmetric function in the variables λ_j.

We shall carry out the proof only for the case $k = 2$, the general case being similar. Thus, in terms of the given basis $e_1 \ldots e_n$, and dual basis $E_1 \ldots E_n$,

$$x^{(1)} \vee x^{(2)} = \sum_{i<j} (x_i^{(1)} x_j^{(2)} - x_i^{(2)} x_j^{(1)}) e_i \vee e_j$$

$$U^{(1)} \wedge U^{(2)} = \sum_{i<j} (U_i^{(1)} U_j^{(2)} - U_i^{(2)} U_j^{(1)}) E_i \wedge E_j$$

so that the induced inner product becomes

$$\langle x^{(1)} x^{(2)} | U^{(1)} U^{(2)} \rangle = \sum_{i<j} (x_i^{(1)} x_j^{(2)} - x_i^{(2)} x_j^{(1)})(U_i^{(1)} U_j^{(2)} - U_i^{(2)} U_j^{(1)}).$$

Applying the linear functional L, this becomes

$$\sum_{i<j} L(x_i^{(1)} x_j^{(2)} U_i^{(1)} U_j^{(2)} + x_i^{(2)} x_j^{(1)} U_i^{(2)} U_j^{(1)}),$$

as the other two terms vanish when L is applied. But it is seen from the definition of L that the above equal $2!a_2$, as desired.

The polynomial

$$L((x^{(1)} \wedge U^{(2)}) \vee (x^{(2)} \wedge U^{(3)}) \vee \cdots \vee (x^{(k)} \wedge U^{(1)}))$$

equal s_k, the power-sum symmetric function in the λ_j.

Again we carry out the proof for $k = 2$, where we find, upon expanding,

$$L((x^{(1)} \wedge U^{(2)}) \vee (x^{(2)} \wedge U^{(1)})) = L\left(\sum_{i,j} x_i^{(1)} U_i^{(2)} x_j^{(2)} U_j^{(1)}\right)$$

all terms with $i \neq j$ vanish, by the definition of L, and this reduces to

$$L\left(\sum_i x_i^{(1)} U_i^{(1)} x_i^{(2)} U_i^{(2)}\right) = \sum_i \lambda_i^2,$$

as desired.

Every polynomial in the inner products $\langle x^{(i)}|U^{(j)}\rangle$ which contains as many occurrences of the vector variables $x^{(i)}$ as of the covector variable $U^{(i)}$ for each i, equals a symmetric function of the λ_k.

Indeed, every such polynomial can be written as a sum of products of disjoint cycles as in (*), and each such cycle equals a symmetric function s_k.

Identities for symmetric functions may have analogs in the Cayley algebra. The analog of Newton's formula, expressing the a_k in terms of the s_k, is obtained as follows. Expanding the inner product defining a_k, we find

$$\langle x_1 \dots x_k | U_1 \dots U_k\rangle = \langle x_1|U_1\rangle\langle x_2 \dots x_k|U_2 \dots U_k\rangle + \sum_{i>1} \pm \langle x_i|U_1\rangle\langle x_1 \dots \hat{x}_i \dots x_k|U_2 \dots U_k\rangle$$

The second term on the right is further expanded, giving $k - 1$ summands of the form

$$(*) \qquad (x_i \wedge U_1) \vee (x_1 \wedge U_i) \vee \langle \hat{x}_1 x_2 \dots \hat{x}_i \dots x_k | \hat{U}_1 U_2 \dots \hat{U}_i \dots U_k\rangle$$

as well as other terms. The remaining terms are further expanded, giving terms of the form

$$(**) \qquad (x_j \wedge U_1) \vee (x_i \wedge U_j) \vee (x_1 \wedge U_i) \vee (\text{Inner Product})$$

as well as other terms. Clearly terms of the form (*) correspond to $s_2 a_{k-2}$, and terms of the form (**) to products $s_3 a_{k-3}$, etc.

Waring's formula, expressing the a_k in terms of the s_k, is even easier. It reduces to the remark that the determinant

$$\langle x_1 \dots x_k | U_1 \dots U_k\rangle = \det\langle x_i|U_j\rangle$$

is a sum of terms, each of which splits into disjoint cycles of a permutation of the indices.

We can define the Schur functions e_μ corresponding to a tableau of shape μ to be L applied to the symmetrized tableau (v. below) of shape μ in the variables x_i and U_i. It is then not difficult to derive the determinant expression for the Schur functions in terms of the elementary symmetric functions a_k. Various results on characters of the symmetric group can be derived and extended by the present approach.

12. Further work

We sketch some lines of work indicated by the present investigations. Some are intended to display applications of the present technique; others are topics which might be further pursued.

(1) *The Gordan–Capelli formula*

The Gordan–Capelli formula is a consequence of the Straightening Formula; we state it without proof—and in greater generality than is found in previous work—avoiding the use of polarization operators which distract from the combinatorial simplicity of the result.

By changing the linear ordering of the variable vectors in all possible ways, and adding the corresponding expressions, one obtains an expansion which is independent of the choice of a linear order, and in some ways simpler. The drawback of such an expansion is that it holds in general only in characteristic zero, unlike the Straightening Formula.

Define a symmetrized tableau $\sigma(T_1|T_2)$ as the sum of all the double tableaux obtained by permuting all the elements of each row of T_1 in turn and independently, repetitions allowed. Thus if a row has k entries, these will be $k!$ terms, even if the row contains repeated entries.

One can show that in characteristic zero the symmetrized tableaux form a basis for $V_{\alpha,\beta}$; this is, in the case of distinct variables, the Gordan–Capelli expansion.

(2) *Strength of identities*

The Birkhoff–Witt theorem can be read as stating that, in an associative algebra, the product xy can be *recovered* from the bracket $xy - yx$; in other words, the bracket is sufficiently strong to give back the product. On the other hand, it is known that the Jordan product $xy + yx$ is in general not strong enough to give back the product. The question can be posed more generally when a given non-commutative polynomial is strong enough to yield another. We hazard the conjecture that these questions can be attacked by the time-ordering device, where $xy - yx$ becomes $(xy|12)$, together with the Straightening Formula.

(3) *Syzygies*

The Cayley algebra analog of the Second Fundamental Theorem of invariant theory is the problem of finding a set of identities on joins and meets which, in a suitable sense, form a basis for the set of all identities.

More important is the problem of the identities between identities, or syzygies of the second order. Little work has been done on this difficult subject.

(4) *Other groups*

There are analogs of the Straightening Formula for the orthogonal and the symplectic groups, which could not be included here. For the orthogonal group it is closely related to identities for spherical harmonics and Hermite polynomials. For the symplectic group, the result is similar to the Straightening Formula, except that determinants are replaced by Pfaffians. One obtains a systematic way of deriving and proving identities for Pfaffians, as well as an explanation of the oft-noted analogy between the two.

(5) *Invariants*

The age-old problem of the computation of projective invariants for sets of linear varieties can be attacked by the present techniques, and we shall limit ourselves to a remark here. Plethysm can be reinterpreted in the Cayley algebra as the relationship between the induced Cayley algebra built on extensors of step k endowed with the bracket obtained from Sylvester's identity, and the given Cayley algebra.

(6) *Word problems and invariant theory*

The version of the Straightening Formula given above is not the most general; we have chosen it because the proof requires fewer notational artifices. A more general version is concerned with words in the Cayley algebra built out of vectors and covectors, and not necessarily of step zero. The result is similar, except that one requires double standard tableaux where the left and right side are not necessarily of the same shape. In this more general version, the Straightening Formula can be viewed as the solution of the word problem in the Cayley algebra for words containing at most vectors and covectors. Several generalizations are suggested by this viewpoint. One may ask in which cases other word problems in the Cayley algebra are solvable, for words containing symbols for extensors of all steps in prescribed numbers. This problem seems not to have ever been previously treated. While it is possible that all such word problems may be solvable, there is one subclass which lends itself to a more straight-forward treatment. This is the word problem for sets of extensors whose supports generate a semimodular lattice of flats in projective space.

(7) *Hopf algebras*

We have neglected the coalgebra structure of the exterior algebra. However, the Hopf algebra structure is indispensable for a better understanding of some of the problems mentioned here especially for syzygies of higher order. The symbolic method of invariant theory is a Hopf algebra technique in disguise.

(8) *Matching Theory*

We have stated elsewhere that matching theory can be systematized by the methods of linear algebra. In support of this contention we sketch a proof of Philip Hall's Marriage Theorem. Thus, given a bipartite graph G on $A \times B$ with the property that every subset of k vertices in A connected to at least k vertices in B, we must show that there exists an injective function $f: A \to B$ such that for every $a \in A$, $(a, f(a))$ is an edge of the graph. The function f is called a matching of A to B.

We define a ring $F(G)$, called the *free ring* of the graph G, following an idea that goes back to Frobenius. Let K' be the free extension of the rational field K obtained by adjoining independent transcendentals $(a_i|x_j)$ as a_i ranges over the set A and x_j over the set B and let $F(G)$ be the homomorphic image of K' obtained by setting $(a_i|x_j) = 0$ whenever the pair (a_i, x_j) is not an edge of the bipartite graph G. We can find a vector space V, and in it vectors a_i and covectors x_j, such that $\langle a_i|x_j\rangle = (a|x_j)$.

The Marriage Theorem states (assuming for simplicity that there are as many vectors as there are covectors) that the matrix of the $(a_i|x_j)$ is non-singular under the stated hypotheses, or equivalently, that the vectors a_i as well as the covectors x_j form a basis under the stated hypothesis.

Proof: Suppose the conclusion fails. Then we can find a minimal dependent set of vectors $a_1, \ldots, a_j$, say, such that $a_1 a_2 \ldots a_j = 0$. Let X be an extensor of step $j - 1$. Expanding

$$X \wedge (a_1 a_2 \ldots a_j) = 0$$

by the alternative law, we find that

$$\sum_{i=1}^{j} \pm \langle a_1 \ldots \hat{a}_i \ldots a_j | X \rangle a_i = 0. \qquad (*)$$

Since $a_i \ldots a_{j-1}$ are independent, we can find covectors $x_1, \ldots, x_{j-1}$, say, such that $\langle a_1 \ldots a_{j-1} | x_1 \ldots x_{j-1} \rangle \neq 0$. Since $a_1, \ldots, a_j$ is a minimal dependent set, it follows that $\langle a_1 \ldots \hat{a}_i \ldots a_j | x_1 \ldots x_{j-1} \rangle \neq 0$ for all i. Expanding in the field $F(G)$, we find

$$\langle a_1 \ldots \hat{a}_i \ldots a_j | x_1 \ldots x_{j-1} \rangle = \det(a_k | x_p) = c_i$$

where $1 \leq k \leq j$ with $k \neq i$, and $1 \leq p \leq j - 1$. Thus

$$\sum_{i=1}^{j} c_i (a_i | x_q) = 0,$$

where $c_i \neq 0$ for all i.

If $q \geq j$ and $(a_i | x_q) \neq 0$ then $(a_i | x_q)$ is transcendental over the field obtained by adjoining the c_i to K. Hence the above equation can hold only if $(a_i | x_q) = 0$ for $1 \leq i \leq j$ and $j \leq q < n$, where n is the dimension of the space. We conclude that the set $a_1 \ldots a_j$ of vertices of A is related at most to the $j - 1$ vertices $x_1, \ldots, x_{j-1}$ of B, contradicting the hypothesis and ending the proof.

(9) *Translating Geometry into Algebra*

The identities developed in Section 6 indicate that the formalism of Cayley algebra should yield a technique for verifying geometric statements by algebraic methods. Such a hope was indeed the moving force behind much of the work on invariant theory carried out during the Nineteenth Century. Strangely, however, this hope remained unfulfilled, and treatises on invariant theory written at the time limit themselves to a few generalities, such as Gram's theorem. This paradoxical situation, which contributed in some measure to the downfall of classical invariant theory, is partly due to the lack of a clearly developed system of first-order logic in which to express geometric statements.

We confine the discussion to joins and meets of subspaces. If $\bar{A}$ and $\bar{B}$ are subspaces of a projective space S then we write $\bar{A} \cap \bar{B}$ for their intersection, and $\bar{A} \cup \bar{B}$ for their *sum*, that is, for the smallest subspace spanned by $\bar{A}$ and $\bar{B}$, at times also called the *join*.

The problem of translating an assertion of projective geometry into an equivalent assertion in the Cayley algebra can be subdivided into two headings:

(1) Develop an algorithm for verifying whether an identity involving intersections and sums (that is, a word in $\cup$ and $\cap$) of subspaces of projective space holds.

(2) Develop a decision procedure for the first-order theory of projective geometry.

Let $L(V)$ be the lattice of subspaces of the vector space V, where lattice-joins and lattice-meets are written $\cup$ and $\cap$. We shall be concerned with translating, and, insofar as possible, verifying a first-order logic statement in the algebra of lattice-joins and meets, into the language of Cayley algebras. We only consider universal sentences. These are sentences constructed from identities in the lattice of subspaces using the logical connectives "and", "not" and "implies", which we shall call *propositions*.

(a) Let the variables $a, b, \ldots, c, x, y, \ldots, z$ denote generic vectors; in other words, any identity in these variables states that the identity holds no matter what values are given to the variables. It follows from the Straightening Formula that the ring of brackets whose entries are generic vectors is an integral domain: it follows further that the *word problem* for any conjunction of identities in the algebra of brackets is *solvable*. Indeed, the proof of the Straightening Formula gives an explicit algorithm for the solution of the word problem (see remarks under *Word Problems*). Thus, if a given proposition can be shown to be equivalent to an identity in the algebra of brackets, then the truth of the proposition can be decided.

(b) It has been shown by Scarpellini and Whiteley that every true proposition in an integral domain is equivalent to the conjunction of equalities and inequalities.

This result is a logical equivalent of Hilbert's Nullstellensatz. It is not known whether, in the special case of the algebra of brackets, the equivalence can be obtained from an explicit algorithm.

An identity involving sums and intersections can be shown to be equivalent to a conjunction of identities and inequalities in the algebra of brackets by the following steps.

(c) An identity of the form

$$\bar{A} \geq \bar{B}$$

in the lattice $L(V)$ can be "translated" into an identity in brackets as follows. Let A and $B = b_1 b_2 \ldots b_k$ be extensors supporting $\bar{A}$ and $\bar{B}$. The above identity is equivalent to the conjunction of the k identities

$$b_i \vee A = 0, \qquad i \leq i \leq k.$$

Completing to brackets if necessary, we see that this is equivalent to a conjunction of bracket identities.

(d) An identity of the form

$$\bar{A} = \bar{B} \cup \bar{C}; \qquad \bar{A}, \bar{B}, \bar{C} \in L(V),$$

can be translated into an identity in brackets as follows. The above is equivalent to the proposition:

or every $\bar{X}$,

$$(*) \qquad \bar{X} \geq \bar{B} \text{ and } \bar{X} \geq \bar{C} \text{ if and only if } \bar{X} \geq \bar{A}.$$

Each of the containment relations is constructively equivalent to a conjunction of bracket identities by (c); further, by (b) the implication is equivalent to a bracket identity.

(e) An identity of the form

$$(**) \qquad \bar{A} = \bar{B} \cap \bar{C}$$

is translated similarly.

(f) A lattice-identity (or inequality) is decomposed into a succession of identities of the form (c), (d), and (e), by introducing extra variables if necessary.

(g) An alternative approach to steps (d) and (e) is the following. In the special case when $\bar{B} \cup \bar{C} = \bar{V}$ then the verification of (**) becomes trivial, as it reduces to checking that $A = B \wedge C$. This can be done constructively, by (c), verifying $\bar{A} \geq \bar{B} \cap \bar{C}$ and $\bar{A} \leq \bar{B} \cup \bar{C}$ in turn. If $\bar{B} \cup \bar{C} \neq V$ then we can use the *reduced bracket* modulo a generic extensor X. Then $\bar{A} = \bar{B} \cap \bar{C}$ if and only if A is equivalent to $B \wedge C$ modulo every extensor X. The definition of $B \wedge C$ depends on the choice of X.

The verification can be cut down to a *finite* number of extensors X by a process that can be considered as the Cayley algebra analog of Herbrand's theorem. In fact, a reduced bracket can be considered as the Cayley algebra analog of a quantifier. Just as in Herbrand's theorem, the reduction to a finite number of X does not yield a decision algorithm.

(h) If a proof of a lattice proposition is available which uses ordinary projective coordinates, then this proof can be translated step by step into the algebra of brackets, and be made to yield constructively a conjunction of identities and inequalities which is equivalent to the lattice proposition. This idea was partially exploited by Whiteley, but can be made very simple in the language of Cayley algebras.

13. Acknowledgments

The idea of a standard tableau made its first appearance with Clebsch, who gave ingenious applications to geometry. With him appeared also the device of polarization, further developed and sharpened by Capelli in the celebrated expansion bearing his name. However, Capelli did not recognize the importance of Clebsch's basic idea. Alfred Young, after careful study of the ideas of Clebsch and Capelli, introduced in 1901 the tableau expansion that bears his name. However, it was not until Young's third paper, published in 1927, that standard tableaux made their reappearance. In this paper one finds the first version of what—suitably generalized—we have called the straightening algorithm, which has been used since in several circumstances.

It seems that Young may have had an inkling of the Straightening Formula. To be sure, double tableaux were used by him for the representations of the octohedral groups, but are nowhere else mentioned in his work. Turnbull, in the short appendix added to his book for the second edition, sketches Young's ideas. Our work grew largely out of trying to understand some of Turnbull's ideas, which are often purely heuristic. The machinery of Cayley algebras was developed under this stimulus. Our statement and proof of the Straightening Formula is, to the best of our knowledge, the first correct and complete one.

The definition of Cayley algebra is new, as is, to the best of our knowledge, the definition of meet. The Scottish convention is inspired by Turnbull, who used it informally. Of the alternative laws, several special cases were known, but we have not found the general case (Theorem 6.5) in the literature.

The brief treatment of symmetric functions was also inspired by some work of Turnbull and Wallace, combined with a linear functional device introduced by Rota.

The time-ordering device was introduced by R. P. Feynman in another context; the present treatment was motivated by the work of P. M. Cohn.

The proof of the Marriage Theorem was arrived at by analyzing some work of Edmonds.

The proofs of the First Fundamental Theorem and of independence of standard tableaux are new.

14. Bibliography

AITKEN, A. C., "A letter to an Edinburgh colleague", *Proc. Edinburgh Math. Soc.* **16** (1968), 166–172.

BOURBAKI, N., *Eléments de Mathématique, Algèbre*, Ch. 3, Hermann, Paris, 1970.

COHN, P. M., "A theorem on the structure of tensor spaces", *Ann. Math.*, **56** (1952), 254–268.
DE PILLIS, JOHN, Grassman algebras as Hilbert space, Journal of Algebra, Vol. **10** (1968), 485–500.
EDMONDS, J., "Systems of distinct representatives and linear algebra", *J. Res. Nat. Bur. Standards, Sect. B*, **71B** (1967), 241–245.
ERICKSEN, W. E., "The intersection of subspaces", *Am. Math. Monthly*, **81** (1974), 154–160.
FEYNMANN, R. P., "An operator calculus having applications in quantum electrodynamics", *Phys. Rev.*, Vol. **84** (1951), 108–128.
FORDER, H. G., *The calculus of extension*, Chelsea, New York, 1960.
HARPER, L. AND ROTA, G.-C., "Matching theory, an introduction", in *Advances in Probability*, Vol. **1**, edited by Peter Ney, Dekker, New York, 1970.
HIGMAN, G., "Representations of general linear groups and varieties of p-groups", *Proc. Int. Conf. Theory of Groups*, Australian Nat. Univ. Canberra, Aug. 1965, 167–173, New York, Gordon & Breach, 1967.
HODGE, W. D. AND PEDOE, D., *Methods of Algebraic Geometry*, Vol. II, Cambridge Univ. Press, 1968.
IGUSA, J., "On the arithmetic normality of the Grassman variety", *Proc. Nat. Acad. Sci.*, **40** (1954), 309–313.
KÄHLER, E., "Der innere Differentialkalkül", *Rediconti di Matematica*, **21** (1962), 425–523.
KNUTH, DONALD E., "Permutations, matrices and generalized Young tableaux", *Pacific J. M.*, **34** (1970), 709–727.
LITTLEWOOD, D. E., *The Theory of Group Characters and Matrix Representations of Groups*, Second edition, Oxford, Clarendon Press, 1950.
MEYER, F., "Bericht über den gegenwärtigen Stand der Invariantentheorie", *Jahresbericht der deutschen Math. Vereinigung*, **I** (1892), 82–292.
MUIR, T. AND METZLER, W. H., *A Treatise on the Theory of Determinants*, Dover, New York, 1960.
ROTA, G.-C., "The number of partitions of a set", *Am. Math. Monthly*, Vol. **71**, 1964, 498–504.
SCARPELLINI, BRUNO, "On the metamathematics of rings and integral domains", *TAMS*. **138** (1969), 71–96.
SEMPLE, J. G. AND ROTH, L., *Introduction to Algebraic Geometry*, Oxford, Clarendon Press, 1949.
STANLEY, RICHARD P., "Theory and application of plane partitions, part 1", *Studies in Appl. Math.*, **L** (1971), 167–188.
TURNBULL, H. W., "The general symbolic notation for the principle of duality, and its application to determinants", *Proc. London Math. Soc. 2*, **22** (1923), 495–507.
TURNBULL, H. W., "A synthetic application of the symbolic invariant theory to geometry", *The Journal of the Indian Math. Soc.*, **XVII** (1927), 1–20.
TURNBULL, H. W., "The invariant theory of a general bilinear form", *Proc. London Math. Soc. 2*, **33** (1931), 1–21.
TURNBULL, H. W., "The projective invariants of four medials", *Proc. Ed. Math. Soc.* (1942), 55–72.
TURNBULL, H. W., AND WALLACE, A. H., "Clebsch–Aronhold symbols and the theory of symmetric functions"; *Proc. Royal Soc. Edinburgh*, **55** (1951), 155–173.
TURNBULL, H. W., *The Theory of Determinants, Matrices and Invariants*, Third Ed., Dover, New York, 1960.
WEYL, HERMANN, *The classical groups*, Princeton U.P., 1946.
WHITELEY, W., *Logic and invariant theory*, Ph. D. thesis, MIT, 1970.
YOUNG, A., "On quantitative substitutional analysis (third paper)", *Proc. London Math. Soc. 2*, **28**, (1927), 255–292.

COLUMBIA UNIVERSITY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, HARVARD UNIVERSITY

(Received May 15, 1974)

On the Foundation of Combinatorial Theory.
X. A Categorical Setting for Symmetric Functions

By Flavio Bonetti, Gian-Carlo Rota,† Domenico Senato,* and Antonietta M. Venezia**

A categorical setting is developed for the theory of symmetric functions.

1. Introduction

It has been said that every generation of mathematicians rewrites the theory of symmetric functions to suit the problems of the day. This work may be viewed as an instance in point.

One of the classical—and perhaps one of the earliest—interpretations of symmetric polynomials in several variables (or symmetric functions, as they have come to be improperly called) was given in terms of the combinatorial theory of distribution and occupancy (sometimes known as "placing balls in boxes"). It has been known for a long time that some of the best-known symmetric functions (for example, the elementary symmetric functions and the complete homogeneous symmetric functions) can be interpreted as generating functions for the number of subsets and of multisets of a finite set. To the best of our knowledge, the first systematic development of this point of view is due to Doubilet, who provides a combinatorial (or "bijective") interpretation of some of the fundamental identities holding among symmetric functions. Doubilet's paper is the starting point of the present work.

Our objective will be to develop the theory of symmetric functions along the lines of the theory of species initiated by André Joyal. More specifically, in keeping with the classical view of symmetric functions as polynomials of sorts,

*Research partially supported by the MPI.
†Research supported by NSF contract MCS8104855.
Professor Gian-Carlo Rota, Room 2-351, M.I.T., Cambridge, MA 02139.

Published by Elsevier Science Publishing Co., Inc. 0022-2526/92/$3.50

we define a functor on the category of sets which may be rightfully viewed as a categorical analog of an ordinary polynomial, in finitely or infinitely many variables. For these functors we propose the name "polynomial species." In fact, we provide a suitable extension of the notion of the generating function to polynomial species, and we verify that the coefficients of the generating function of a polynomial species is indeed a polynomial (in general, a polynomial in infinitely many variables).

We believe the notion of polynomial species to be of independent interest in the theory of enumeration. In the present work, it is systematically exploited to achieve only one objective: the interpretation in categorical language of the theory of symmetric functions. More specifically, our objective is to associate to every identity holding among symmetric functions a categorical (and thus "natural") identity among polynomial species. The classical identities are recovered by taking generating functions. This objective is attained in the present work for identities with nonnegative coefficients (requiring infinite sums and possibly infinite products) holding among the classical symmetric functions, with the notable exception of identities involving the Schur functions, which we hope to treat in a subsequent paper.

Our starting point was all but forced upon us by reflecting on what is to be meant by a symmetric function. Previous authors have been content to consider symmetric functions in a finite set of variables, or to solve the sometimes delicate limiting problems that arise when the number of variables is infinite, by linearly ordering the variables as a sequence $x_1, x_2, x_3, \ldots$. These ad hoc and sometimes misleading devices had to be replaced by a more subtle technique that led to a definition of the notion of a symmetric function in any unordered set of variables. Symmetric functions, in the present general setting, are elements of a ring which we call the ring of *formal polynomials*. A formal polynomial in the set of variables X consists, intuitively speaking, of arbitrary sums of monomials in the variables belonging to the set X, subject only to the requirement that it become an ordinary polynomial whenever any cofinite set of variables is set equal to zero. We devote the first section to the study of the ring $Z[(X)]$ of formal polynomials with integer coefficients, defined as the completion of the ring of ordinary polynomials in the variables X in an obvious topology. Symmetric functions are then defined as symmetric formal polynomials.

The simplicity of the topology of the ring of formal polynomials leads to a simple but useful theory of convergent infinite sums and—what is more important—of infinite products (Theorem 1).

The next section gives the main result of this work, namely, the definition of a polynomial species. This definition can be motivated as follows. The "categorical" analog of a monomial is a function defined on a subset of a set E, with values in a set of "variables" X. Thus, a "sum" of monomials has as its categorical counterpart the disjoint sum

$$\bigcup_{A \subseteq E} \operatorname{Hom}(A, X) = H[E].$$

Thus, H is a functor from the category of finite sets (of which E is an object) to the category of sets.

Next, we consider the categorical analog of a sum of monomials, each with a suitable coefficient (that is, the categorical analog of a polynomial). This is done in two steps. First, one chooses a functor M from the category of finite sets and bijections to the category of sets (generalizing the notion of species in the sense of Joyal). The functor M will be the categorical analog of the set from which the coefficients are chosen. Second, one must naturally choose for each finite set E a subset of the product

$$M[E] \times H[E] \tag{1}$$

which will correspond to the assignment of "coefficients" to each monomial. Fortunately, the categorical notion of subfunctor comes to our aid here. We define a polynomial species with coefficients in the functor M to be a subfunctor of the functor (1). Actually, the definition given in Section 3 is slightly more technical because of finiteness conditions (which are in turn a categorical counterpart of the finiteness conditions that define formal polynomials).

To every polynomial species one assigns "generating polynomials," which turn out to be formal polynomials. In a "natural" sense, every homogeneous formal polynomial with nonnegative integer coefficients turns out to be the generating polynomial of a polynomial species. In this sense, we may claim that the notion of a polynomial species is indeed the categorical counterpart of the notion of a polynomial.

A symmetric polynomial species is then defined as a polynomial species which is invariant under all automorphisms of the set of "variables" X.

In Section 4 we define infinite sums and products of families of polynomial species. Again, the motivation here is to give categorical definition of these operations among polynomial species that correspond—via generating polynomials—to infinite sums and products of formal polynomials. We find it surprising that this objective can be achieved within the category of finite sets.

The last operation we introduce is an assembly of a polynomial species, in Section 6. Here we closely follow Joyal's ideas and show that an assembly of polynomial species does indeed correspond to the exponential of their generating polynomials.

In Section 5 we introduce the species-theoretic equivalents of the classical symmetric functions, to wit:

the elementary symmetric species **A**, namely, the functor such that for any finite set E, the set $\mathbf{A}[E]$ consists of all injective functions defined on E and with values in X;

the species **H** of dispositions, such that $\mathbf{H}[E]$ is the set of all dispositions of E to X, that is, of all enriched functions from E to X bearing a permutation on each fiber;

the cyclic species **C**, such that $\mathbf{C}[E]$ consists of all functions from E to X taking only one value, and enriched by a cyclic permutation of the set E.

One easily shows that the generating polynomials of these polynomial species are, respectively, the elementary symmetric functions, the complete homogeneous symmetric functions, and the sums-of-powers symmetric functions.

We also define two nonsymmetric polynomial species: the species $\mathbf{A}_x$, defined for every element x of X, whose generating polynomial is the polynomial x, and the species $\mathbf{H}_x$, whose generating polynomials have $1/(1-x)$ as sum.

In Section 7 we prove the following identity:

$$\prod_{x \in X} (1+\mathbf{A}_x) = \mathbf{A}. \tag{2}$$

This is the categorical counterpart of the well-known infinite-product expression for the generating function of the elementary symmetric functions. We stress the fact that the identity (2) for polynomial species is stronger, since it provides a set-theoretic, or "bijective," interpretation for such an infinite product.

Similarly, for the species of dispositions we prove the identity

$$\prod_{x \in X} (1+\mathbf{H}_x) = \mathbf{H},$$

which again provides a bijective interpretation of the infinite-product expression for the homogeneous product sums. Finally, we give a bijective version of Waring's formula by the identity

$$\mathbf{H} = \mathrm{Exp}(\mathbf{C}).$$

In closing, we stress the preliminary and introductory character of the present work. Despite the elegant identities above, most of the spadework remains to be done, notably the interpretation in a categorical setting of identities among symmetric functions with alternating signs (which Doubilet elegantly interprets by Möbius inversion, and which we conjecture can be carried to a categorical setting by a "super"-theory of symmetric functions with positively and negatively signed variables), and most importantly the categorical interpretation of the Schur functions. We surmise that such interpretation will require a combinatorial theory of "super"-symmetric functions.

2. Formal polynomials

Let X be a set, possibly infinite. The elements of X will be called *variables*, and X will remain fixed throughout.

To define the ring Λ of symmetric function with integer coefficients in infinitely many variables, ordinarily one assumes that X is numerable and linearly ordered: $X = \{x_1, x_2, x_3, \ldots\}$. One can define Λ by either of two devices: first, as the direct sum of the $\mathbf{Z}$-modules Λ^h of symmetric functions of degree h, where Λ^h is defined as the inverse limit of the $\mathbf{Z}$-modules $\Lambda^h[x_1, \ldots, x_n]$ as

$n \to +\infty$; or second, as a subring of the ring of formal power series in the variables $x_1, x_2, x_3, \dots$.

We follow another approach, which we believe to be close to the combinatorics of symmetric functions, by introducing a suitable generalization of the notion of polynomial to any set X of variables, which we call *formal polynomial.*

Recall that a *multiset* m on X is a pair $(X, \overline{m}: X \to \mathbf{Z})$ where $\overline{m}$ is a function such that $\overline{m}(x) \geq 0$ for any $x \in X$. In other words, a multiset is a set X together with a function $\overline{m}$ which is to be interpreted as the *multiplicity* of every element of X.

The *support* of a multiset m on X is the set

$$\operatorname{supp}(m) = \{x \in X : \overline{m}(x) > 0\}.$$

A *finite multiset* is a multiset whose support is a finite (possibly empty) set.

We shall denote by $\mathscr{M}$ the set of finite multisets on X, and by m_0 the multiset whose support is empty.

The *cardinality* of a finite multiset m is the following integer:

$$|m| = \sum_{x \in \operatorname{supp}(m)} \overline{m}(x).$$

The *sum* $m_1 + m_2$ and *product* $m_1 \cdot m_2$ of multisets are defined as follows:

$$m_1 + m_2 = (X, \overline{m}_1 + \overline{m}_2 : X \to \mathbf{Z}),$$

where

$$(\overline{m}_1 + \overline{m}_2)(x) = \overline{m}_1(x) + \overline{m}_2(x),$$

and

$$m_1 \cdot m_2 = (X, \overline{m}_1 \cdot \overline{m}_2 : X \to \mathbf{Z}),$$

where

$$(\overline{m}_1 \cdot \overline{m}_2)(x) = \overline{m}_1(x) \cdot \overline{m}_2(x).$$

The sum and product of several multisets are defined similarly.

We omit the well-known construction that, given a set X, gives the ring $\mathbf{Z}[X]$ of all polynomials in the variables of the set X. A basis of the module $\mathbf{Z}[X]$ is given by all monomials

$$x^m = \prod_{x \in \operatorname{supp}(m)} x^{\overline{m}(x)}$$

as m ranges over all finite multisets on X.

Thus we write a polynomial p as follows:

$$p = \sum_m \text{coeff}\langle p, m \rangle x^m$$

as m ranges over all finite multisets on X, where the integers $\text{coeff}\langle p, m \rangle$ equal 0 for almost all $m \in \mathscr{M}$.

When F is a finite subset of X, we denote by ϵ_F the linear operator from $\mathbf{Z}[X]$ to $\mathbf{Z}[X]$ defined as follows:

$$\epsilon_F(p) = q,$$

where

$$q = \sum_m \text{coeff}\langle q, m \rangle x^m$$

and

$$\text{coeff}\langle q, m \rangle = \begin{cases} \text{coeff}\langle p, m \rangle & \text{if} \quad \text{supp}(m) \subset F, \\ 0 & \text{otherwise}. \end{cases}$$

In other words, the polynomial q is obtained from the polynomial p by setting to zero all coefficients of all monomials not supported on F.

Let J be a directed set (that is, a partially ordered set, whose order relation is denoted by $\geq$, such that for all $i, j \in J$ there is a $h \in J$ such that $h \geq i$ and $h \geq j$). We say that a family

$$(p_j)_{j \in J}$$

of polynomials is *Cauchy* when for every finite subset F of X there exists an element $j(F) \in J$ such that for all $j \geq j(F)$ the integers

$$\text{coeff}\big\langle \epsilon_F(p_j), m \big\rangle$$

depend on F and on m but not on j, in other words, when the family of polynomials

$$\big(\epsilon_F(p_j)\big)_{j \in J}$$

is eventually constant.

This condition for directed sets defines a Hausdorff topology on the ring $\mathbf{Z}[X]$, which with this topology is a topological ring. Clearly the operators ϵ_F are continuous in this topology. The *ring of formal polynomials* $\mathbf{Z}[(X)]$ is defined to be the completion of the ring $\mathbf{Z}[X]$ in this topology.

We shall write

$$\lim_j p_j = f$$

whenever f is a formal polynomial and $(p_j)_{j \in J}$ is a generalized Cauchy sequence in $\mathbf{Z}[X]$ that converges to f.

The continuous linear operator

$$\epsilon_F : \mathbf{Z}[(X)] \to \mathbf{Z}[(X)]$$

defined (by an abuse of notation) as

$$\epsilon_F(f) = \lim_j \epsilon_F(p_i), \qquad p_j \in \mathbf{Z}[X]$$

extends the continuous linear operator $\epsilon_F : \mathbf{Z}[X] \to \mathbf{Z}[X]$ to the space $\mathbf{Z}[(X)]$ of formal polynomials. The operator

$$\epsilon_F : \mathbf{Z}[(X)] \to \mathbf{Z}[(X)]$$

leads to a simple expression for the condition for a family of formal polynomials to be a (generalized) Cauchy sequence, to wit: a family $(f_j)_{j \in J}$ of formal polynomials is Cauchy if and only if for any finite subset F of X the family $(\epsilon_F(f_j))_{j \in J}$ of polynomials is eventually constant. Similarly, a family

$$(f_j)_{j \in J}$$

of formal polynomials converges to a formal polynomial f if and only if for any finite subset F of X, the family

$$\left(\epsilon_F(f_j)\right)_{j \in J}$$

of polynomials converges to

$$\epsilon_F(f).$$

Every formal polynomial f is the limit of the family of polynomials

$$\left(\epsilon_F(f)\right)_F$$

as F ranges over the set of all finite subsets of X. In symbols,

$$\lim_F \epsilon_F(f) = f.$$

In fact if F, F' are subsets of X such that $F \subseteq F'$, we have

$$\epsilon_F(\epsilon_{F'}(f)) = \epsilon_F(f),$$

namely,

$$\text{coeff}\langle \epsilon_F(f), m \rangle = \text{coeff}\langle \epsilon_{F'}(f), m \rangle$$

for any m whose support is contained in F.

Let $(p_F)_F$ be a family of polynomials (indexed by the finite subsets F of X) such that if $F \subseteq F'$ then $\epsilon_F(p_{F'}) = p_F$. Thus $(p_F)_F$ is a generalized Cauchy sequence in $\mathbf{Z}[X]$ indexed by the directed set of all finite subsets of X. The formal polynomial

$$f = \lim_F p_F$$

is well defined [that is, the family $(p_F)_F$ converges], and we have

$$\epsilon_F(f) = p_F$$

for every finite subset F of X.

Let

$$(f_i)_{i \in I}$$

be an indexed family of formal polynomials when I is any set. We say it is *summable* if the directed family

$$\left(\sum_{i \in J} f_i \right)_J ,$$

as J ranges over the directed set of all finite subsets of I, is convergent in the topology of $\mathbf{Z}[(X)]$. Under these conditions, the limit is denoted by

$$\sum_{i \in I} f_i .$$

PROPOSITION 2.1. *A family $(f_i)_{i \in I}$ of formal polynomials is summable if and only if the set*

$$I(F) = \{ i \in I : \epsilon_F(f_i) \neq 0 \}$$

is finite for every finite subset F of X.

Proof: For every finite subset F of X, the definition of summability gives a finite subset $J(F) \subset I$ such that

$$\epsilon_F\left(\sum_{i \in J(F)} f_i\right) = \epsilon_F\left(\sum_{i \in J} f_i\right),$$

where J is a finite subset of I, and $J(F) \subseteq J$.

If $i \notin J(F)$, take $J = J(F) \cup \{i\}$ to obtain the conclusion. The converse follows by the same argument.

PROPOSITION 2.2. *Every formal polynomial f is a convergent sum of monomials, i.e.,*

$$f = \sum_{m \in \mathscr{M}} \operatorname{coeff}\langle f, m\rangle x^m,$$

where

$$\operatorname{coeff}\langle f, m\rangle = \operatorname{coeff}\left\langle \epsilon_{\operatorname{supp}(m)}(f), m\right\rangle.$$

Proof: Set

$$p_I = \sum_{m \in I} \operatorname{coeff}\left\langle \epsilon_{\operatorname{supp}(m)}(f), m\right\rangle x^m,$$

where I is a finite subset of $\mathscr{M}$. We prove that

$$\lim_I p_I = f,$$

namely, that the family $(\epsilon_F(p_I))_I$ converges to $\epsilon_F(f)$ for every finite subset F of X. In fact, if F is a finite subset of X, the set

$$I(F) = \left\{m : \operatorname{coeff}\left\langle \epsilon_F(f), m\right\rangle \neq 0\right\}$$

is finite, since $\epsilon_F(f)$ is a polynomial. Namely, the set

$$I(F) = \left\{m : \operatorname{supp}(m) \subseteq F, \operatorname{coeff}\left\langle \epsilon_{\operatorname{supp}(m)}(f), m\right\rangle \neq 0\right\}$$

is finite and we have

$$\epsilon_F(p_I) = \epsilon_F(p_{I(F)}) = \sum_m \operatorname{coeff}\left\langle \epsilon_F(f), m\right\rangle x^m = \epsilon_F(f)$$

for any $I \supseteq I(F)$. Thus $f = \lim_I p_I$, and from the definition of summability we have

$$f = \sum_m \operatorname{coeff}\langle \epsilon_{\operatorname{supp}(m)}(f), m \rangle x^m.$$

This completes the proof of the proposition.

A formal polynomial

$$f = \sum_m \operatorname{coeff}\langle f, m \rangle x^m$$

is said to be *positive* whenever $f \neq 0$ and $\operatorname{coeff}\langle f, m \rangle \geq 0$ for all multisets m. When $\operatorname{coeff}\langle f, m_0 \rangle = 0$, we say that the formal polynomial f is *without constant term*.

The identity of the ring of the formal polynomials is the formal polynomial 1 such that

$$\operatorname{coeff}\langle 1, m_0 \rangle = 1,$$

$$\operatorname{coeff}\langle 1, m \rangle = 0 \qquad \text{otherwise}.$$

Let $f = \Sigma_m \operatorname{coeff}\langle f, m \rangle x^m$ and $g = \Sigma_m \operatorname{coeff}\langle g, m \rangle x^m$. The *sum* and *product* of the formal polynomials f and g are defined as the formal polynomials having the following coefficients:

$$\operatorname{coeff}\langle f + g, m \rangle = \operatorname{coeff}\langle f, m \rangle + \operatorname{coeff}\langle g, m \rangle,$$

$$\operatorname{coeff}\langle f \cdot g, m \rangle = \sum_{m' + m'' = m} \operatorname{coeff}\langle f, m' \rangle \operatorname{coeff}\langle g, m'' \rangle.$$

A formal polynomial f such that $\operatorname{coeff}\langle f, m \rangle \neq 0$ only if

$$|m| = h$$

is said to be *homogeneous of degree h*.

Let $\sigma : X \to X$ be a bijection. Then σ induces an automorphism of the ring $\mathbf{Z}[(X)]$, which we again denote by σ. If f is a formal polynomial which is a finite sum of homogeneous formal polynomials and such that $\sigma f = f$, we say that f is a *symmetric function*.

Let $(f_i)_{i \in I}$ be a family of formal polynomials without constant term. We say it is *multipliable* when the family of products:

$$\left(\prod_{i \in J} (1 + f_i) \right)_J$$

converges, as J ranges over the directed set of finite subsets of I. When such is the case, we say that the infinite product

$$\prod_{i \in I} (1+f_i)$$

is *convergent*.

We come to the main result of this section.

THEOREM 2.1. *A family of formal polynomials*

$$(f_i)_{i \in I}$$

without constant term is multipliable if and only if it is summable.

Proof: We first prove that the family is multipliable, under the assumption that it is summable.

From the definition of summability follows that the set

$$I(F) = \{i \in I : \epsilon_F(f_i) \neq 0\}$$

is finite for any $F \subseteq X$. Thus if J is a finite subset of I and $I(F) \subseteq J$, set

$$I(F) \cup J' = J \quad \text{and} \quad J' \cap I(F) = \varnothing.$$

Then we have

$$\prod_{i \in J} (1+f_i) = \prod_{i \in I(F)} (1+f_i) + \left(\prod_{i \in J'} (1+f_i) - 1\right) \prod_{i \in I(F)} (1+f_i)$$

On the other hand

$$\operatorname{coeff}\left\langle \epsilon_F\left(\prod_{i \in J'} (1+f_i) - 1\right), m \right\rangle = 0;$$

hence

$$\operatorname{coeff}\left\langle \epsilon_F\left(\left(\prod_{i \in J'} (1+f_i) - 1\right) \prod_{i \in I(f)} (1+f_i)\right), m \right\rangle = 0.$$

Thus

$$\operatorname{coeff}\left\langle \epsilon_F\left(\prod_{i \in I(F)} (1+f_i)\right), m \right\rangle = \operatorname{coeff}\left\langle \epsilon_F\left(\prod_{i \in J} (1+f_i)\right), m \right\rangle,$$

as desired.

Conversely, let $(f_i)_{i \in I}$ be a multipliable family of formal polynomials. Then for every F there exists a finite subset $J(F)$ of I such that

$$\epsilon_F\Big(\prod_{i \in J(F)}(1+f_i)\Big) = \epsilon_F\Big(\prod_{i \in J}(1+f_i)\Big)$$

for any $J \supseteq J(F)$. We shall prove that the set

$$I(F) = \{i \in I : \epsilon_F(f_i) \neq 0\}$$

is finite, whence the conclusion by Proposition 2.1.

Given $j \notin J(F)$, set

$$J = J(F) \cup \{j\}.$$

Then we have

$$\epsilon_F\Big(\prod_{i \in J(F)}(1+f_i)\Big) = \epsilon_F\Big(\prod_{i \in J}(1+f_i)\Big) = \epsilon_F\Big(\prod_{i \in J(F)}(1+f_i)\Big) \cdot \epsilon_F(1+f_j),$$

from which

$$1 = \epsilon_F(1+f_j) = \epsilon_F(1) + \epsilon_F(f_j) = 1 + \epsilon_F(f_j),$$

and hence

$$\epsilon_F(f_j) = 0$$

for every $j \notin J(F)$. This proves that

$$I(F) \subseteq J(F)$$

and hence that the set $I(F)$ is finite.

3. Polynomial species

The central notion of the present work is that of polynomial species. We shall see that this notion is the set-theoretic counterpart of the notion of a formal polynomial. In fact, the ordinary algebraic operations of sum and product of formal polynomials (both finite and infinite) turn out to have fitting set-theoretic counterparts for polynomial species.

Let **Ens** be the category of sets and functions, and let $\mathscr{B}$ be the category of the finite sets and bijections. We define a covariant functor H from $\mathscr{B}$ to **Ens** as follows: first,

$$H[E] = \{f: A \to X : A \subseteq E\},$$

where E is a finite set and A ranges over all subsets of E. Next, when $u: E \to E'$ is a bijection, and $f: A \to X$ is an arbitrary function, we define

$$H[u](f) = f \circ u^{-1}: u(A) \to X.$$

Thus $H[E]$ is the set of all functions defined on some subset of E, with values in the set of variables X.

Recall that, by convention, there exists a unique function from the empty set to the set X. We shall call it the *empty function* and we shall denote it by $f_{\varnothing}$.

For all functors

$$M: \mathscr{B} \to \mathbf{Ens}$$

we denote by Pol(M) the functor from $\mathscr{B}$ to **Ens** that associates the set

$$M[E] \times H[E]$$

to every finite set E. When $u: E \to E'$ is a morphism, the definition of Pol(M)[u] is obvious.

We shall write

$$P \subseteq \mathrm{Pol}(M)$$

whenever the functor P is a subfunctor of the functor Pol(M).

Let $P \subseteq \mathrm{Pol}(M)$, and let F be a finite subset of X. We define the functor

$$\epsilon_F(P): \mathscr{B} \to \mathbf{Ens}$$

by setting

$$\epsilon_F(P)[E] = \{(s, f) \in P[E]: \mathrm{Im}\, f \subseteq F\},$$

where $\mathrm{Im}\, f$ denotes the set of all elements $f(e)$ where $e \in A$ and where f is a function from A ($\subseteq E$) to X. In other words, $\epsilon_F(P)[E]$ is the subset of $P[E]$ whose elements are the pairs (s, f) such that $s \in M[E]$ and f is a function from some subset of E to X whose image is contained in F.

A *polynomial species* **P** *with coefficients on* M is defined to be a subfunctor of the functor Pol(M) such that the set

$$\epsilon_F(P)[E]$$

is finite for any finite set E and for any finite subset F of X. Note that ϵ_F is an endofunction in a category that will be defined shortly.

It is easy to see that

(1) $F \subseteq F'$ implies $\epsilon_F(\mathbf{P})[E] \subseteq \epsilon_{F'}(\mathbf{P})[E]$;
(2) $\mathbf{P}[E] = \bigcup_F \epsilon_F(\mathbf{P})[E]$.

The *functor of coefficients* of a polynomial species $\mathbf{P}$ is defined to be the functor P from $\mathscr{B}$ to **Ens** given by

$$P[E] = \{s : (s, f) \in \mathbf{P}[E]\}.$$

One verifies that if $\mathbf{P} \subseteq \mathrm{Pol}(M)$ then $P \subseteq M$ and $\mathbf{P} \subseteq \mathrm{Pol}(P)$. Thus P is "minimal" among functors $M : \mathscr{B} \to \mathbf{Ens}$ such that

$$\mathbf{P} \subseteq \mathrm{Pol}(M).$$

A polynomial species $\mathbf{P}$ is said to be *symmetric* when for every bijection $\sigma : X \to X$ we have

$$(s, \sigma \circ f) \in \mathbf{P}[E]$$

whenever $(s, f) \in \mathbf{P}[E]$.

Our next objective is to associate to every polynomial species a formal power series with coefficients in the ring $\mathbf{Z}[(X)]$ of formal polynomials. To this end, we proceed as follows. First, if $f : A \to X$ and A is a finite set, we denote by $\mathrm{gen}(f)$ the monomial in $\mathbf{Z}[X]$:

$$\mathrm{gen}(f) = \prod_{a \in A} f(a) = \prod_{x \in X} x^{|f^{-1}(x)|},$$

where $|f^{-1}(x)|$ denotes the number of elements of the set $f^{-1}(x)$. The right side is a finite product, since for almost all $x \in X$ we have $|f^{-1}(x)| = 0$.

Next, set

$$\mathrm{gen}(s, f) = \mathrm{gen}(f),$$

where (s, f) is an ordered pair, with s an element of an arbitrary set.

Finally, let $\mathbf{P}$ be a polynomial species. The family

$$(\mathrm{gen}(s, f)),$$

as (s, f) ranges in $\mathbf{P}[E]$, is summable in $\mathbf{Z}[(X)]$. In fact, for every finite subset F of X, the set

$$\epsilon_F(\mathbf{P})[E] = \{(s, f) \in \mathbf{P}[E] : \epsilon_F(\mathrm{gen}(s, f)) \neq 0\}$$

is finite. Therefore the family

$$(\mathrm{gen}(s,f))_{(s,f)\in \mathbf{P}[E]}$$

verifies the summability condition (see Proposition 2.1). We can therefore set

$$\mathrm{gen}(\mathbf{P}[E]) = \sum_{(s,f)\in \mathbf{P}[E]} \mathrm{gen}(s,f).$$

If $\mathbf{P}$ is a symmetric species, then $\mathrm{gen}(\mathbf{P}[E])$ is a symmetric function.

Note that $\mathrm{gen}(\mathbf{P}[E])$ depends only on the cardinality of E. We are therefore justified in writing

$$\mathrm{gen}(\mathbf{P}[E]) = \mathrm{gen}(\mathbf{P},n) \qquad \text{with} \quad |E| = n.$$

We define the *generating function of the polynomial species* $\mathbf{P}$ to be the formal power series

$$\mathrm{Gen}(\mathbf{P},z) = \sum_{n\geq 0} \mathrm{gen}(\mathbf{P},n)\frac{z^n}{n!}.$$

The generating function of a polynomial species $\mathbf{P}$ is a formal power series in a new variable z, whose coefficients are elements of $\mathbf{Z}[(X)]$, that is, formal polynomials.

As promised, we now define a *category of polynomial species* by setting $\mathrm{Hom}(\mathbf{P},\mathbf{Q})$ equal to the set of all natural equivalences τ between $\mathbf{P}$ and $\mathbf{Q}$ whose components $\tau_E:\mathbf{P}[E]\to\mathbf{Q}[E]$ are bijections such that

$$\tau_E(s,f) = (t,f).$$

In other words, $\tau_E(s,f)$ is an element of $\mathbf{Q}[E]$ whose second coordinate is f.

We write $\mathbf{P}=\mathbf{Q}$, when $\mathbf{P}$ and $\mathbf{Q}$ are naturally equivalent in the category of polynomial species. Clearly, if $\mathbf{P}=\mathbf{Q}$ then $\mathrm{Gen}(\mathbf{P},z)=\mathrm{Gen}(\mathbf{Q},z)$. One verifies that $\boldsymbol{\epsilon}_F$ is a functor from the category of polynomial species to itself that associates to every polynomial species $\mathbf{P}$ the polynomial species $\boldsymbol{\epsilon}_F(\mathbf{P})$.

The functor $\boldsymbol{\epsilon}_F$ is the set-theoretic analog of the linear operator

$$\epsilon_F:\mathbf{Z}[(X)] \to \mathbf{Z}[(X)].$$

In fact, it is easy to prove that

$$\epsilon_F(\mathrm{gen}(\mathbf{P}[E])) = \mathrm{gen}(\boldsymbol{\epsilon}_F(\mathbf{P})[E]).$$

4. Operations on polynomial species

We define the sum and product of polynomial species, as well as a notion of convergence. By passing to generating functions, these notions are seen to be the set-theoretic counterparts of the corresponding notions for formal polynomials.

A. Sum and product of a finite family of polynomial species

Let I be a finite set. The *sum* $\Sigma_{i \in I} P_i$ of the family of polynomial species $(\mathbf{P}_i)_{i \in I}$ is defined as follows:

$$\sum_{i \in I} \mathbf{P}_i[E] = \{((s,i),f) : (s,f) \in \mathbf{P}_i[E] \text{ for some } i \in I\}.$$

That is, for each element s such that $(s,f) \in \mathbf{P}_i[E]$, and for any $i \in I$, the pair $((s,i),f)$ belongs to the set

$$\sum_i \mathbf{P}_i[E].$$

Passing to generating functions, we clearly have

$$\mathrm{Gen}\Big(\sum_{i \in I} \mathbf{P}_i, z\Big) = \sum_{i \in I} \mathrm{Gen}(\mathbf{P}_i, z).$$

The *product* of polynomial species is defined using the notion of composition, which we introduce next.

A *composition* of a set E indexed by a set I (not necessarily finite) is a function

$$k : I \to \mathscr{P}(E)$$

such that:

(1) $k(i) \cap k(j) = \varnothing$ if $i \neq j$;
(2) $\bigcup_{i \in I} k(i) = E$.

In some instances we shall denote a composition of E by

$$\sum_{i \in I} E_i,$$

it being understood that

$$E_i \cap E_j = \varnothing \qquad \text{if} \quad i \neq j$$

and

$$E = \bigcup_{i \in I} E_i.$$

when $E_i \neq \varnothing$ for all $i \in I$, we say that the composition is *strict*.

Let I be a finite set and $(\mathbf{P}_i)_{i\in I}$ be a family of polynomial species. The product $\prod_{i\in I}\mathbf{P}_i$ is defined as follows. For every finite set E, let $\prod_{i\in I}\mathbf{P}_i[E]$ be the set of all pairs (s,f) obtained by the following steps:

(1) Choose a composition k of E indexed by the elements of I.

(2) Choose $(s_i,f_i)\in\mathbf{P}_i[k(i)]$ for every $i\in I$.

(3) Set $s=(k,(s_i)_{i\in I})$, and define f to be the function whose restriction to the set $k(i)$ is f_i; in symbols,

$$f|_{k(i)} = f_i \qquad \text{for every} \quad i\in I.$$

Note that the set $\prod_{i\in I}\mathbf{P}_i[E]$ is empty whenever there is an index $i\in I$ such that $\mathbf{P}_i[k(i)]=\varnothing$ for every composition k.

Proposition 4.1. $\operatorname{Gen}(\prod_{i\in I}\mathbf{P}_i,z)=\prod_{i\in I}\operatorname{Gen}(\mathbf{P}_i,z)$.

Proof: Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be polynomial species. We prove that

$$\operatorname{gen}(\mathbf{P}_1\times\mathbf{P}_2,n) = \sum_{i+j=n}\binom{n}{i}\operatorname{gen}(\mathbf{P}_1,i)\operatorname{gen}(\mathbf{P}_2,j).$$

When $|E|=n$, we have

$$\begin{aligned}
\operatorname{gen}(\mathbf{P}_1\times\mathbf{P}_2,n) &= \sum_{(s,f)\in\mathbf{P}_1\times\mathbf{P}_2[E]}\operatorname{gen}(s,f)\\
&= \sum_{E_1+E_2=E}\left(\sum_{\substack{(s_1,f_1)\in\mathbf{P}_1[E_1]\\(s_2,f_2)\in\mathbf{P}_2[E_2]}}\prod_{x\in X}\left(x^{|f_1^{-1}(x)|+|f_2^{-1}(x)|}\right)\right)\\
&= \sum_{E_1+E_2=E}\left(\sum_{\substack{(s_1,f_1)\in\mathbf{P}_1[E_1]\\(s_2,f_2)\in\mathbf{P}_2[E_2]}}\operatorname{gen}(s_1,f_1)\operatorname{gen}(s_2,f_2)\right)\\
&= \sum_{E_1+E_2=E}\left(\sum_{(s_1,f_1)\in\mathbf{P}_1[E_1]}\operatorname{gen}(s_1,f_1)\sum_{(s_2,f_2)\in\mathbf{P}_2[E_2]}\operatorname{gen}(s_2,f_2)\right)\\
&= \sum_{i+j=n}\binom{n}{i}\operatorname{gen}(\mathbf{P}_1,i)\operatorname{gen}(\mathbf{P}_2,j).
\end{aligned}$$

It is easy to verify that

$$\boldsymbol{\epsilon}_F(\mathbf{P}_1+\mathbf{P}_2) = \boldsymbol{\epsilon}_F(\mathbf{P}_1)+\boldsymbol{\epsilon}_F(\mathbf{P}_2),$$

$$\boldsymbol{\epsilon}_F(\mathbf{P}_1\cdot\mathbf{P}_2) = \boldsymbol{\epsilon}_F(\mathbf{P}_1)\cdot\boldsymbol{\epsilon}_F(\mathbf{P}_2).$$

B. *Infinite sum and infinite product of polynomial species*

We first give the definition of the limit of a family of polynomial species.

Let J be a directed set. We say that the family of polynomial species $(\mathbf{P}_j)_{j \in J}$ *converges to the polynomial species* $\mathbf{P}$ when there exist an index $j(n,F) \in J$ (dependent on $n \geq 0$ and $F \subseteq X$ alone) and a bijection

$$\phi(j,F,E) : \epsilon_F(\mathbf{P}_j)[E] \to \epsilon_F(\mathbf{P})[E] \qquad \text{for every} \quad j \geq j(n,F)$$

$$\text{gen}(\mathbf{C}[E]) = (|E|-1)!\,p_{|E|}(X),$$

such that:

(1) $\phi(j,F,E)(s,f) = (s',f)$. In other words, $\phi(j,F,E)(s,f)$ is an element of $\epsilon_F(\mathbf{P})[E]$ whose second coordinate is f.

(2) For every bijection $u : E \to E'$ the diagram

$$\begin{array}{ccc} \epsilon_F(\mathbf{P}_j)[E] & \xrightarrow{\phi(j,F,E)} & \epsilon_F(\mathbf{P})[E] \\ \downarrow{\scriptstyle \mathbf{P}_j[u]} & & \downarrow{\scriptstyle \mathbf{P}[u]} \\ \epsilon_F(\mathbf{P}_j)[E'] & \xrightarrow{\phi(j,F,E')} & \epsilon_F(\mathbf{P})[E'] \end{array}$$

is commutative.

When $\mathbf{P}$ is the limit of the family $(\mathbf{P}_j)_{j \in J}$ we write

$$\lim_j \mathbf{P}_j = \mathbf{P}.$$

We remark that if $\mathbf{P}$ and $\mathbf{P}'$ are limits of the family $(\mathbf{P}_j)_{j \in J}$, there exists a natural equivalence in the category of polynomial species between $\mathbf{P}$ and $\mathbf{P}'$.

PROPOSITION 4.2. $\lim_j \text{Gen}(\mathbf{P}_j, z) = \text{Gen}(\lim_j \mathbf{P}_j, z)$.

Proof: The family $(\text{gen}(\mathbf{P}_j, n))_j$ converges to $\text{gen}(\mathbf{P}, n)$. In fact, for every F and for every $j \geq j(n,F)$ we have

$$\epsilon_F(\text{gen}(\mathbf{P}_j,n)) = \text{gen}(\epsilon_F(\mathbf{P}_j),n) = \text{gen}(\epsilon_F(\mathbf{P}),n) = \epsilon_F(\text{gen}(\mathbf{P},n)).$$

We note that every polynomial species is the limit of the family

$$(\epsilon_F(\mathbf{P}))_F$$

as F ranges on the (directed) set of all finite subsets of X.

We say that a family $(\mathbf{P}_i)_{i \in I}$ of polynomial species is *summable* when the family

$$\left(\sum_{i \in J} \mathbf{P}_i\right)_J,$$

as J ranges on the set of all finite subsets of I, converges. We denote the limit of the family $(\mathbf{P}_i)_{i \in I}$ by

$$\sum_{i \in I} \mathbf{P}_i .$$

PROPOSITION 4.3. *A family* $(\mathbf{P}_i)_{i \in I}$ *of polynomial species is summable if and only if the set*

$$I(n,F) = \{i \in I : \epsilon_F(\mathbf{P}_i)[E] \neq \varnothing\}$$

is finite for every finite subset F *of* X *and for every* $n = |E|$.

Proof: Suppose that the family $(\mathbf{P}_i)_{i \in I}$ is summable. Then there exists the limit

$$\lim_J \sum_{i \in J} \mathbf{P}_i = \mathbf{P}.$$

This means that whenever F is a finite subset of X, there exist a finite subset $J(n,F)$ of I and a bijection

$$\phi(J,F,E) : \epsilon_F\Big(\sum_{i \in J} \mathbf{P}_i\Big)[E] \to \epsilon_F(\mathbf{P})[E]$$

for every finite subset J of I that contains $J(n,F)$.

Now, if $j \notin J(n,F)$, taking $J = J(n,F) \cup \{j\}$, we have that the following map is also bijective:

$$\phi(J,F,E)^{-1} \circ \phi(J(n,F),F,E) : \epsilon_F\Big(\sum_{i \in J(n,F)} \mathbf{P}_i\Big)[E] \to \epsilon_F\Big(\sum_{i \in J} \mathbf{P}_i\Big)[E]. \tag{4.4}$$

On the other hand,

$$\epsilon_F\Big(\sum_{i \in J} \mathbf{P}_i\Big) = \epsilon_F\Big(\sum_{i \in J(n,F)} \mathbf{P}_i\Big) + \epsilon_F(\mathbf{P}_j) \tag{4.5}$$

Thus, from (4.4) and (4.5), we infer

$$\epsilon_F(\mathbf{P}_j)[E] = \varnothing$$

for every $j \notin J(n,F)$.

Conversely, if the set $I(n,F)$ is finite for every n and F, the polynomial species $\mathbf{P}$ defined as

$$\mathbf{P}[E] = \bigcup_F \left\{(s,f) \in \left(\sum_{i \in I(n,F)} \mathbf{P}_i\right)[E] : \operatorname{Im} f = F\right\},$$

where F ranges on the set of all finite sets of X, is clearly the sum of the family $(\mathbf{P}_i)_{i \in I}$.

We say that a polynomial species $\mathbf{P}$ is *without constant term* if $\mathbf{P}[\varnothing] = \varnothing$.
We denote by $\mathbf{1}$ the polynomial species defined as follows:

$$\mathbf{1}[\varnothing] = \{(\varnothing; f_\varnothing)\},$$

$$\mathbf{1}[E] = \varnothing \quad \text{if} \quad E \neq \varnothing.$$

The species $\mathbf{1}$ is not a species without constant term.
A family $(\mathbf{P}_i)_{i \in I}$ of polynomial species, almost all without constant term, is said to be *multipliable* when the family

$$\left(\prod_{i \in J} (\mathbf{1} + \mathbf{P}_i)\right)_J,$$

as J ranges on the direct set of all finite subsets of I, converges. When this is the case, we denote the limit by

$$\prod_{i \in I} (\mathbf{1} + \mathbf{P}_i).$$

Proposition 4.6. *A family of polynomial species $(\mathbf{P}_i)_{i \in I}$, all without constant term, is summable if and only if it is multipliable.*

Proof: Suppose that the family $(\mathbf{P}_i)_{i \in I}$ is summable. From Proposition 4.3 it follows that the set

$$I(n,F) = \{i \in I : \epsilon_F(\mathbf{P}_i)[E] \neq \varnothing\}$$

is finite for every F and $n = |E|$. Thus the polynomial species $\mathbf{P}$ defined as

$$\mathbf{P}[E] = \bigcup_F \left\{(s,f) \in \left(\prod_{i \in I(n,F)} \mathbf{1} + \mathbf{P}_i\right)[E] : \operatorname{Im} f = F\right\},$$

as F ranges on the set of all finite subsets of X, is the product of the family $(\mathbf{P}_i)_{i \in I}$.

Conversely, if the family $(\mathbf{P}_i)_{i\in I}$ is multipliable, then the limit

$$\lim_J\Big(\prod_{i\in J}\mathbf{1}+\mathbf{P}_i\Big) = \mathbf{P}$$

exists. Hence, whenever F is a finite subset of X, there exists a finite subset $J(n,F)$ of I and a bijection

$$\phi(J,F,E):\boldsymbol{\epsilon}_F\Big(\prod_{i\in J}\mathbf{P}_i\Big)[E] \to \boldsymbol{\epsilon}_F(\mathbf{P})[E]$$

for every finite subset J of I that contains $J(n,F)$.

If $j\notin J(n,F)$, take $J=J(n,F)\cup\{j\}$. Then the following map is also bijective:

$$\phi(J,F,E)^{-1}\circ\phi(J(n,F),F,E):\boldsymbol{\epsilon}_F\Big(\prod_{i\in J(n,F)}\mathbf{1}+\mathbf{P}_i\Big)[E]$$

$$\to \boldsymbol{\epsilon}_F\Big(\prod_{i\in J}\mathbf{1}+\mathbf{P}_i\Big)[E]. \tag{4.7}$$

On the other hand,

$$\boldsymbol{\epsilon}_F\Big(\prod_{i\in J}\mathbf{1}+\mathbf{P}_i\Big) = \boldsymbol{\epsilon}_F\Big(\prod_{i\in J(n,F)}\mathbf{1}+\mathbf{P}_i\Big)\cdot\boldsymbol{\epsilon}_F(\mathbf{1}+\mathbf{P}_j). \tag{4.8}$$

Thus from (4.7) and (4.8) we have

$$\boldsymbol{\epsilon}_F(\mathbf{P}_j)[E] = \varnothing$$

for every $j\notin J(n,F)$, that is, the family $(\mathbf{P}_i)_{i\in I}$ is summable.

We remark that when $(\mathbf{P}_i)_{i\in I}$ is a multipliable family of polynomial species, then the infinite sum and the infinite product are explicitly given as follows:

$$\Big(\sum_{i\in I}\mathbf{P}_i\Big)[E] = \bigcup_F\Big\{(s,f)\in\sum_{i\in I(n,F)}\mathbf{P}_i[E]:\operatorname{Im}f=F\Big\},$$

and

$$\Big(\prod_{i\in I}\mathbf{1}+\mathbf{P}_i\Big)[E] = \bigcup_F\Big\{(s,f)\in\Big(\prod_{i\in I(n,F)}\mathbf{1}+\mathbf{P}_i\Big)[E]:\operatorname{Im}f=F\Big\},$$

where $I(n,F)=\{i\in I:\boldsymbol{\epsilon}_F(\mathbf{P}_i)[E]\neq\varnothing\}$.

5. Some special polynomial species

Example 5.1: We define the polynomial species $\mathbf{A}_x$, for every $x \in X$, as follows:

$$\mathbf{A}_x[E] = \begin{cases} E \times \mathrm{Hom}(E, \{x\}) & \text{if } \quad |E| = 1, \\ \varnothing & \text{if } \quad |E| \neq 1. \end{cases}$$

The generating function of the polynomial species $\mathbf{A}_x$ is clearly

$$\mathrm{Gen}(\mathbf{A}_x, z) = xz.$$

Example 5.2: (The elementary symmetric species): This polynomial species, denoted by **A**, is defined as follows:

$$\mathbf{A}[\varnothing] = \{(\varnothing, f_\varnothing)\},$$

$$\mathbf{A}[E] = \{(E, f) : f : E \to X \text{ is an injective function}\}.$$

One verifies that

$$\mathrm{gen}(\mathbf{A}[E]) = |E|! e_{|E|}(X),$$

where $e_n(X)$ is the elementary symmetric function defined by

$$e_n(X) = \sum_m x^m,$$

where the sum ranges over all multisets m such that $|m| = n = |\mathrm{supp}(m)|$ (that is, over all multisets that "are" sets).

The generating function of the elementary symmetric species is

$$\mathrm{Gen}(\mathbf{A}, z) = \sum_{n \geq 0} n! e_n(X) \frac{z^n}{n!} = \sum_{n \geq 0} e_n(X) z^n.$$

Example 5.3: The species $\mathbf{H}_x$ is defined, for every $x \in X$, as follows:

$$\mathbf{H}_x[\varnothing] = \varnothing,$$

and if $E \neq \varnothing$,

$$\mathbf{H}_x[E] = \{(\sigma, f)\}$$

where $f : E \to X$ is the function taking the constant value x, and σ is a permutation of the set E. One verifies that

$$\operatorname{gen}((\mathbf{1}+\mathbf{H}_x)[E]) = |E|!x^{|E|},$$

and hence that

$$\operatorname{Gen}(\mathbf{1}+\mathbf{H}_x, z) = \sum_{n\geq 0} n!x^n\frac{z^n}{n!} = (1-xz)^{-1}.$$

We digress to present what we believe to be a fundamental combinatorial result.

THEOREM 5.4. *Let m be a finite multiset on X, and let E be a finite set such that $|m|=|E|$. Denote by*

$$[m;E]$$

the set of all pairs (d,f) where:

(1) *$f: E \to X$ is a function such that $\operatorname{gen}(f)=x^m$;*

(2) *$d=\{\sigma_B : B\in \operatorname{Ker} f\}$ is a family of permutations, one defined on each block B of $\operatorname{Ker} f$.*

Under these conditions, the cardinality of the set $[m;E]$ is $|E|!$.

We stress the fact that the cardinality of the set $[m;E]$ does not depend on m.

Proof: Set

$$x^m = \prod_{x\in F} x^{\overline{m}(x)},$$

where F is the finite subset of X such that $\overline{m}(x)>0$ for every $x\in F$. Then we have

$$\sum_{x\in F} \overline{m}(x) = |m|.$$

Let

$$[[m;E]]$$

be the set of all pairs

$$(\tau, k)$$

where:

(1) k is a composition of E indexed on F such that $|k(x)| = \overline{m}(x)$ for every $x \in F$;

(2) τ is a family $\{\sigma_x\}_{x \in F}$ of permutations, where σ_x is a permutation of the set $k(x)$.

We claim that the sets $[m; E]$ and $[[m; E]]$ are naturally bijective. To see this, it suffices to realize that every function f from E to F defines a composition k by setting

$$k(x) = f^{-1}(x).$$

Conversely, given a composition k, we obtain a function $f: E \to F$ by setting

$$f(s) = x$$

whenever $s \in k(x)$.

Now, the cardinality of the set $[[m; E]]$ is

$$\frac{|E|!}{\prod_{x \in F} \overline{m}(x)!} \prod_{x \in F} \overline{m}(x)! = |E|!,$$

since $|E|!/\prod_{x \in F} \overline{m}(x)!$ is the number of compositions k of the set E such that

$$|k(x)| = \overline{m}(x).$$

Let us now continue to define new polynomial species.

Example 5.5: The *symmetric species* **H** *of dispositions* is defined as follows:

$$\mathbf{H}[\varnothing] = \{(\varnothing, f_\varnothing)\}$$

and if $E \neq \varnothing$

$$\mathbf{H}[E] = \{(d, f)\},$$

where $f: E \to X$ is an arbitrary function, and for each fiber $f^{-1}(x)$, a permutation is defined on $f^{-1}(x)$. The set of such permutations is denoted by d.

The generating function of the polynomial species of dispositions is

$$\mathrm{Gen}(\mathbf{H}, z) = \sum_{n \geq 0} n! h_n(X) \frac{z^n}{n!} = \sum_{n \geq 0} h_n(X) z^n,$$

where $h_n(X)$ is the complete homogeneous symmetric function of degree n, that is,

$$h_n(X) = \sum_m x^m,$$

where the sum ranges over all multisets m such that $|m| = n$. In fact, the preceding theorem gives

$$\operatorname{gen}(\mathbf{H}[E]) = |E|!h_{|E|}(X).$$

Example 5.6: The *cyclic polynomial species* **C** is defined as

$$\mathbf{C}[\varnothing] = \varnothing$$

and if $E \neq \varnothing$

$$\mathbf{C}[E] = \{(\gamma, f)\},$$

where $f: E \to X$ is a function such that

$$f(e) = f(e') \quad \text{for all} \quad e, e' \in E,$$

and $\gamma: E \to E$ is a cyclic permutation.

One verifies that

$$\operatorname{gen}(\mathbf{C}[E]) = (|E| - 1)!p_{|E|}(X),$$

where $p_n(X)$ is the power-sum symmetric functions, defined as

$$p_n(X) = \sum_{x \in X} x^n;$$

thus

$$\operatorname{Gen}(\mathbf{C}, z) = \sum_{n \geq 1} (n-1)!p_n(X)\frac{z^n}{n!} = \sum_{n \geq 1} p_n(X)\frac{z^n}{n}.$$

6. Assemblies of polynomial species

Let **P** be a polynomial species without constant term, and let E be a finite set.

An *assembly of polynomial species* **P** *on the partition* $\pi = \{B\}_{B \in \pi}$ of E is a pair (r, f) defined as follows:

(1) Choose a pair (s_B, f_B) in $\mathbf{P}[B]$, for every block $B \in \pi$.
(2) Set $r = (\pi, (s_B)_B)$.
(3) Let f be the function defined as follows: $f|_B = f_B$ for every $B \in \pi$.

We define the hth *divided power* $\mathbf{P}^{(h)}$ of the polynomial species **P** without constant term, by setting $\mathbf{P}^{(h)}[E]$ equal to the set of all polynomial assemblies (r, f) on all partition π of E with h blocks.

Note that $\mathbf{P}^{(h)}[\varnothing] = \varnothing$. We set $\mathbf{P}^{(0)} = \mathbf{1}$. We have now

$$\operatorname{Gen}(\mathbf{P}^{(h)}, z) = \frac{\operatorname{Gen}(\mathbf{P}^{(h)}, z)}{h!}. \tag{6.1}$$

It is easy to prove the identity (6.1) by the following remark. Let k and k' be strict compositions of E indexed on I and I' respectively. Say that the compositions k and k' are equivalent when there exists a bijection

$$\phi : I \to I'$$

such that $k' \circ \phi = k$. An equivalence class of strict compositions determines a partition of E, and conversely.

From (6.1) and from the product identity for generating functions we have

$$\operatorname{Gen}(\mathbf{P}^{h}, z) = \operatorname{Gen}(\mathbf{P}, z)^{h}$$

and hence

$$\operatorname{Gen}(\mathbf{P}^{(h)}, z) = \frac{\operatorname{Gen}(\mathbf{P}, z)^{h}}{h!}.$$

The family $\{\mathbf{P}^{(h)} : h \geq 0\}$ is summable, since for any finite set E, the sets

$$\mathbf{P}^{(h)}[E]$$

are almost all equal to the empty set.

We can therefore define the *exponential species* $\operatorname{Exp}(\mathbf{P})$ of the polynomial species $\mathbf{P}$ as the sum of the family $(\mathbf{P}^{(h)})$, that is,

$$\operatorname{Exp}(\mathbf{P}) = \sum_{h \geq 0} \mathbf{P}^{(h)}.$$

Note that $\operatorname{Exp}(\mathbf{P})[E] = \bigcup_{h \geq 0} \mathbf{P}^{(h)}[E]$. Clearly

$$\operatorname{Gen}(\operatorname{Exp}(\mathbf{P}), z) = \sum_{h \geq 0} \frac{\operatorname{Gen}(\mathbf{P}, z)^{h}}{h!} x^{h} = \exp(\operatorname{Gen}(\mathbf{P}, z)).$$

7. Some identities on symmetric species

We first prove the following species-theoretic generalization of the classical product identity for elementary symmetric functions:

PROPOSITION 7.1. $\prod_{x \in X}(\mathbf{1} + \mathbf{A}_x) = \mathbf{A}$.

Proof: Recall that the equality sign stands for natural equivalence in the category of polynomial species.

We note that the family $(\mathbf{A}_x)_{x \in X}$ is multipliable. Indeed, we have:

(1) $\mathbf{A}_x$ is without constant term.

(2) The set $I(n,F)=\{x \in X : \epsilon_F(\mathbf{A}_x)[E] \neq \varnothing\}$ is equal to F; thus $I(n,F)$ is finite and does not depend on $|E|$.

The infinite product of the family $(\mathbf{A}_x)_{x \in X}$ is the polynomial species $\mathbf{P}$ defined by

$$\mathbf{P}[E] = \bigcup_F \left\{(s,f) \in \prod_{x \in F} (\mathbf{1}+\mathbf{A}_x)[E] : \operatorname{Im} f = F\right\},$$

and if $(s,f) \in \mathbf{P}[E]$, then the function f is injective.

On the other hand, there is a natural bijection τ_E between the set $\mathbf{P}[E]$ and the set $\mathbf{A}[E]$ given by

$$\tau_E(s,f) = (E,f),$$

where $(s,f) \in \mathbf{P}[E]$. Thus $\mathbf{P} = \mathbf{A}$.

We have thus provided a bijective interpretation of the classical identity

$$\sum_{n \geq 0} n! e_n(X) \frac{z^n}{n!} = \prod_{x \in X} (1 + xz).$$

PROPOSITION 7.2. $\mathbf{H} = \prod_{x \in X}(\mathbf{1}+\mathbf{H}_x)$.

Proof: Note that the family $(\mathbf{H}_x)_{x \in X}$ is multipliable. Indeed, we have:

(1) The polynomial species $\mathbf{H}_x$ is without constant term.

(2) The set $I(n,F)=\{x \in X : \epsilon_F(\mathbf{H}_x)[E] \neq \varnothing\}$ is equal to F; thus $I(n,F)$ is finite and does not depend on $|E|$.

The infinite product of the family $(\mathbf{H}_x)_{x \in X}$ is the polynomial species $\mathbf{P}$ defined by

$$\mathbf{P}[E] = \bigcup_F \left\{(s,f) \in \prod_{x \in F} (\mathbf{1}+\mathbf{H}_x)[E] : \operatorname{Im} f = F\right\}.$$

We define a natural equivalence between $\mathbf{P}$ and $\mathbf{H}$ as follows. Let $(s,f) \in \mathbf{P}[E]$; then (s,f) is an element of $\prod_{x \in F}(\mathbf{1}+\mathbf{H}_x)[E]$, where $F = \operatorname{Im} f$. That is, $s = (k,\{\sigma_x\}_{x \in F})$, where k is a composition of E indexed on F, and σ_x is a permutation of $k(x) = f(x)$ when $k(x) \neq \varnothing$.

Let d be the set defined by

$$d = \left\{\sigma_x : \sigma_x \text{ is a permutation of } f^{-1}(x)\right\}.$$

Thus the bijection that associates the pair $(s,f) \in \mathbf{P}[E]$ to the disposition (d,f) is natural.

We have thus obtained a bijective interpretation of the symmetric-function identity

$$\sum_{n\geq 0} n!h_n(X)\frac{z^n}{n!} = \prod_{x\in X}(1-xz)^{-1}.$$

Our last example will be a bijective interpretation of Waring's formula.

PROPOSITION 7.3. $\mathbf{H} = \mathrm{Exp}(\mathbf{C})$.

Proof: Let α be an element of $\mathrm{Exp}(\mathbf{C})[E]$. Thus, α contains the following data:

(1) a partition π of E,
(2) for each $B \in \pi$, a pair (γ_B, f_B), where $f_B : B \to X$ is a function taking a constant value, and where γ_B is a cyclic permutation of the set B.

Now, for every $x \in X$, set

$$B(x) = \bigcup \{B : f_B(B) = x\}.$$

The set of sets

$$\{B(x) : x \in X,\ B(x) \neq \varnothing\}$$

is a partition of E.
The set of cyclic permutations

$$\{\gamma_B : B \in \pi,\ B \subseteq B(x)\}$$

defines a permutation of the set $B(x)$, which we denote by $\sigma(x)$.
We define a function $f : E \to X$ by setting

$$f|_B = f_B.$$

Then $f|_{B(x)}$ takes the constant value x, and the pair

$$(\{\sigma(x) : B(x) \neq \varnothing\}, f)$$

is a disposition and thus an element of $\mathbf{H}[E]$.
Conversely, let $(d, f) \in \mathbf{H}[E]$ be a disposition. Let

$$\pi = \{f^{-1}(x) : f^{-1}(x) \neq \varnothing\}.$$

Thus π is a partition of E. On every block $B \in \pi$ a permutation $\sigma(B)$ is defined. Let $\pi(B)$ be the partition of B such that for every $B' \in \pi(B)$ the restriction $\sigma(B')$ of $\sigma(B)$ to B' is a cyclic permutation.

Clearly, for every $B' \in \pi(B)$ the restriction $f|_{B'}$ is a function taking a constant value. Thus,

$$(\sigma(B'), f|_{B'}) \in \mathbf{C}[B']$$

defines an assembly of cyclic species on the partition $\{B' : B' \in \pi(B),\ B \in \pi\}$. The algorithm is clearly bijective.

We have thus provided a bijective interpretation of Waring's formula:

$$\sum_{n \geq 0} n! h_n(x) \frac{z^n}{n!} = \sum_{n \geq 0} h_x(x) z^n = \exp\left(\sum_{h \geq 1} (h-1)! p_h(x) \frac{z^h}{h!} \right)$$

$$= \exp\left(\sum_{h \geq 1} p_h(x) \frac{z^h}{h} \right).$$

Our interpretation is valid for any set X of variables.

References

1. M. Aigner, *Combinatorial Theory*, Springer-Verlag, New York, 1979.
2. F. Bergeron, A Combinatorial Outlook on Symmetric Functions, Rapport de Recherche 31, Univ. du Québec, 1987.
3. F. Bonetti, G.-C. Rota, A. M. Venezia, and D. Senato, Symmetric functions and symmetric species, *Ann. Discrete Math.* 30:107–114 (1986).
4. N. Bourbaki, *Algèbre*, Hermann, Paris, 1968.
5. N. Bourbaki, *Topologie Générale*, Hermann, Paris, 1971.
6. P. Doubilet, On the foundations of combinatorial theory VII: Symmetric functions through the theory of distribution and occupancy, *Stud. Appl. Math.* 51:377–396 (1972).
7. P. Doubilet, G.-C. Rota, and R. Stanley, On the foundations of combinatorial theory VI: The idea of generating function, in *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability,* Univ. of California Press, vol. 2, 1972, pp. 267–318.
8. C. Ehresmann, *Categories et Structures*, Dunod, Paris, 1965.
9. D. Foata (Ed.), *Combinatoire et Représentation du Groupe Symétrique*, Strasbourg, 1976, Lecture Notes in Math. 579, Springer-Verlag, New York, 1977.
10. A. Joyal, Une théorie combinatoire des séries formelles, *Adv. Math.* 42:1–82 (1981).
11. A. Joyal, *Foncteurs Analytiques et Espèces de Structures*, Lecture Notes in Math. 1234, 1986, pp. 126–159.
12. A. Lascoux and M. P. Schutzemberger, *Formulaire Raisonné de Functions Symétriques*, 1985.
13. I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Claredon, Oxford, 1979.
14. S. MacLane, *Categories for the Working Mathematicians*, Springer-Verlag, New York, 1971.
15. N. Metropolis, G. Nicoletti, and G.-C. Rota, A new class of symmetric functions, *Adv. in Math. Suppl. Stud.* 7B:563–575 (1981).
16. G.-C. Rota, Baxter algebras and combinatorial identities, Parts 1, 2, *Bull. Amer. Math. Soc.*, 1969, pp. 325–329, 330–334.
17. D. Senato, A bijective proof of the fundamental theorem on symmetric functions, *Stud. Appl. Math.* 81:1–5 (1989).
18. R. Stanley, Acyclic orientations of graphs, *Discrete Math.* 5:171–178 (1973).
19. R. Stanley, Theory and applications of plane partitions, parts 1, 2, *Stud. Appl. Math.* 50:167–188, 259–279 (1971).

Università di Bologna
Massachusetts Institute of Technology
Università di Napoli
Università "La Sapienza" Roma

(Received April 27, 1989)