TOPICS IN ALGEBRAIC COMBINATORICS

Richard P. Stanley

Course notes for 18.312 (Algebraic Combinatorics)
M.I.T., Spring 2010
Preliminary (incomplete) version of 24 April 2010

Acknowledgment. I am grateful to Sergey Fomin for his careful reading of the manuscript and for several helpful suggestions.

1 Walks in graphs.

Given a finite set S and integer $k \geq 0$, let $\binom{S}{k}$ denote the set of k-element subsets of S, and let $\left(\!\binom{S}{k}\!\right)$ denote the set of k-element multisubsets (sets with repeated elements) on S. For instance, if $S = \{1, 2, 3\}$ then (using abbreviated notation),

$$\binom{S}{2} = \{12, 13, 23\}, \quad \left(\!\!\binom{S}{2}\!\!\right) = \{11, 22, 33, 12, 13, 23\}.$$

A (finite) *graph* G consists of a *vertex set* $V = \{v_1, \dots, v_p\}$ and *edge set* $E = \{e_1, \dots, e_q\}$, together with a function $\varphi : E \to \left(\!\binom{V}{2}\!\right)$. We think that if $\varphi(e) = uv$ (short for $\{u, v\}$), then e connects u and v or equivalently e is *incident* to u and v. If there is at least one edge incident to u and v then we say that the vertices u and v are *adjacent*. If $\varphi(e) = vv$, then we call e a *loop* at v. If several edges $e_1, \dots, e_j$ $(j > 1)$ satisfy $\varphi(e_1) = \cdots = \varphi(e_j) = uv$, then we say that there is a *multiple edge* between u and v. A graph without loops or multiple edges is called *simple*. In this case we can think of E as just a subset of $\binom{V}{2}$ [why?].

The *adjacency matrix* of the graph G is the $p \times p$ matrix $\boldsymbol{A} = \boldsymbol{A}(G)$, over the field of complex numbers, whose (i, j)-entry a_{ij} is equal to the number of edges incident to v_i and v_j. Thus $\boldsymbol{A}$ is a real symmetric matrix (and hence has real eigenvalues) whose trace is the number of loops in G.

A *walk* in G of *length* ℓ from vertex u to vertex v is a sequence $v_1, e_1, v_2, e_2, \ldots$, v_ℓ, e_ℓ, $v_{\ell+1}$ such that:

- each v_i is a vertex of G
- each e_j is an edge of G
- the vertices of e_i are v_i and v_{i+1}, for $1 \leq i \leq \ell$
- $v_1 = u$ and $v_{\ell+1} = v$.

1.1 Theorem. *For any integer $\ell \geq 1$, the (i,j)-entry of the matrix $\boldsymbol{A}(G)^\ell$ is equal to the number of walks from v_i to v_j in G of length ℓ.*

Proof. This is an immediate consequence of the definition of matrix multiplication. Let $\boldsymbol{A} = (a_{ij})$. The (i,j)-entry of $\boldsymbol{A}(G)^\ell$ is given by

$$(\boldsymbol{A}(G)^\ell)_{ij} = \sum a_{ii_1} a_{i_1 i_2} \cdots a_{i_{\ell-1} j},$$

where the sum ranges over all sequences $(i_1, \ldots, i_{\ell-1})$ with $1 \leq i_k \leq p$. But since a_{rs} is the number of edges between v_r and v_s, it follows that the summand $a_{ii_1} a_{i_1 i_2} \cdots a_{i_{\ell-1} j}$ in the above sum is just the number (which may be 0) of walks of length ℓ from v_i to v_j of the form

$$v_i, e_1, v_{i_1}, e_2, \ldots, v_{i_{\ell-1}}, e_\ell, v_j$$

(since there are a_{ii_1} choices for e_1, $a_{i_1 i_2}$ choices for e_2, etc.) Hence summing over all $(i_1, \ldots, i_{\ell-1})$ just gives the total number of walks of length ℓ from v_i to v_j, as desired. $\square$

We wish to use Theorem 1.1 to obtain an explicit formula for the number $(\boldsymbol{A}(G)^\ell)_{ij}$ of walks of length ℓ in G from v_i to v_j. The formula we give will depend on the eigenvalues of $\boldsymbol{A}(G)$. The eigenvalues of $\boldsymbol{A}(G)$ are also called simply the *eigenvalues of* G. Recall that a real symmetric $p \times p$ matrix M has p linearly independent real eigenvectors, which can in fact be chosen to be orthonormal (i.e., orthogonal and of unit length). Let $u_1, \ldots, u_p$ be real orthonormal unit eigenvectors for M, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_p$. All vectors u will be regarded as $p \times 1$ *column* vectors. We let ^{t} denote transpose, so u^t is a $1 \times p$ *row* vector. Thus the dot (or scalar or inner) product of the vectors u and v is given by $u^t v$ (ordinary matrix multiplication). In particular, $u_i^t u_j = \delta_{ij}$ (the Kronecker delta). Let $U = (u_{ij})$ be the matrix whose columns are $u_1, \dots, u_p$, denoted $U = [u_1, \dots, u_p]$. Thus U is an orthogonal matrix and

$$U^t = U^{-1} = \left[\begin{array}{c} u_1^t \\ \cdot \\ \cdot \\ \cdot \\ u_p^t \end{array} \right],$$

the matrix whose rows are $u_1^t, \dots, u_p^t$. Recall from linear algebra that the matrix U *diagonalizes* M, i.e.,

$$U^{-1} M U = \mathrm{diag}(\lambda_1, \dots, \lambda_p),$$

where $\mathrm{diag}(\lambda_1, \dots, \lambda_p)$ denotes the diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_p$. In fact, we have

$$MU = [\lambda_1 u_1, \dots, \lambda_p u_p]$$

$$(U^{-1} M U)_{ij} = (U^t M U)_{ij} = \lambda_j u_i^t u_j = \lambda_j \delta_{ij}.$$

1.2 Corollary. *Given the graph G as above, fix the two vertices v_i and v_j. Let $\lambda_1, \dots, \lambda_p$ be the eigenvalues of the adjacency matrix $\boldsymbol{A}(G)$. Then there exist real numbers $c_1, \dots, c_p$ such that for all $\ell \geq 1$, we have*

$$(\boldsymbol{A}(G)^\ell)_{ij} = c_1 \lambda_1^\ell + \cdots + c_p \lambda_p^\ell.$$

In fact, if $U = (u_{rs})$ is a real orthogonal matrix such that

$$U^{-1} \boldsymbol{A} U = \mathrm{diag}(\lambda_1, \dots, \lambda_p),$$

then we have

$$c_k = u_{ik} u_{jk}.$$

Proof. We have [why?]

$$U^{-1} \boldsymbol{A}^\ell U = \mathrm{diag}(\lambda_1^\ell, \dots, \lambda_p^\ell).$$

Hence

$$\boldsymbol{A}^\ell = U \cdot \mathrm{diag}(\lambda_1^\ell, \ldots, \lambda_p^\ell) U^{-1}.$$

Taking the (i, j)-entry of both sides (and using $U^{-1} = U^t$) gives [why?]

$$(\boldsymbol{A}^\ell)_{ij} = \sum_k u_{ik} \lambda_k^\ell u_{jk},$$

as desired. □

In order for Corollary 1.2 to be of any use we must be able to compute the eigenvalues $\lambda_1, \ldots, \lambda_p$ as well as the diagonalizing matrix U (or eigenvectors u_i). There is one interesting special situation in which it is not necessary to compute U. A *closed walk* in G is a walk that ends where it begins. The number of closed walks in G of length ℓ starting at v_i is therefore given by $(\boldsymbol{A}(G)^\ell)_{ii}$, so the *total* number $f_G(\ell)$ of closed walks of length ℓ is given by

$$\begin{aligned} f_G(\ell) &= \sum_{i=1}^p (\boldsymbol{A}(G)^\ell)_{ii} \\ &= \mathrm{tr}(\boldsymbol{A}(G)^\ell), \end{aligned}$$

where tr denotes trace (sum of the main diagonal entries). Now recall that the trace of a square matrix is the sum of its eigenvalues. If the matrix M has eigenvalues $\lambda_1, \ldots, \lambda_p$ then [why?] M^ℓ has eigenvalues $\lambda_1^\ell, \ldots, \lambda_p^\ell$. Hence we have proved the following.

1.3 Corollary. *Suppose $\boldsymbol{A}(G)$ has eigenvalues $\lambda_1, \ldots, \lambda_p$. Then the number of closed walks in G of length ℓ is given by*

$$f_G(\ell) = \lambda_1^\ell + \cdots + \lambda_p^\ell.$$

We now are in a position to use various tricks and techniques from linear algebra to count walks in graphs. Conversely, it is sometimes possible to count the walks by combinatorial reasoning and use the resulting formula to determine the eigenvalues of G. As a first simple example, we consider the *complete graph* K_p with vertex set $V = \{v_1, \ldots, v_p\}$, and one edge between any two *distinct* vertices. Thus K_p has p vertices and $\binom{p}{2} = \frac{1}{2}p(p-1)$ edges.

1.4 Lemma. *Let J denote the $p \times p$ matrix of all 1's. Then the eigenvalues of J are p (with multiplicity one) and 0 (with multiplicity $p-1$).*

Proof. Since all rows are equal and nonzero, we have $\mathrm{rank}(J) = 1$. Since a $p \times p$ matrix of rank $p-m$ has at least m eigenvalues equal to 0, we conclude that J has at least $p-1$ eigenvalues equal to 0. Since $\mathrm{tr}(J) = p$ and the trace is the sum of the eigenvalues, it follows that the remaining eigenvalue of J is equal to p. $\square$

1.5 Proposition. *The eigenvalues of the complete graph K_p are as follows: an eigenvalue of -1 with multiplicity $p-1$, and an eigenvalue of $p-1$ with multiplicity one.*

Proof. We have $\boldsymbol{A}(K_p) = J - I$, where I denotes the $p \times p$ identity matrix. If the eigenvalues of a matrix M are $\mu_1, \ldots, \mu_p$, then the eigenvalues of $M + cI$ (where c is a scalar) are $\mu_1 + c, \ldots, \mu_p + c$ [why?]. The proof follows from Lemma 1.4. $\square$

1.6 Corollary. *The number of closed walks of length ℓ in K_p from some vertex v_i to itself is given by*

$$(\boldsymbol{A}(K_p)^\ell)_{ii} = \frac{1}{p}((p-1)^\ell + (p-1)(-1)^\ell). \qquad (1)$$

(Note that this is also the number of sequences $(i_1, \ldots, i_\ell)$ of numbers $1, 2, \ldots, p$ such that $i_1 = i$, no two consecutive terms are equal, and $i_\ell \neq i_1$ [why?].)

Proof. By Corollary 1.3 and Proposition 1.5, the total number of closed walks in K_p of length ℓ is equal to $(p-1)^\ell + (p-1)(-1)^\ell$. By the symmetry of the graph K_p, the number of closed walks of length ℓ from v_i to itself does not depend on i. (All vertices "look the same.") Hence we can divide the total number of closed walks by p (the number of vertices) to get the desired answer. $\square$

What about non-closed walks in K_p? It's not hard to diagonalize explicitly the matrix $\boldsymbol{A}(K_p)$ (or equivalently, to compute its eigenvectors), but there is an even simpler special argument. We have

$$(J-I)^\ell = \sum_{k=0}^{\ell} (-1)^{\ell-k} \binom{\ell}{k} J^k, \tag{2}$$

by the binomial theorem. Now for $k>0$ we have $J^k = p^{k-1}J$ [why?], while $J^0 = I$. (It is not clear *a priori* what is the "correct" value of J^0, but in order for equation (2) to be valid we must take $J^0 = I$.) Hence

$$(J-I)^\ell = \sum_{k=1}^{\ell} (-1)^{\ell-k} \binom{\ell}{k} p^{k-1} J + (-1)^\ell I.$$

Again by the binomial theorem we have

$$\begin{aligned} (J-I)^\ell &= \frac{1}{p}((p-1)^\ell J - (-1)^\ell J) + (-1)^\ell I \\ &= \frac{1}{p}(p-1)^\ell J + \frac{(-1)^\ell}{p}(pI - J). \end{aligned} \tag{3}$$

Taking the (i,j)-entry of each side when $i \neq j$ yields

$$(\boldsymbol{A}(K_p)^\ell)_{ij} = \frac{1}{p}((p-1)^\ell - (-1)^\ell). \tag{4}$$

If we take the (i,i)-entry of (3) then we recover equation (1). Note the curious fact that if $i \neq j$ then

$$(\boldsymbol{A}(K_p)^\ell)_{ii} - (\boldsymbol{A}(K_p)^\ell)_{ij} = (-1)^\ell.$$

We could also have deduced (4) from Corollary 1.6 using

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \left(\boldsymbol{A}(K_p)^\ell\right)_{ij} = p(p-1)^\ell,$$

the total number of walks of length ℓ in K_p. Details are left to the reader.

We now will show how equation (1) itself determines the eigenvalues of $\boldsymbol{A}(K_p)$. Thus if (1) is proved without first computing the eigenvalues of $\boldsymbol{A}(K_p)$ (which in fact is what we did two paragraphs ago), then we have another means to compute the eigenvalues. The argument we will give can be applied to any graph G, not just K_p. We begin with a simple lemma.

1.7 Lemma. *Suppose $\alpha_1, \ldots, \alpha_r$ and $\beta_1, \ldots, \beta_s$ are* nonzero *complex numbers such that for all positive integers ℓ, we have*

$$\alpha_1^\ell + \cdots + \alpha_r^\ell = \beta_1^\ell + \cdots + \beta_s^\ell. \tag{5}$$

Then $r = s$ and the α's are just a permutation of the β's.

Proof. We will use the powerful method of *generating functions.* Let x be a complex number whose absolute value is close to 0. Multiply (5) by x^ℓ and sum on all $\ell \geq 1$. The geometric series we obtain will converge, and we get

$$\frac{\alpha_1 x}{1-\alpha_1 x} + \cdots + \frac{\alpha_r x}{1-\alpha_r x} = \frac{\beta_1 x}{1-\beta_1 x} + \cdots + \frac{\beta_s x}{1-\beta_s x}. \tag{6}$$

This is an identity valid for sufficiently small (in modulus) complex numbers. By clearing denominators we obtain a polynomial identity. But if two polynomials in x agree for infinitely many values, then they are the same polynomial [why?]. Hence equation (6) is actually valid for *all* complex numbers x (ignoring values of x which give rise to a zero denominator).

Fix a complex number $\gamma \neq 0$. Multiply (6) by $1 - \gamma x$ and let $x \to 1/\gamma$. The left-hand side becomes the number of α_i's which are equal to γ, while the right-hand side becomes the number of β_j's which are equal to γ [why?]. Hence these numbers agree for all γ, so the lemma is proved. □

1.8 Example. Suppose that G is a graph with 12 vertices, and that the number of closed walks of length ℓ in G is equal to $3 \cdot 5^\ell + 4^\ell + 2(-2)^\ell + 4$. Then it follows from Corollary 1.3 and Lemma 1.7 [why?] that the eigenvalues of $\boldsymbol{A}(G)$ are given by $5, 5, 5, 4, -2, -2, 1, 1, 1, 1, 0, 0$.

2 Cubes and the Radon transform.

Let us now consider a more interesting example of a graph G, one whose eigenvalues have come up in a variety of applications. Let $\mathbb{Z}_2$ denote the cyclic group of order 2, with elements 0 and 1, and group operation being addition modulo 2. Thus $0+0=0$, $0+1=1+0=1$, $1+1=0$. Let $\mathbb{Z}_2^n$ denote the direct product of $\mathbb{Z}_2$ with itself n times, so the elements of $\mathbb{Z}_2^n$ are n-tuples $(a_1,\dots,a_n)$ of 0's and 1's, under the operation of component-wise addition. Define a graph C_n, called the *n-cube*, as follows: The vertex set of C_n is given by $V(C_n)=\mathbb{Z}_2^n$, and two vertices u and v are connected by an edge if they differ in exactly one component. Equivalently, $u+v$ has exactly one nonzero component. If we regard $\mathbb{Z}_2^n$ as consisting of *real* vectors, then these vectors form the set of vertices of an n-dimensional cube. Moreover, two vertices of the cube lie on an edge (in the usual geometric sense) if and only if they form an edge of C_n. This explains why C_n is called the n-cube. We also see that walks in C_n have a nice geometric interpretation — they are simply walks along the edges of an n-dimensional cube.

We want to determine explicitly the eigenvalues and eigenvectors of C_n. We will do this by a somewhat indirect but extremely useful and powerful technique, the finite Radon transform. Let $\mathcal{V}$ denote the set of all functions $f:\mathbb{Z}_2^n\to\mathbb{R}$, where $\mathbb{R}$ denotes the field of real numbers. (Note: For groups other than $\mathbb{Z}_2^n$ it is necessary to use complex numbers rather than real numbers. We could use complex numbers here, but there is no need to do so.) Note that $\mathcal{V}$ is a vector space over $\mathbb{R}$ of dimension 2^n [why?]. If $u=(u_1,\dots,u_n)$ and $v=(v_1,\dots,v_n)$ are elements of $\mathbb{Z}_2^n$, then define their *dot product* by

$$u\cdot v=u_1v_1+\cdots+u_nv_n,$$

where the computation is performed modulo 2. Thus we regard $u\cdot v$ as an element of $\mathbb{Z}_2$. The expression $(-1)^{u\cdot v}$ is defined to be the *real number* $+1$ or -1, depending on whether $u\cdot v=0$ or 1, respectively. Since for integers k the value of $(-1)^k$ depends only on k (mod 2), it follows that we can treat u and v as integer vectors without affecting the value of $(-1)^{u\cdot v}$. Thus, for instance, formulas such as

$$(-1)^{u\cdot(v+w)}=(-1)^{u\cdot v+u\cdot w}=(-1)^{u\cdot v}(-1)^{u\cdot w}$$

are well-defined and valid.

We now define two important bases of the vector space $\mathcal{V}$. There will be one basis element of each basis for each $u \in \mathbb{Z}_2^n$. The first basis, denoted B_1, has elements f_u defined as follows:

$$f_u(v) = \delta_{uv}, \tag{7}$$

the Kronecker delta. It is easy to see that B_1 is a basis, since any $g \in \mathcal{V}$ satisfies

$$g = \sum_{u \in \mathbb{Z}_2^n} g(u) f_u \tag{8}$$

[why?]. Hence B_1 spans $\mathcal{V}$, so since $|B_1| = \dim \mathcal{V} = 2^n$, it follows that B_1 is a basis. The second basis, denoted B_2, has elements χ_u defined as follows:

$$\chi_u(v) = (-1)^{u \cdot v}.$$

In order to show that B_2 is a basis, we will use an inner product on $\mathcal{V}$ (denoted $\langle \cdot, \cdot \rangle$) defined by

$$\langle f, g \rangle = \sum_{u \in \mathbb{Z}_2^n} f(u) g(u).$$

Note that this inner product is just the usual dot product with respect to the basis B_1.

2.1 Lemma. *The set $B_2 = \{\chi_u : u \in \mathbb{Z}_2^n\}$ forms a basis for $\mathcal{V}$.*

Proof. Since $|B_2| = \dim \mathcal{V}$ $(= 2^n)$, it suffices to show that B_2 is linearly independent. In fact, we will show that the elements of B_2 are orthogonal. We have

$$\begin{aligned} \langle \chi_u, \chi_v \rangle &= \sum_{w \in \mathbb{Z}_2^n} \chi_u(w) \chi_v(w) \\ &= \sum_{w \in \mathbb{Z}_2^n} (-1)^{(u+v) \cdot w}. \end{aligned}$$

It is left as an easy exercise to the reader to show that for any $y \in \mathbb{Z}_2^n$, we have

$$\sum_{w \in \mathbb{Z}_2^n} (-1)^{y \cdot w} = \begin{cases} 2^n, & \text{if } y = \mathbf{0} \\ 0, & \text{otherwise.} \end{cases}$$

where $\mathbf{0}$ denotes the identity element of $\mathbb{Z}_2^n$ (the vector $(0,0,\ldots,0)$). Thus $\langle \chi_u, \chi_v \rangle = 0$ if and only $u + v = \mathbf{0}$, i.e., $u = v$, so the elements of B_2 are orthogonal (and nonzero). Hence they are linearly independent as desired. □

We now come to the key definition of the Radon transform.

2.2 Definition. Given a subset Γ of $\mathbb{Z}_2^n$ and a function $f \in \mathcal{V}$, define a new function $\Phi_\Gamma f \in \mathcal{V}$ by

$$\Phi_\Gamma f(v) = \sum_{w \in \Gamma} f(v + w).$$

The function $\Phi_\Gamma f$ is called the (*discrete* or *finite*) *Radon transform* of f (on the group $\mathbb{Z}_2^n$, with respect to the subset Γ).

We have defined a map $\Phi_\Gamma : \mathcal{V} \to \mathcal{V}$. It is easy to see that Φ_Γ is a linear transformation; we want to compute its eigenvalues and eigenvectors.

2.3 Theorem. *The eigenvectors of Φ_Γ are the functions χ_u, where $u \in \mathbb{Z}_2^n$. The eigenvalue λ_u corresponding to χ_u (i.e., $\Phi_\Gamma \chi_u = \lambda_u \chi_u$) is given by*

$$\lambda_u = \sum_{w \in \Gamma} (-1)^{u \cdot w}.$$

Proof. Let $v \in \mathbb{Z}_2^n$. Then

$$\begin{aligned}
\Phi_\Gamma \chi_u(v) &= \sum_{w \in \Gamma} \chi_u(v + w) \\
&= \sum_{w \in \Gamma} (-1)^{u \cdot (v + w)} \\
&= \left(\sum_{w \in \Gamma} (-1)^{u \cdot w} \right) (-1)^{u \cdot v} \\
&= \left(\sum_{w \in \Gamma} (-1)^{u \cdot w} \right) \chi_u(v).
\end{aligned}$$

Hence

$$\Phi_\Gamma \chi_u = \left(\sum_{w\in\Gamma} (-1)^{u\cdot w} \right) \chi_u,$$

as desired. □

Note that because the χ_u's form a basis for $\mathcal{V}$ by Lemma 2.1, it follows that Theorem 2.3 yields a complete set of eigenvalues and eigenvectors for Φ_Γ. Note also that the eigenvectors χ_u of Φ_Γ are independent of Γ; only the eigenvalues depend on Γ.

Now we come to the payoff. Let $\Delta = \{\delta_1, \dots, \delta_n\}$, where δ_i is the ith unit coordinate vector (i.e., δ_i has a 1 in position i and 0's elsewhere). Note that the jth coordinate of δ_i is just δ_{ij} (the Kronecker delta), explaining our notation δ_i. Let $[\Phi_\Delta]$ denote the matrix of the linear transformation $\Phi_\Delta : \mathcal{V} \to \mathcal{V}$ with respect to the basis B_1 of $\mathcal{V}$ given by (7).

2.4 Lemma. *We have* $[\Phi_\Delta] = \boldsymbol{A}(C_n)$*, the adjacency matrix of the* n*-cube.*

Proof. Let $v \in \mathbb{Z}_2^n$. We have

$$\begin{aligned} \Phi_\Delta f_u(v) &= \sum_{w\in\Delta} f_u(v+w) \\ &= \sum_{w\in\Delta} f_{u+w}(v), \end{aligned}$$

since $u = v + w$ if and only if $u + w = v$. There follows [why?]

$$\Phi_\Delta f_u = \sum_{w\in\Delta} f_{u+w}. \qquad (9)$$

Equation (9) says that the (u, v)-entry of the matrix Φ_Δ is given by

$$(\Phi_\Delta)_{uv} = \begin{cases} 1, & \text{if } u+v \in \Delta \\ 0, & \text{otherwise.} \end{cases}$$

Now $u + v \in \Delta$ if and only if u and v differ in exactly one coordinate. This is just the condition for uv to be an edge of C_n, so the proof follows. □

2.5 Corollary. *The eigenvectors E_u ($u \in \mathbb{Z}_2^n$) of $\boldsymbol{A}(C_n)$ (regarded as linear combinations of the vertices of C_n, i.e., of the elements of $\mathbb{Z}_2^n$) are given by*

$$E_u = \sum_{v \in \mathbb{Z}_2^n} (-1)^{u \cdot v} v. \tag{10}$$

The eigenvalue λ_u corresponding to the eigenvector E_u is given by

$$\lambda_u = n - 2\omega(u), \tag{11}$$

where $\omega(u)$ is the number of 1*'s in u. ($\omega(u)$ is called the* Hamming weight *or simply the* weight *of u.) Hence $\boldsymbol{A}(C_n)$ has $\binom{n}{i}$ eigenvalues equal to $n-2i$, for each $0 \le i \le n$.*

Proof. For any function $g \in \mathcal{V}$ we have by (8) that

$$g = \sum_v g(v) f_v.$$

Applying this equation to $g = \chi_u$ gives

$$\chi_u = \sum_v \chi_u(v) f_v = \sum_v (-1)^{u \cdot v} f_v. \tag{12}$$

Equation (12) expresses the eigenvector χ_u of Φ_Δ (or even Φ_Γ for any $\Gamma \subseteq \mathbb{Z}_2^n$) as a linear combination of the functions f_v. But Φ_Δ has the same matrix with respect to the basis of the f_v's as $\boldsymbol{A}(C_n)$ has with respect to the vertices v of C_n. Hence the expansion of the eigenvectors of Φ_Δ in terms of the f_v's has the same coefficients as the expansion of the eigenvectors of $\boldsymbol{A}(C_n)$ in terms of the v's, so equation (10) follows.

According to Theorem 2.3 the eigenvalue λ_u corresponding to the eigenvector χ_u of Φ_Δ (or equivalently, the eigenvector E_u of $\boldsymbol{A}(C_n)$) is given by

$$\lambda_u = \sum_{w \in \Delta} (-1)^{u \cdot w}. \tag{13}$$

Now $\Delta = \{\delta_1, \ldots, \delta_n\}$, and $\delta_i \cdot u$ is 1 if u has a one in its ith coordinate and is 0 otherwise. Hence the sum in (13) has $n - \omega(u)$ terms equal to $+1$ and $\omega(u)$ terms equal to -1, so $\lambda_u = (n - \omega(u)) - \omega(u) = n - 2\omega(u)$, as claimed. □

We have all the information needed to count walks in C_n.

2.6 Corollary. *Let $u, v \in \mathbb{Z}_2^n$, and suppose that $\omega(u+v) = k$ (i.e., u and v disagree in exactly k coordinates). Then the number of walks of length ℓ in C_n between u and v is given by*

$$(\boldsymbol{A}^\ell)_{uv} = \frac{1}{2^n} \sum_{i=0}^{n} \sum_{j=0}^{k} (-1)^j \binom{k}{j} \binom{n-k}{i-j} (n-2i)^\ell, \tag{14}$$

where we set $\binom{n-k}{i-j} = 0$ if $j > i$. In particular,

$$(\boldsymbol{A}^\ell)_{uu} = \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} (n-2i)^\ell. \tag{15}$$

Proof. Let E_u and λ_u be as in Corollary 2.5. In order to apply Corollary 1.2, we need the eigenvectors to be of *unit* length (where we regard the f_v's as an orthonormal basis of $\mathcal{V}$). By equation (10), we have

$$|E_u|^2 = \sum_{v \in \mathbb{Z}_2^n} ((-1)^{u \cdot v})^2 = 2^n.$$

Hence we should replace E_u by $E'_u = \frac{1}{2^{n/2}} E_u$ to get an orthonormal basis. According to Corollary 1.2, we thus have

$$(\boldsymbol{A}^\ell)_{uv} = \frac{1}{2^n} \sum_{w \in \mathbb{Z}_2^n} E_{uw} E_{vw} \lambda_w^\ell.$$

Now E_{uw} by definition is the coefficient of f_w in the expansion (10), i.e., $E_{uw} = (-1)^{u+w}$ (and similarly for E_v), while $\lambda_w = n - 2\omega(w)$. Hence

$$(\boldsymbol{A}^\ell)_{uv} = \frac{1}{2^n} \sum_{w \in \mathbb{Z}_2^n} (-1)^{(u+v) \cdot w} (n - 2\omega(w))^\ell. \tag{16}$$

The number of vectors w of Hamming weight i which have j 1's in common with $u+v$ is $\binom{k}{j}\binom{n-k}{i-j}$, since we can choose the j 1's in $u+v$ which agree with w in $\binom{k}{j}$ ways, while the remaining $i-j$ 1's of w can be inserted in the

$n-k$ remaining positions in $\binom{n-k}{i-j}$ ways. Since $(u+v)\cdot w \equiv j \pmod 2$, the sum (16) reduces to (14) as desired. Clearly setting $u=v$ in (14) yields (15), completing the proof. □

It is possible to give a direct proof of (15) avoiding linear algebra. Thus by Corollary 1.3 and Lemma 1.7 (exactly as was done for K_n) we have another determination of the eigenvalues of C_n. With a little more work one can also obtain a direct proof of (14). Later in Example 9.9.12, however, we will use the eigenvalues of C_n to obtain a combinatorial result for which no nonalgebraic proof is known.

2.7 Example. Setting $k=1$ in (14) yields

$$\begin{aligned}(\boldsymbol{A}^{\ell})_{uv} &= \frac{1}{2^n}\sum_{i=0}^{n}\left[\binom{n-1}{i}-\binom{n-1}{i-1}\right](n-2i)^{\ell}\\ &= \frac{1}{2^n}\sum_{i=0}^{n-1}\binom{n-1}{i}\frac{(n-2i)^{\ell+1}}{n-i}.\ \square\end{aligned}$$

3 Random walks.

Let G be a finite graph. We consider a random walk on the vertices of G of the following type. Start at a vertex u. (The vertex u could be chosen randomly according to some probability distribution or could be specified in advance.) Among all the edges incident to u, choose one uniformly at random (i.e., if there are k edges incident to u, then each of these edges is chosen with probability $1/k$). Travel to the vertex v at the other end of the chosen edge and continue as before from v. Readers with some familiarity with probability theory will recognize this random walk as a special case of a finite state Markov chain. Many interesting questions may be asked about such walks; the basic one is to determine the probability of being at a given vertex after a given number ℓ of steps.

Suppose vertex u has *degree* d_u, i.e., there are d_u edges incident to u (counting loops at u once only). Let $\boldsymbol{M} = \boldsymbol{M}(G)$ be the matrix whose rows and columns are indexed by the vertex set $\{v_1, \dots, v_p\}$ of G, and whose (u, v)-entry is given by

$$M_{uv} = \frac{\mu_{uv}}{d_u},$$

where μ_{uv} is the number of edges between u and v (which for simple graphs will be 0 or 1). Thus M_{uv} is just the probability that if one starts at u, then the next step will be to v. An elementary probability theory argument (equivalent to Theorem 1.1) shows that if ℓ is a positive integer, then $(\boldsymbol{M}^\ell)_{uv}$ is equal to probability that one ends up at vertex v in ℓ steps given that one has started at u. Suppose now that the starting vertex is not specified, but rather we are given probabilities ρ_u summing to 1 and that we start at vertex u with probability ρ_u. Let P be the row vector $P = [\rho_{v_1}, \dots, \rho_{v_p}]$. Then again an elementary argument shows that if $P\boldsymbol{M}^\ell = [\sigma_{v_1}, \dots, \sigma_{v_p}]$, then σ_v is the probability of ending up at v in ℓ steps (with the given starting distribution). By reasoning as in Section 1, we see that if we know the eigenvalues and eigenvectors of $\boldsymbol{M}$, then we can compute the crucial probabilities $(\boldsymbol{M}^\ell)_{uv}$ and σ_u.

Since the matrix $\boldsymbol{M}$ is not the same as the adjacency matrix $\boldsymbol{A}$, what does all this have to do with adjacency matrices? The answer is that in one important case $\boldsymbol{M}$ is just a scalar multiple of $\boldsymbol{A}$. We say that the graph G

is *regular of degree* d if each $d_u = d$, i.e., each vertex is incident to d edges. In this case it's easy to see that $\boldsymbol{M}(G) = \frac{1}{d}\boldsymbol{A}(G)$. Hence the eigenvectors E_u of $\boldsymbol{M}(G)$ and $\boldsymbol{A}(G)$ are the same, and the eigenvalues are related by $\lambda_u(\boldsymbol{M}) = \frac{1}{d}\lambda_u(\boldsymbol{A})$. Thus random walks on a regular graph are closely related to the adjacency matrix of the graph.

3.1 Example. Consider a random walk on the n-cube C_n which begins at the "origin" (the vector $(0,\ldots,0)$). What is the probability p_ℓ that after ℓ steps one is again at the origin? Before applying any formulas, note that after an even (respectively, odd) number of steps, one must be at a vertex with an even (respectively, odd) number of 1's. Hence $p_\ell = 0$ if ℓ is odd. Now note that C_n is regular of degree n. Thus by (11), we have

$$\lambda_u(\boldsymbol{M}(C_n)) = \frac{1}{n}(n - 2\omega(u)).$$

By (15) we conclude that

$$p_\ell = \frac{1}{2^n n^\ell}\sum_{i=0}^{n}\binom{n}{i}(n-2i)^\ell.$$

Note that the above expression for p_ℓ does indeed reduce to 0 when ℓ is odd.

4 The Sperner property.

In this section we consider a surprising application of certain adjacency matrices to some problems in extremal set theory. An important role will also be played by finite groups. In general, extremal set theory is concerned with finding (or estimating) the most or least number of sets satisfying given set-theoretic or combinatorial conditions. For example, a typical easy problem in extremal set theory is the following: What is the most number of subsets of an n-element set with the property that any two of them intersect? (Can you solve this problem?) The problems to be considered here are most conveniently formulated in terms of partially ordered sets, or posets for short. Thus we begin with discussing some basic notions concerning posets.

4.1 Definition. A *poset* (short for partially ordered set) P is a finite set, also denoted P, together with a binary relation denoted $\leq$ satisfying the following axioms:

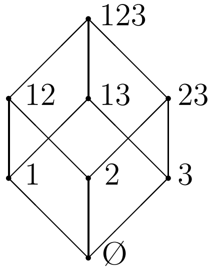
(P1) (reflexivity) $x \leq x$ for all $x \in P$

(P2) (antisymmetry) If $x \leq y$ and $y \leq x$, then $x = y$.

(P3) (transitivity) If $x \leq y$ and $y \leq z$, then $x \leq z$.

One easy way to obtain a poset is the following. Let P be any collection of sets. If $x, y \in P$, then define $x \leq y$ in P if $x \subseteq y$ as sets. It is easy to see that this definition of $\leq$ makes P into a poset. If P consists of *all* subsets of an n-element set S, then P is called a (finite) *boolean algebra* of *rank* n and is denoted by B_S. If $S = \{1, 2, \ldots, n\}$, then we denote B_S simply by B_n. Boolean algebras will play an important role throughout this section.

There is a simple way to represent small posets pictorially. The *Hasse diagram* of a poset P is a planar drawing, with elements of P drawn as dots. If $x < y$ in P (i.e., $x \leq y$ and $x \neq y$), then y is drawn "above" x (i.e., with a larger vertical coordinate). An edge is drawn between x and y if y *covers* x, i.e., $x < y$ and no element z is in between, i.e., no z satisfies $x < z < y$. By the transitivity property (P3), all the relations of a finite poset are determined by the cover relations, so the Hasse diagram determines P. (This is not true for infinite posets; for instance, the real numbers $\mathbb{R}$ with their usual order is a poset with no cover relations.) The Hasse diagram of the boolean algebra B_3 looks like



We say that two posets P and Q are *isomorphic* if there is a bijection (one-to-one and onto function) $\varphi : P \to Q$ such that $x \leq y$ in P if and only if $\varphi(x) \leq \varphi(y)$ in Q. Thus one can think that two posets are isomorphic if they differ only in the names of their elements. This is exactly analogous to the notion of isomorphism of groups, rings, etc. It is an instructive exercise to draw Hasse diagrams of the one poset of order (number of elements) one (up to isomorphism), the two posets of order two, the five posets of order three, and the sixteen posets of order four. More ambitious readers can try the 63 posets of order five, the 318 of order six, the 2045 of order seven, the 16999 of order eight, the 183231 of order nine, the 2567284 of order ten, the 46749427 of order eleven, the 1104891746 of order twelve, the 33823827452 of order thirteen, the 1338193159771 of order fourteen, the 68275077901156 of order fifteen, and the 4483130665195087 of order sixteen. Beyond this the number is not currently known.

A *chain* C in a poset is a totally ordered subset of P, i.e., if $x, y \in C$ then either $x \leq y$ or $y \leq x$ in P. A finite chain is said to have *length* n if it has $n+1$ elements. Such a chain thus has the form $x_0 < x_1 < \cdots < x_n$. We say that a finite poset is *graded of rank* n if every maximal chain has length n. (A chain is *maximal* if it's contained in no larger chain.) For instance, the boolean algebra B_n is graded of rank n [why?]. A chain $y_0 < y_1 < \cdots < y_j$ is said to be *saturated* if each y_{i+1} covers y_i. Such a chain need not be maximal since there can be elements of P smaller than y_0 or greater than y_j. If P is graded of rank n and $x \in P$, then we say that x has *rank* j, denoted $\rho(x) = j$, if some (or equivalently, every) saturated chain of P with top element x has length j. Thus [why?] if we let $P_j = \{x \in P : \rho(x) = j\}$, then P is a *disjoint* union $P = P_0 \cup P_1 \cup \cdots \cup P_n$, and every maximal chain of P has the form $x_0 < x_1 < \cdots < x_n$ where $\rho(x_j) = j$. We write $p_j = |P_j|$, the number of elements of P of rank j. For example, if $P = B_n$ then $\rho(x) = |x|$ (the cardinality of x as a set) and

$$p_j = \#\{x \subseteq \{1, 2, \cdots, n\} : |x| = j\} = \binom{n}{j}.$$

(Note that we use both $|S|$ and $\#S$ for the cardinality of the finite set S.)

We say that a graded poset P of rank n (always assumed to be finite) is *rank-symmetric* if $p_i = p_{n-i}$ for $0 \le i \le n$, and *rank-unimodal* if $p_0 \le p_1 \le \cdots \le p_j \ge p_{j+1} \ge p_{j+2} \ge \cdots \ge p_n$ for some $0 \le j \le n$. If P is both rank-symmetric and rank-unimodal, then we clearly have

$$p_0 \le p_1 \le \cdots \le p_m \ge p_{m+1} \ge \cdots \ge p_n, \text{ if } n = 2m$$

$$p_0 \le p_1 \le \cdots \le p_m = p_{m+1} \ge p_{m+2} \ge \cdots \ge p_n, \text{ if } n = 2m+1.$$

We also say that the sequence $p_0, p_1, \ldots, p_n$ itself or the polynomial $F(q) = p_0 + p_1 q + \cdots + p_n q^n$ is *symmetric* or *unimodal*, as the case may be. For instance, B_n is rank-symmetric and rank-unimodal, since it is well-known (and easy to prove) that the sequence $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ (the nth row of Pascal's triangle) is symmetric and unimodal. Thus the polynomial $(1+q)^n$ is symmetric and unimodal.

A few more definitions, and then finally some results! An *antichain* in a poset P is a subset A of P for which no two elements are comparable, i.e., we can never have $x, y \in A$ and $x < y$. For instance, in a graded poset P the "levels" P_j are antichains [why?]. We will be concerned with the problem of finding the largest antichain in a poset. Consider for instance the boolean algebra B_n. The problem of finding the largest antichain in B_n is clearly equivalent to the following problem in extremal set theory: Find the largest collection of subsets of an n-element set such that no element of the collection contains another. A good guess would be to take all the subsets of cardinality $\lfloor n/2 \rfloor$ (where $\lfloor x \rfloor$ denotes the greatest integer $\le x$), giving a total of $\binom{n}{\lfloor n/2 \rfloor}$ sets in all. But how can we actually prove there is no larger collection? Such a proof was first given by Emmanuel Sperner in 1927 and is known as *Sperner's* *theorem.* We will give three proofs of Sperner's theorem in this section: one proof uses linear algebra and will be applied to certain other situations; the second proof is an elegant combinatorial argument due to David Lubell in 1966; while the third proof is another combinatorial argument closely related to the linear algebra proof. We present the last two proofs for their "cultural value." Our extension of Sperner's theorem to certain other situations will involve the following crucial definition.

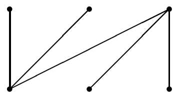
4.2 Definition. Let P be a graded poset of rank n. We say that P has the *Sperner property* or is a *Sperner poset* if

$$\max\{|A| : A \text{ is an antichain of } P\} = \max\{|P_i| : 0 \leq i \leq n\}.$$

In other words, no antichain is larger than the largest level P_i.

Thus Sperner's theorem is equivalent to saying that B_n has the Sperner property. Note that if P has the Sperner property there may still be antichains of maximum cardinality other than the biggest P_i; there just can't be any bigger antichains.

4.3 Example. A simple example of a graded poset that fails to satisfy the Sperner property is the following:



We now will discuss a simple combinatorial condition which guarantees that certain graded posets P are Sperner. We define an *order-matching* from P_i to P_{i+1} to be a *one-to-one* function $\mu : P_i \to P_{i+1}$ satisfying $x < \mu(x)$ for all $x \in P_i$. Clearly if such an order-matching exists then $p_i \leq p_{i+1}$ (since μ is one-to-one). Easy examples show that the converse is false, i.e., if $p_i \leq p_{i+1}$ then there need not exist an order-matching from P_i to P_{i+1}. We similarly define an order-matching from P_i to P_{i-1} to be a one-to-one function $\mu : P_i \to P_{i-1}$ satisfying $\mu(x) < x$ for all $x \in P_i$.

4.4 Proposition. *Let P be a graded poset of rank n. Suppose there exists an integer $0 \leq j \leq n$ and order-matchings*

$$P_0 \to P_1 \to P_2 \to \cdots \to P_j \leftarrow P_{j+1} \leftarrow P_{j+2} \leftarrow \cdots \leftarrow P_n. \tag{17}$$

Then P is rank-unimodal and Sperner.

Proof. Since order-matchings are one-to-one it is clear that

$$p_0 \leq p_1 \leq \cdots \leq p_j \geq p_{j+1} \geq p_{j+2} \geq \cdots \geq p_n.$$

Hence P is rank-unimodal.

Define a graph G as follows. The vertices of G are the elements of P. Two vertices x, y are connected by an edge if one of the order-matchings μ in the statement of the proposition satisfies $\mu(x) = y$. (Thus G is a subgraph of the Hasse diagram of P.) Drawing a picture will convince you that G consists of a disjoint union of paths, including single-vertex paths not involved in any of the order-matchings. The vertices of each of these paths form a chain in P. Thus we have partitioned the elements of P into disjoint chains. Since P is rank-unimodal with biggest level P_j, all of these chains must pass through P_j [why?]. Thus the number of chains is exactly p_j. Any antichain A can intersect each of these chains at most once, so the cardinality $|A|$ of A cannot exceed the number of chains, i.e., $|A| \leq p_j$. Hence by definition P is Sperner. □

It is now finally time to bring some linear algebra into the picture. For any (finite) set S, we let $\mathbb{R}S$ denote the real vector space consisting of all formal linear combinations (with real coefficients) of elements of S. Thus S is a basis for $\mathbb{R}S$, and in fact we could have simply defined $\mathbb{R}S$ to be the real vector space with basis S. The next lemma relates the combinatorics we have just discussed to linear algebra and will allow us to prove that certain posets are Sperner by the use of linear algebra (combined with some finite group theory).

4.5 Lemma. *Suppose there exists a linear transformation $U : \mathbb{R}P_i \rightarrow \mathbb{R}P_{i+1}$ (U stands for "up") satisfying:*

- *U is one-to-one.*
- *For all $x \in P_i$, $U(x)$ is a linear combination of elements $y \in P_{i+1}$ satisfying $x < y$. (We then call U an* order-raising operator.*)*

Then there exists an order-matching $\mu : P_i \to P_{i+1}$.

Similarly, suppose there exists a linear transformation $U : \mathbb{R}P_i \to \mathbb{R}P_{i+1}$ satisfying:

- *U is onto.*
- *U is an order-raising operator.*

Then there exists an order-matching $\mu : P_{i+1} \to P_i$.

Proof. Suppose $U : \mathbb{R}P_i \to \mathbb{R}P_{i+1}$ is a one-to-one order-raising operator. Let $[U]$ denote the matrix of U with respect to the bases P_i of $\mathbb{R}P_i$ and P_{i+1} of $\mathbb{R}P_{i+1}$. Thus the rows of $[U]$ are indexed by the elements $x_1, \ldots, x_{p_i}$ of P_i (in some order) and the columns by the elements $y_1, \ldots, y_{p_{i+1}}$ of P_{i+1}. Since U is one-to-one, the rank of $[U]$ is equal to p_i (the number of rows). Since the row rank of a matrix equals its column rank, $[U]$ must have p_i linearly independent columns. Say we have labelled the elements of P_{i+1} so that the first p_i columns of $[U]$ are linearly independent.

Let $A = (a_{ij})$ be the $p_i \times p_i$ matrix whose columns are the first p_i columns of $[U]$. (Thus A is a square submatrix of $[U]$.) Since the columns of A are linearly independent, we have

$$\det(A) = \sum \pm a_{1\pi(1)} \cdots a_{p_i \pi(p_i)} \neq 0,$$

where the sum is over all permutations π of $1, \ldots, p_i$. Thus some term $\pm a_{1\pi(1)} \cdots a_{p_i \pi(p_i)}$ of the above sum in nonzero. Since U is order-raising, this means that [why?] $x_k < y_{\pi(k)}$ for $1 \leq k \leq p_i$. Hence the map $\mu : P_i \to P_{i+1}$ defined by $\mu(x_k) = y_{\pi(k)}$ is an order-matching, as desired.

The case when U is onto rather than one-to-one is proved by a completely analogous argument. □

We now want to apply Proposition 4.4 and Lemma 4.5 to the boolean algebra B_n. For each $0 \leq i < n$, we need to define a linear transformation $U_i : \mathbb{R}(B_n)_i \to \mathbb{R}(B_n)_{i+1}$, and then prove it has the desired properties. We simply define U_i to be the simplest possible order-raising operator, namely, for $x \in (B_n)_i$, let

$$U_i(x) = \sum_{\substack{y \in (B_n)_{i+1} \\ y > x}} y. \tag{18}$$

Note that since $(B_n)_i$ is a basis for $\mathbb{R}(B_n)_i$, equation (18) does indeed define a unique linear transformation $U_i : \mathbb{R}(B_n)_i \to \mathbb{R}(B_n)_{i+1}$. By definition U_i is order-raising; we want to show that U_i is one-to-one for $i < n/2$ and onto for $i \geq n/2$. There are several ways to show this using only elementary linear algebra; we will give what is perhaps the simplest proof, though it is quite tricky. The idea is to introduce "dual" operators $D_i : \mathbb{R}(B_n)_i \to \mathbb{R}(B_n)_{i-1}$ to the U_i's (D stands for "down"), defined by

$$D_i(y) = \sum_{\substack{x \in (B_n)_{i-1} \\ x < y}} x, \tag{19}$$

for all $y \in (B_n)_i$. Let $[U_i]$ denote the matrix of U_i with respect to the bases $(B_n)_i$ and $(B_n)_{i+1}$, and similarly let $[D_i]$ denote the matrix of D_i with respect to the bases $(B_n)_i$ and $(B_n)_{i-1}$. A key observation which we will use later is that

$$[D_{i+1}] = [U_i]^t, \tag{20}$$

i.e., the matrix $[D_{i+1}]$ is the transpose of the matrix $[U_i]$ [why?]. Now let $I_i : \mathbb{R}(B_n)_i \to \mathbb{R}(B_n)_i$ denote the identity transformation on $\mathbb{R}(B_n)_i$, i.e., $I_i(u) = u$ for all $u \in \mathbb{R}(B_n)_i$. The next lemma states (in linear algebraic terms) the fundamental combinatorial property of B_n which we need. For this lemma set $U_n = 0$ and $D_0 = 0$ (the 0 linear transformation between the appropriate vector spaces).

4.6 Lemma. *Let $0 \leq i \leq n$. Then*

$$D_{i+1}U_i - U_{i-1}D_i = (n - 2i)I_i. \tag{21}$$

(Linear transformations are multiplied right-to-left, so $AB(u) = A(B(u))$.)

Proof. Let $x \in (B_n)_i$. We need to show that if we apply the left-hand side of (21) to x, then we obtain $(n - 2i)x$. We have

$$D_{i+1}U_i(x) \;=\; D_{i+1}\left(\sum_{\substack{|y|=i+1 \\ x \subset y}} y \right)$$

$$= \sum_{\substack{|y|=i+1 \\ x\subset y}} \sum_{\substack{|z|=i \\ z\subset y}} z.$$

If $x, z \in (B_n)_i$ satisfy $|x \cap z| < i-1$, then there is no $y \in (B_n)_{i+1}$ such that $x \subset y$ and $z \subset y$. Hence the coefficient of z in $D_{i+1}U_i(x)$ when it is expanded in terms of the basis $(B_n)_i$ is 0. If $|x \cap z| = i-1$, then there is one such y, namely, $y = x \cup z$. Finally if $x = z$ then y can be any element of $(B_n)_{i+1}$ containing x, and there are $n-i$ such y in all. It follows that

$$D_{i+1}U_i(x) = (n-i)x + \sum_{\substack{|z|=i \\ |x\cap z|=i-1}} z. \qquad (22)$$

By exactly analogous reasoning (which the reader should check), we have for $x \in (B_n)_i$ that

$$U_{i-1}D_i(x) = ix + \sum_{\substack{|z|=i \\ |x\cap z|=i-1}} z. \qquad (23)$$

Subtracting (23) from (22) yields $(D_{i+1}U_i - U_{i-1}D_i)(x) = (n-2i)x$, as desired. □

4.7 Theorem. *The operator U_i defined above is one-to-one if $i < n/2$ and is onto if $i \geq n/2$.*

Proof. Recall that $[D_i] = [U_{i-1}]^t$. From linear algebra we know that a (rectangular) matrix times its transpose is *positive semidefinite* (or just *semidefinite* for short) and hence has nonnegative (real) eigenvalues. By Lemma 4.6 we have

$$D_{i+1}U_i = U_{i-1}D_i + (n-2i)I_i.$$

Thus the eigenvalues of $D_{i+1}U_i$ are obtained from the eigenvalues of $U_{i-1}D_i$ by adding $n-2i$. Since we are assuming that $n-2i > 0$, it follows that the eigenvalues of $D_{i+1}U_i$ are strictly positive. Hence $D_{i+1}U_i$ is invertible (since it has no 0 eigenvalues). But this implies that U_i is one-to-one [why?], as desired.

The case $i \geq n/2$ is done by a "dual" argument (or in fact can be deduced directly from the $i < n/2$ case by using the fact that the poset B_n is "self-dual," though we will not go into this). Namely, from the fact that

$$U_iD_{i+1} = D_{i+2}U_{i+1} + (2i+2-n)I_{i+1}$$

we get that U_iD_{i+1} is invertible, so now U_i is onto, completing the proof. $\square$

Combining Proposition 4.4, Lemma 4.5, and Theorem 4.7, we obtain the famous theorem of Sperner.

4.8 Corollary. *The boolean algebra B_n has the Sperner property.*

It is natural to ask whether there is a less indirect proof of Corollary 4.8. In fact, several nice proofs are known; we give one due to David Lubell, mentioned before Definition 4.2.

Lubell's proof of Sperner's theorem. First we count the total number of maximal chains $\emptyset = x_0 < x_1 < \cdots < x_n = \{1, \ldots, n\}$ in B_n. There are n choices for x_1, then $n-1$ choices for x_2, etc., so there are $n!$ maximal chains in all. Next we count the number of maximal chains $x_0 < x_1 < \cdots < x_i = x < \cdots < x_n$ which contain a given element x of rank i. There are i choices for x_1, then $i-1$ choices for x_2, up to one choice for x_i. Similarly there are $n-i$ choices for x_{i+1}, then $n-i+1$ choices for x_{i+2}, etc., up to one choice for x_n. Hence the number of maximal chains containing x is $i!(n-i)!$.

Now let A be an antichain. If $x \in A$, then let C_x be the set of maximal chains of B_n which contain x. Since A is an antichain, the sets C_x, $x \in A$ are pairwise disjoint. Hence

$$\begin{aligned} |\bigcup_{x \in A} C_x| &= \sum_{x \in A} |C_x| \\ &= \sum_{x \in A} (\rho(x))!(n-\rho(x))! \end{aligned}$$

Since the total number of maximal chains in the C_x's cannot exceed the total number $n!$ of maximal chains in B_n, we have

$$\sum_{x \in A} (\rho(x))!(n-\rho(x))! \leq n!$$

Divide both sides by $n!$ to obtain

$$\sum_{x \in A} \frac{1}{\binom{n}{\rho(x)}} \leq 1.$$

Since $\binom{n}{i}$ is maximized when $i = \lfloor n/2 \rfloor$, we have

$$\frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \leq \frac{1}{\binom{n}{\rho(x)}},$$

for all $x \in A$ (or all $x \in B_n$). Thus

$$\sum_{x \in A} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \leq 1,$$

or equivalently,

$$|A| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Since $\binom{n}{\lfloor n/2 \rfloor}$ is the size of the largest level of B_n, it follows that B_n is Sperner. □

There is another nice way to show directly that B_n is Sperner, namely, by constructing an explicit order-matching $\mu : (B_n)_i \to (B_n)_{i+1}$ when $i < n/2$. We will define μ by giving an example. Let $n = 21$, $i = 9$, and $S = \{3, 4, 5, 8, 12, 13, 17, 19, 20\}$. We want to define $\mu(S)$. Let $(a_1, a_2, \ldots, a_{21})$ be a sequence of ± 1's, where $a_i = 1$ if $i \in S$, and $a_i = -1$ if $i \notin S$. For the set S above we get the sequence (writing $-$ for -1)

$$- - 1\,1\,1 - -1 - - - 1\,1 - - - 1 - 1\,1 - .$$

Replace any two consecutive terms $1-$ with $0\,0$:

$$- - 1\,1\,0\,0 - 0\,0 - -1\,0\,0 - -0\,0\,1\,0\,0.$$

Ignore the 0's and replace any two consecutive terms $1-$ with $0\,0$:

$$- - 1\,0\,0\,0\,0\,0\,0 - -0\,0\,0\,0 - 0\,0\,1\,0\,0.$$

Continue:

$$- - 0\,0\,0\,0\,0\,0\,0\,0 - 0\,0\,0\,0 - 0\,0\,1\,0\,0.$$

At this stage no further replacement is possible. The nonzero terms consist of a sequence of $-$'s followed by a sequence of 1's. There is at least one $-$ since $i < n/2$. Let k be the position (coordinate) of the last $-$; here $k = 16$. Define $\mu(S) = S \cup \{k\} = S \cup \{16\}$. The reader can check that this procedure gives an order-matching. In particular, why is μ injective (one-to-one), i.e., why can we recover S from $\mu(S)$?

In view of the above elegant proof of Lubell and the explicit description of an order-matching $\mu : (B_n)_i \to (B_n)_{i+1}$, the reader may be wondering what was the point of giving a rather complicated and indirect proof using linear algebra. Admittedly, if all we could obtain from the linear algebra machinery we have developed was just another proof of Sperner's theorem, then it would have been hardly worth the effort. But in the next section we will show how Theorem 4.7, when combined with a little finite group theory, can be used to obtain many interesting combinatorial results for which simple, direct proofs are not known.

5 Group actions on boolean algebras.

Let us begin by reviewing some facts from group theory. Suppose that X is an n-element set and that G is a group. We say that G *acts on* the set X if for every element π of G we associate a permutation (also denoted π) of X, such that for all $x \in X$ and $\pi, \sigma \in G$ we have

$$\pi(\sigma(x)) = (\pi\sigma)(x).$$

Thus [why?] an action of G on X is the same as a homomorphism $\varphi : G \to \mathfrak{S}_X$, where $\mathfrak{S}_X$ denotes the symmetric group of all permutations of X. We sometimes write $\pi \cdot x$ instead of $\pi(x)$.

5.1 Example. (a) Let the real number α act on the xy-plane by rotation counterclockwise around the origin by an angle of α radians. It is easy to check that this defines an action of the group $\mathbb{R}$ of real numbers (under addition) on the xy-plane.

(b) Now let $\alpha \in \mathbb{R}$ act by translation by a distance α to the right (i.e., adding $(\alpha, 0)$). This yields a completely different action of $\mathbb{R}$ on the xy-plane.

(c) Let $X = \{a, b, c, d\}$ and $G = \mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0,0), (0,1), (1,0), (1,1)\}$. Let G act as follows:

$$(0,1) \cdot a = b, \quad (0,1) \cdot b = a, \quad (0,1) \cdot c = c, \quad (0,1) \cdot d = d$$

$$(1,0) \cdot a = a, \quad (1,0) \cdot b = b, \quad (1,0) \cdot c = d, \quad (1,0) \cdot d = c.$$

The reader should check that this does indeed define an action. In particular, since $(1,0)$ and $(0,1)$ generate G, we don't need to define the action of $(0,0)$ and $(1,1)$ — they are uniquely determined.

(d) Let X and G be as in (c), but now define the action by

$$(0,1) \cdot a = b, \quad (0,1) \cdot b = a, \quad (0,1) \cdot c = d, \quad (0,1) \cdot d = c$$

$$(1,0) \cdot a = c, \quad (1,0) \cdot b = d, \quad (1,0) \cdot c = a, \quad (1,0) \cdot d = b.$$

Again one can check that we have an action of $\mathbb{Z}_2 \times \mathbb{Z}_2$ on $\{a, b, c, d\}$.

Recall what is meant by an *orbit* of the action of a group G on a set X. Namely, we say that two elements x, y of X are *G-equivalent* if $\pi(x) = y$ for some $\pi \in G$. The relation of G-equivalence is an equivalence relation, and the equivalence classes are called orbits. Thus x and y are in the same orbit if $\pi(x) = y$ for some $\pi \in G$. The orbits form a *partition* of X, i.e, they are pairwise-disjoint, nonempty subsets of X whose union is X. The orbit containing x is denoted Gx; this is sensible notation since Gx consists of all elements $\pi(x)$ where $\pi \in G$. Thus $Gx = Gy$ if and only if x and y are G-equivalent (i.e., in the same G-orbit). The set of all G-orbits is denoted X/G.

5.2 Example. (a) In Example 5.1(a), the orbits are circles with center $(0, 0)$ (including the degenerate circle whose only point is $(0, 0)$).

(b) In Example 5.1(b), the orbits are horizontal lines. Note that although in (a) and (b) the same group G acts on the same set X, the orbits are different.

(c) In Example 5.1(c), the orbits are $\{a, b\}$ and $\{c, d\}$.

(d) In Example 5.1(d), there is only one orbit $\{a, b, c, d\}$. Again we have a situation in which a group G acts on a set X in two different ways, with different orbits.

We wish to consider the situation where $X = B_n$, the boolean algebra of rank n (so $|B_n| = 2^n$). We begin by defining an *automorphism* of a poset P to be an isomorphism $\varphi : P \rightarrow P$. (This definition is exactly analogous to the definition of an automorphism of a group, ring, etc.) The set of all automorphisms of P forms a group, denoted $\mathrm{Aut}(P)$ and called the *automorphism group* of P, under the operation of composition of functions (just as is the case for groups, rings, etc.)

Now consider the case $P = B_n$. Any permutation π of $\{1, \ldots, n\}$ acts on B_n as follows: If $x = \{i_1, i_2, \ldots, i_k\} \in B_n$, then

$$\pi(x) = \{\pi(i_1), \pi(i_2), \ldots, \pi(i_k)\}. \tag{24}$$

This action of π on B_n is an automorphism [why?]; in particular, if $|x| = i$, then also $|\pi(x)| = i$. Equation (24) defines an action of the symmetric group

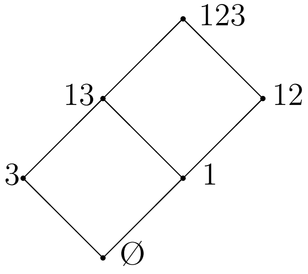
$\mathfrak{S}_n$ of all permutations of $\{1, \ldots, n\}$ on B_n [why?]. (In fact, it is not hard to show that *every* automorphism of B_n is of the form (24) for $\pi \in \mathfrak{S}_n$.) In particular, any subgroup G of $\mathfrak{S}_n$ acts on B_n *via* (24) (where we restrict π to belong to G). In what follows this action is always meant.

5.3 Example. Let $n = 3$, and let G be the subgroup of $\mathfrak{S}_3$ with elements e and $(1, 2)$. Here e denotes the identity permutation, and (using disjoint cycle notation) $(1, 2)$ denotes the permutation which interchanges 1 and 2, and fixes 3. There are six orbits of G (acting on B_3). Writing e.g. 13 as short for $\{1, 3\}$, the six orbits are $\{\emptyset\}$, $\{1, 2\}$, $\{3\}$, $\{12\}$, $\{13, 23\}$, and $\{123\}$.

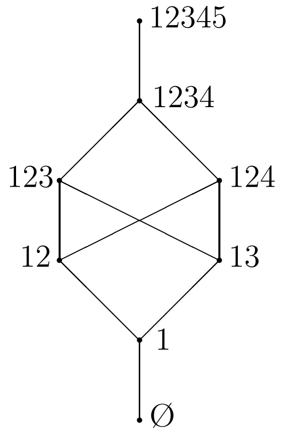
We now define the class of posets which will be of interest to us here. Later we will give some special cases of particular interest.

5.4 Definition. Let G be a subgroup of $\mathfrak{S}_n$. Define the *quotient poset* B_n/G as follows: The elements of B_n/G are the orbits of G. If $\mathcal{O}$ and $\mathcal{O}'$ are two orbits, then define $\mathcal{O} \leq \mathcal{O}'$ in B_n/G if there exist $x \in \mathcal{O}$ and $y \in \mathcal{O}'$ such that $x \leq y$ in B_n. (It's easy to check that this relation $\leq$ is indeed a partial order.)

5.5 Example. (a) Let $n = 3$ and G be the group of order two generated by the cycle $(1, 2)$, as in Example 5.3. Then the Hasse diagram of B_3/G is shown below, where each element (orbit) is labeled by one of its elements.



(b) Let $n = 5$ and G be the group of order five generated by the cycle $(1, 2, 3, 4, 5)$. Then B_5/G has Hasse diagram



One simple property of a quotient poset B_n/G is the following.

5.6 Proposition. *The quotient poset B_n/G defined above is graded of rank n and rank-symmetric.*

Proof. We leave as an exercise the easy proof that B_n/G is graded of rank n, and that the rank of an element $\mathcal{O}$ of B_n/G is just the rank in B_n of any of the elements x of $\mathcal{O}$. Thus the number of elements $p_i(B_n/G)$ of rank i is equal to the number of orbits $\mathcal{O} \in (B_n)_i/G$. If $x \in B_n$, then let $\bar{x}$ denote the set-theoretic complement of x, i.e.,

$$\bar{x} = \{1, \ldots, n\} - x = \{1 \leq i \leq n : i \notin x\}.$$

Then $\{x_1, \ldots, x_j\}$ is an orbit of i-element subsets of $\{1, \ldots, n\}$ if and only if $\{\bar{x}_1, \ldots, \bar{x}_j\}$ is an orbit of $(n-i)$-element subsets [why?]. Hence $|(B_n)_i/G| = |(B_n)_{n-i}/G|$, so B_n/G is rank-symmetric. □

Let $\pi \in \mathfrak{S}_n$. We associate with π a linear transformation (still denoted π)
$\pi : \mathbb{R}(B_n)_i \to \mathbb{R}(B_n)_i$ by the rule

$$\pi\left(\sum_{x \in (B_n)_i} c_x x\right) = \sum_{x \in (B_n)_i} c_x \pi(x),$$

where each c_x is a real number. (This defines an action of $\mathfrak{S}_n$, or of any subgroup G of $\mathfrak{S}_n$, on the vector space $\mathbb{R}(B_n)_i$.) The matrix of π with respect to the basis $(B_n)_i$ is just a *permutation matrix*, i.e., a matrix with one 1 in every row and column, and 0's elsewhere. We will be interested in elements of $\mathbb{R}(B_n)_i$ which are fixed by every element of a subgroup G of $\mathfrak{S}_n$. The set of all such elements is denoted $\mathbb{R}(B_n)_i^G$, so

$$\mathbb{R}(B_n)_i^G = \{v \in \mathbb{R}(B_n)_i : \pi(v) = v \text{ for all } \pi \in G\}.$$

5.7 Lemma. *A basis for $\mathbb{R}(B_n)_i^G$ consists of the elements*

$$v_{\mathcal{O}} := \sum_{x \in \mathcal{O}} x,$$

where $\mathcal{O} \in (B_n)_i/G$, the set of G-orbits for the action of G on $(B_n)_i$.

Proof. First note that if $\mathcal{O}$ is an orbit and $x \in \mathcal{O}$, then by definition of orbit we have $\pi(x) \in \mathcal{O}$ for all $\pi \in G$ (or all $\pi \in \mathfrak{S}_n$). Since π permutes the elements of $(B_n)_i$, it follows that π permutes the elements of $\mathcal{O}$. Thus $\pi(v_{\mathcal{O}}) = v_{\mathcal{O}}$, so $v_{\mathcal{O}} \in \mathbb{R}(B_n)_i^G$. It is clear that the $v_{\mathcal{O}}$'s are linearly independent since any $x \in (B_n)_i$ appears with nonzero coefficient in exactly one $v_{\mathcal{O}}$.

It remains to show that the $v_{\mathcal{O}}$'s span $\mathbb{R}(B_n)_i^G$, i.e., any $v = \sum_{x \in (B_n)_i} c_x x \in \mathbb{R}(B_n)_i^G$ can be written as a linear combination of $v_{\mathcal{O}}$'s. Given $x \in (B_n)_i$, let $G_x = \{\pi \in G : \pi(x) = x\}$, the *stabilizer* of x. We leave as an exercise the standard fact that $\pi(x) = \sigma(x)$ (where $\pi, \sigma \in G$) if and only if π and σ belong to the same left coset of G_x, i.e., $\pi G_x = \sigma G_x$. It follows that in the multiset of elements $\pi(x)$, where π ranges over all elements of G and x is fixed, every element y in the orbit Gx appears $\#G_x$ times, and no other elements appear. In other words,

$$\sum_{\pi \in G} \pi(x) = |G_x| \cdot v_{Gx}.$$

(Do not confuse the orbit Gx with the subgroup G_x!) Now apply π to v and sum on all $\pi \in G$. Since $\pi(v) = v$ (because $v \in \mathbb{R}(B_n)_i^G$), we get

$$\begin{aligned} |G| \cdot v &= \sum_{\pi \in G} \pi(v) \\ &= \sum_{\pi \in G} \left(\sum_{x \in (B_n)_i} c_x \pi(x) \right) \end{aligned}$$

$$\begin{aligned} &= \sum_{x\in(B_n)_i} c_x \left(\sum_{\pi\in G}\pi(x)\right) \\ &= \sum_{x\in(B_n)_i} c_x\cdot|G_x|\cdot v_{Gx}. \end{aligned}$$

Dividing by $|G|$ expresses v as a linear combination of the elements v_{Gx} (or $v_{\mathcal{O}}$), as desired. □

Now let us consider the effect of applying the order-raising operator U_i to an element v of $\mathbb{R}(B_n)_i^G$.

5.8 Lemma. *If $v\in\mathbb{R}(B_n)_i^G$, then $U_i(v)\in\mathbb{R}(B_n)_{i+1}^G$.*

Proof. Note that since $\pi\in G$ is an automorphism of B_n, we have $x<y$ in B_n if and only if $\pi(x)<\pi(y)$ in B_n. It follows [why?] that if $x\in(B_n)_i$ then

$$U_i(\pi(x))=\pi(U_i(x)).$$

Since U_i and π are linear transformations, it follows by linearity that $U_i\pi(u)=\pi U_i(u)$ for all $u\in\mathbb{R}(B_n)_i$. (In other words, $U_i\pi=\pi U_i$.) Then

$$\begin{aligned} \pi(U_i(v)) &= U_i(\pi(v)) \\ &= U_i(v), \end{aligned}$$

so $U_i(v)\in\mathbb{R}(B_n)_{i+1}^G$, as desired. □

We come to the main result of this section, and indeed our main result on the Sperner property.

5.9 Theorem. *Let G be a subgroup of $\mathfrak{S}_n$. Then the quotient poset B_n/G is graded of rank n, rank-symmetric, rank-unimodal, and Sperner.*

Proof. Let $P=B_n/G$. We have already seen in Proposition 5.6 that P is graded of rank n and rank-symmetric. We want to define order-raising operators $\hat{U}_i:\mathbb{R}P_i\to\mathbb{R}P_{i+1}$ and order-lowering operators $\hat{D}_i:\mathbb{R}P_i\to\mathbb{R}P_{i-1}$. Let us first consider just $\hat{U}_i$. The idea is to identify the basis element $v_{\mathcal{O}}$ of $\mathbb{R}B_n^G$ with the basis element $\mathcal{O}$ of $\mathbb{R}P$, and to let $\hat{U}_i:\mathbb{R}P_i\to\mathbb{R}P_{i+1}$ correspond to the usual order-raising operator $U_i:\mathbb{R}(B_n)_i\to\mathbb{R}(B_n)_{i+1}$. More precisely, suppose that the order-raising operator U_i for B_n given by (18) satisfies

$$U_i(v_{\mathcal{O}}) = \sum_{\mathcal{O}' \in (B_n)_{i+1}/G} c_{\mathcal{O},\mathcal{O}'} v_{\mathcal{O}'}, \tag{25}$$

where $\mathcal{O} \in (B_n)_i/G$. (Note that by Lemma 5.8, $U_i(v_{\mathcal{O}})$ does indeed have the form given by (25).) Then define the linear operator $\hat{U}_i : \mathbb{R}((B_n)_i/G) \rightarrow \mathbb{R}((B_n)_i/G)$ by

$$\hat{U}_i(\mathcal{O}) = \sum_{\mathcal{O}' \in (B_n)_{i+1}/G} c_{\mathcal{O},\mathcal{O}'} \mathcal{O}'.$$

We claim that $\hat{U}_i$ is order-raising. We need to show that if $c_{\mathcal{O},\mathcal{O}'} \neq 0$, then $\mathcal{O}' > \mathcal{O}$ in B_n/G. Since $v_{\mathcal{O}'} = \sum_{x' \in \mathcal{O}'} x'$, the only way $c_{\mathcal{O},\mathcal{O}'} \neq 0$ in (25) is for some $x' \in \mathcal{O}'$ to satisfy $x' > x$ for some $x \in \mathcal{O}$. But this is just what it means for $\mathcal{O}' > \mathcal{O}$, so $\hat{U}_i$ is order-raising.

Now comes the heart of the argument. We want to show that $\hat{U}_i$ is one-to-one for $i < n/2$. Now by Theorem 4.7, U_i is one-to-one for $i < n/2$. Thus the restriction of U_i to the subspace $\mathbb{R}(B_n)_i^G$ is one-to-one. (The restriction of a one-to-one function is always one-to-one.) But U_i and $\hat{U}_i$ are exactly the same transformation, except for the names of the basis elements on which they act. Thus $\hat{U}_i$ is also one-to-one for $i < n/2$.

An exactly analogous argument can be applied to D_i instead of U_i. We obtain one-to-one order-lowering operators $\hat{D}_i : \mathbb{R}(B_n)_i^G \rightarrow \mathbb{R}(B_n)_{i-1}^G$ for $i > n/2$. It follows from Proposition 4.4, Lemma 4.5, and (20) that B_n/G is rank-unimodal and Sperner, completing the proof. □

We will consider two interesting applications of Theorem 5.9. For our first application, we let $n = \binom{m}{2}$ for some $m \geq 1$, and let $M = \{1, \ldots, m\}$. Let $X = \binom{M}{2}$, the set of all two-element subsets of M. Think of the elements of X as (possible) edges of a graph with vertex set M. If B_X is the boolean algebra of all subsets of X (so B_X and B_n are isomorphic), then an element x of B_X is a collection of edges on the vertex set M, in other words, just a simple graph on M. Define a subgroup G of $\mathfrak{S}_X$ as follows: Informally, G consists of all permutations of the edges $\binom{M}{2}$ that are induced from permutations of the vertices M. More precisely, if $\pi \in \mathfrak{S}_m$, then define $\hat{\pi} \in \mathfrak{S}_X$ by $\hat{\pi}(\{i,j\}) = \{\pi(i), \pi(j)\}$. Thus G is isomorphic to $\mathfrak{S}_m$.

When are two graphs $x, y \in B_X$ in the same orbit of the action of G on B_X? Since the elements of G just permute vertices, we see that x and y are in the same orbit if we can obtain x from y by permuting vertices. This is just what it means for two simple graphs x and y to be *isomorphic* — they are the same graph except for the names of the vertices (thinking of edges as pairs of vertices). Thus the elements of B_X/G are *isomorphism classes* of simple graphs on the vertex set M. In particular, $\#(B_X/G)$ is the number of nonisomorphic m-vertex simple graphs, and $\#((B_X/G)_i)$ is the number of nonisomorphic such graphs with i edges. We have $x \leq y$ in B_X/G if there is some way of labelling the vertices of x and y so that every edge of x is an edge of y. Equivalently, some *spanning subgraph* of y (i.e., a subgraph of y with all the vertices of y) is isomorphic to x. Hence by Theorem 5.9 there follows the following result, which is by no means obvious and has no known non-algebraic proof.

5.10 Theorem. (a) *Fix $m \geq 1$. Let p_i be the number of nonisomorphic simple graphs with m vertices and i edges. Then the sequence $p_0, p_1, \ldots, p_{\binom{m}{2}}$ is symmetric and unimodal.*

(b) *Let T be a collection of simple graphs with m vertices such that no element of T is isomorphic to a spanning subgraph of another element of T. Then $|T|$ is maximized by taking T to consist of all nonisomorphic simple graphs with $\lfloor \frac{1}{2}\binom{m}{2} \rfloor$ edges.*

Our second example of the use of Theorem 5.9 is somewhat more subtle and will be the topic of the next section.

Digression: edge reconstruction. Much work has been done on "reconstruction problems," that is, trying to reconstruct a mathematical structure such as a graph from some of its substructures. The most famous of such problems is *vertex reconstruction*: given a simple graph G on n vertices $v_1, \ldots, v_p$, let G_i be the subgraph obtained by deleting vertex v_i (and all incident edges). Given the multiset $\{G_1, \ldots, G_p\}$ of vertex-deleted subgraphs graphs, can G be uniquely reconstructed? It is important to realize that the vertices are *unlabelled*, so given G_i we don't know for any j which vertex is v_j. The famous *vertex reconstruction conjecture* (still open) states that for $n \geq 3$ any graph G can be reconstructed from the multiset $\{G_1, \ldots, G_p\}$.

Here we will be concerned with *edge* reconstruction, another famous open problem. Given a simple graph G with edges $e_1, \ldots, e_q$, let $H_i = G - e_i$, the graph obtained from G by removing the edge e_i.

Edge Reconstruction Conjecture. A simple graph G can be uniquely reconstructed from its number of vertices and the multiset $\{H_1, \ldots, H_q\}$ of edge-deleted subgraphs.

NOTE. As in the case of vertex-reconstruction, the subgraphs H_i are unlabelled. The reason for including the number of vertices is that for a graph with no edges, we have $\{H_1, \ldots, H_q\} = \emptyset$, so we need to specify the number of vertices to obtain G.

NOTE. It can be shown that if G can be vertex-reconstructed, then G can be edge reconstructed. Hence the vertex-reconstruction conjecture implies the edge-reconstruction conjecture.

The techniques developed above to analyze group actions on boolean algebra can be used to prove a special case of the edge-reconstruction conjecture. Note that a simple graph with p vertices has at most $\binom{p}{2}$ edges.

5.11 Theorem. Let G be a simple graph with p vertices and $q > \frac{1}{2}\binom{p}{2}$ edges. Then G is edge-reconstructible.

Proof. Let P_i be the set of all simple graphs with i edges on the vertex set $[p]$, so $\#P_i = \binom{\binom{p}{2}}{i}$. Let $\mathbb{R}P_i$ denote the real vector space with basis P_i. Define a linear transformation $\psi_i : \mathbb{R}P_i \to \mathbb{R}P_{i-1}$ by

$$\psi_i(\Gamma) = \Gamma_1 + \cdots + \Gamma_i,$$

where $\Gamma_1, \ldots, \Gamma_i$ are the (labelled) graphs obtained from Γ by deleting a single edge. By Theorem 4.7, ψ_i is injective for $i > \frac{1}{2}\binom{p}{2}$. (Think of ψ_i as adding edges to the *complement* of Γ, i.e., the graph with vertex set $[p]$ and edge set $\binom{[p]}{2} - E(\Gamma)$.)

The symmetric group $\mathfrak{S}_p$ acts on P_q by permuting the vertices, and hence acts on $\mathbb{R}P_q$, the real vector space with basis P_q. A basis for the fixed space $(\mathbb{R}P_q)^{\mathfrak{S}_p}$ consists of the distinct sums $\tilde{\Gamma} = \sum_{\pi \in \mathfrak{S}_p} \pi(\Gamma)$, where $\Gamma \in \ P_q$. We may identify $\tilde{\Gamma}$ with the *unlabelled* graph isomorphic to Γ, since $\tilde{\Gamma} = \tilde{\Gamma}'$ if and only if Γ and Γ' are isomorphic. Just as in the proof of Theorem 5.9, when we restrict ψ_q to $(\mathbb{R}P_q)^{\mathfrak{S}_p}$ for $q > \frac{1}{2}\binom{p}{2}$ we obtain an injection $\psi_q : (\mathbb{R}P_q)^{\mathfrak{S}_p} \rightarrow (\mathbb{R}P_{q-1})^{\mathfrak{S}_p}$. In particular, for nonisomorphic unlabelled graphs $\tilde{\Gamma}, \tilde{\Gamma}'$ with p vertices, we have

$$\tilde{\Gamma}_1 + \cdots + \tilde{\Gamma}_q = \psi_q(\tilde{\Gamma}) \neq \psi_q(\tilde{\Gamma}') = \tilde{\Gamma}'_1 + \cdots + \tilde{\Gamma}'_q.$$

Hence the unlabelled graphs $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_q$ determine $\tilde{\Gamma}$, as desired. □

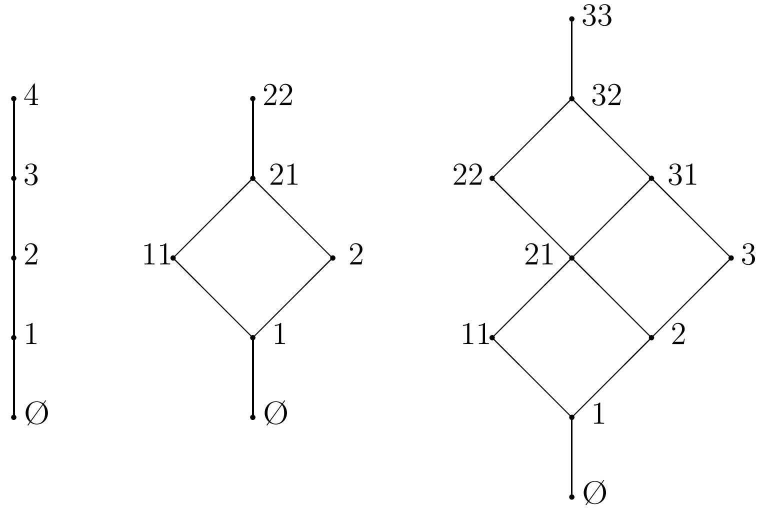
Theorem 5.11 was first proved by L. Lovász using the Principle of Inclusion-Exclusion. The proof given above is due to R. Stanley. W. Müller found an improvement of Lovász's argument, showing that a graph with p vertices and $q > p \log p$ edges is edge-reconstructible. I. Krasikov and Y. Roditty later found an improvement of our proof of Theorem 5.11 that gave another proof of Müller's result.

6 Young diagrams and q-binomial coefficients.

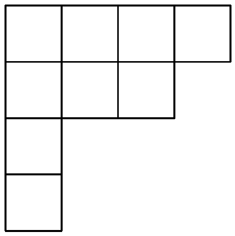
A *partition* λ of an integer $n \geq 0$ is a sequence $\lambda = (\lambda_1, \lambda_2, \ldots)$ of integers $\lambda_i \geq 0$ satisfying $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\sum_{i \geq 1} \lambda_i = n$. Thus all but finitely many λ_i are equal to 0. Each $\lambda_i > 0$ is called a *part* of λ. We sometimes suppress 0's from the notation for λ, e.g., $(5,2,2,1)$, $(5,2,2,1,0,0,0)$, and $(5,2,2,1,0,0,\ldots)$ all represent the same partition λ (of 10, with four parts). If λ is a partition of n, then we denote this by $\lambda \vdash n$ or $|\lambda| = n$.

6.1 Example. There are seven partitions of 5, namely (writing e.g. 221 as short for $(2,2,1)$): 5, 41, 32, 311, 221, 2111, and 11111.

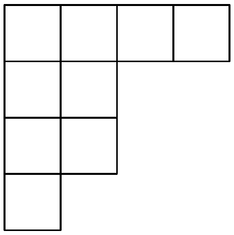
The subject of partitions of integers has been extensively developed, and we will only be concerned here with a small part related to our previous discussion. Given positive integers m and n, let $L(m,n)$ denote the set of all partitions with at most m parts and with largest part at most n. For instance, $L(2,3) = \{\emptyset, 1, 2, 3, 11, 21, 31, 22, 32, 33\}$. (Note that we are denoting by $\emptyset$ the unique partition $(0,0,\ldots)$ with no parts.) If $\lambda = (\lambda_1, \lambda_2, \ldots)$ and $\mu = (\mu_1, \mu_2, \ldots)$ are partitions, then define $\lambda \leq \mu$ if $\lambda_i \leq \mu_i$ for all i. This makes the set of all partitions into a very interesting poset, denoted Y and called *Young's lattice* (named after the British mathematician Alfred Young, 1873–1940). (It is called "Young's lattice" rather than "Young's poset" because it turns out to have certain properties which define a *lattice*. However, these properties are irrelevant to us here, so we will not bother to define the notion of a lattice.) We will be looking at some properties of Y in Section 8. The partial ordering on Y, when restricted to $L(m,n)$, makes $L(m,n)$ into a poset which also has some fascinating properties. The diagrams below show $L(1,4)$, $L(2,2)$, and $L(2,3)$.



There is a nice geometric way of viewing partitions and the poset $L(m, n)$. The *Young diagram* (sometimes just called the *diagram*) of a partition λ is a left-justified array of squares, with λ_i squares in the ith row. For instance, the Young diagram of $(4, 3, 1, 1)$ looks like:

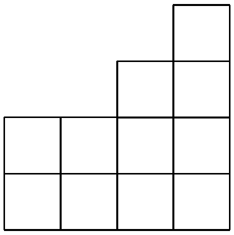


If dots are used instead of boxes, then the resulting diagram is called a *Ferrers diagram.* The advantage of Young diagrams over Ferrers diagrams is that we can put numbers in the boxes of a Young diagram, which we will do in Section 7. Observe that $L(m, n)$ is simply the set of Young diagrams D fitting in an $m \times n$ rectangle (where the upper-left (northwest) corner of D is the same as the northwest corner of the rectangle), ordered by inclusion. *We will always assume that when a Young diagram D is contained in a rectangle R, the northwest corners agree.* It is also clear from the Young diagram point of view that $L(m, n)$ and $L(n, m)$ are isomorphic partially ordered sets, the isomorphism being given by transposing the diagram (i.e., interchanging rows and columns). If λ has Young diagram D, then the partition whose diagram is D^t (the transpose of D) is called the *conjugate* of λ and is denoted λ'. For instance, $(4,3,1,1)' = (4,2,2,1)$, with diagram



6.2 Proposition. *$L(m,n)$ is graded of rank mn and rank-symmetric. The rank of a partition λ is just $|\lambda|$ (the sum of the parts of λ or the number of squares in its Young diagram).*

Proof. As in the proof of Proposition 5.6, we leave to the reader everything except rank-symmetry. To show rank-symmetry, consider the complement $\bar{\lambda}$ of λ in an $m \times n$ rectangle R, i.e., all the squares of R except for λ. (Note that $\bar{\lambda}$ depends on m and n, and not just λ.) For instance, in $L(4,5)$, the complement of $(4,3,1,1)$ looks like



If we rotate the diagram of $\bar{\lambda}$ by 180° then we obtain the diagram of a partition $\tilde{\lambda} \in L(m,n)$ satisfying $|\lambda| + |\tilde{\lambda}| = mn$. This correspondence between λ and $\tilde{\lambda}$ shows that $L(m,n)$ is rank-symmetric. □

Our main goal in this section is to show that $L(m,n)$ is rank-unimodal and Sperner. Let us write $p_i(m,n)$ as short for $p_i(L(m,n))$, the number of elements of $L(m,n)$ of rank i. Equivalently, $p_i(m,n)$ is the number of partitions of i with largest part at most n and with at most m parts, or, in other words, the number of distinct Young diagrams with i squares which fit inside an $m \times n$ rectangle (with the same northwest corner, as explained previously). Though not really necessary for this goal, it is nonetheless interesting to obtain some information on these numbers $p_i(m,n)$. First let us consider the total number $|L(m,n)|$ of elements in $L(m,n)$.

6.3 Proposition. *We have* $|L(m,n)| = \binom{m+n}{m}$.

Proof. We will give an elegant combinatorial proof, based on the fact that $\binom{m+n}{m}$ is equal to the number of sequences $a_1, a_2, \ldots, a_{m+n}$, where each a_j is either N or E, and there are m N's (and hence n E's) in all. We will associate a Young diagram D contained in an $m \times n$ rectangle R with such a sequence as follows. Begin at the lower left-hand corner of R, and trace out the southeast boundary of D, ending at the upper right-hand corner of R. This is done by taking a sequence of unit steps (where each square of R is one unit in length), each step either north or east. Record the sequence of steps, using N for a step to the north and E for a step to the east.

Example. Let $m = 5$, $n = 6$, $\lambda = (4,3,1,1)$. Then R and D are given by:

×	×	×	×		
×	×	×			
×					
×					

The corresponding sequence of N's and E's is $NENNEEENENEE$.

It is easy to see (left to the reader) that the above correspondence gives a bijection between Young diagrams D fitting in an $m \times n$ rectangle R, and sequences of m N's and n E's. Hence the number of diagrams is equal to $\binom{m+n}{m}$, the number of sequences. □

We now consider how many elements of $L(m,n)$ have rank i. To this end, let q be an indeterminate; and given $j \geq 1$ define $[j] = 1 + q + q^2 + \cdots + q^{j-1}$. Thus $[1] = 1$, $[2] = 1 + q$, $[3] = 1 + q + q^2$, etc. Note that $[j]$ is a polynomial in q whose value at $q = 1$ is just j (denoted $[j]_{q=1} = j$). Next define $[j]! = [1][2]\cdots[j]$ for $j \geq 1$, and set $[0]! = 1$. Thus $[1]! = 1$, $[2]! = 1+q$, $[3]! = (1+q)(1+q+q^2) = 1 + 2q + 2q^2 + q^3$, etc., and $[j]!_{q=1} = j!$. Finally define for $k \geq j \geq 0$,

$$\begin{bmatrix} k \\ j \end{bmatrix} = \frac{[k]!}{[j]![k-j]!}.$$

The expression $\begin{bmatrix} k \\ j \end{bmatrix}$ is called a *q-binomial coefficient* (or *Gaussian coefficient*). Since $[r]!_{q=1} = r!$, it is clear that

$$\begin{bmatrix} k \\ j \end{bmatrix}_{q=1} = \binom{k}{j}.$$

One sometimes says that $\begin{bmatrix} k \\ j \end{bmatrix}$ is a "q-analogue" of the binomial coefficient $\binom{k}{j}$.

6.4 Example. We have $\begin{bmatrix} k \\ j \end{bmatrix} = \begin{bmatrix} k \\ k-j \end{bmatrix}$ [why?]. Moreover,

$$\begin{bmatrix} k \\ 0 \end{bmatrix} = \begin{bmatrix} k \\ k \end{bmatrix} = 1$$

$$\begin{bmatrix} k \\ 1 \end{bmatrix} = \begin{bmatrix} k \\ k-1 \end{bmatrix} = [k] = 1 + q + q^2 + \cdots + q^{k-1}$$

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \frac{[4][3][2][1]}{[2][1][2][1]} = 1 + q + 2q^2 + q^3 + q^4$$

$$\begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} = 1 + q + 2q^2 + 2q^3 + 2q^4 + q^5 + q^6.$$

In the above example, $\begin{bmatrix} k \\ j \end{bmatrix}$ was always a polynomial in q (and with nonnegative integer coefficients). It is not obvious that this is always the case, but it will follow easily from the following lemma.

6.5 Lemma. *We have*

$$\begin{bmatrix} k \\ j \end{bmatrix} = \begin{bmatrix} k-1 \\ j \end{bmatrix} + q^{k-j} \begin{bmatrix} k-1 \\ j-1 \end{bmatrix}, \tag{26}$$

whenever $k \geq 1$, with the "initial conditions" $\left[{0 \atop 0}\right] = 1$, $\left[{k \atop j}\right] = 0$ if $j < 0$ or $j > k$ (the same intial conditions satisfied by the binomial coefficients $\binom{k}{j}$).

Proof. This is a straightforward computation. Specifically, we have

$$\begin{aligned}
\begin{bmatrix} k-1 \\ j \end{bmatrix} + q^{k-j} \begin{bmatrix} k-1 \\ j-1 \end{bmatrix} &= \frac{[k-1]!}{[j]![k-1-j]!} + q^{k-j} \frac{[k-1]!}{[j-1]![k-j]!} \\
&= \frac{[k-1]!}{[j-1]![k-1-j]!} \left(\frac{1}{[j]} + \frac{q^{k-j}}{[k-j]} \right) \\
&= \frac{[k-1]!}{[j-1]![k-1-j]!} \frac{[k-j] + q^{k-j}[j]}{[j][k-j]} \\
&= \frac{[k-1]!}{[j-1]![k-1-j]!} \frac{[k]}{[j][k-j]} \\
&= \begin{bmatrix} k \\ j \end{bmatrix}. \; \square
\end{aligned}$$

Note that if we put $q = 1$ in (26) we obtain the well-known formula

$$\binom{k}{j} = \binom{k-1}{j} + \binom{k-1}{j-1},$$

which is just the recurrence defining Pascal's triangle. Thus equation (26) may be regarded as a q-analogue of the Pascal triangle recurrence.

We can regard equation (26) as a recurrence relation for the q-binomial coefficients. Given the initial conditions of Lemma 6.5, we can use (26) inductively to compute $\left[{k \atop j}\right]$ for any k and j. From this it is obvious by induction that the q-binomial coefficient $\left[{k \atop j}\right]$ is a polynomial in q with nonnegative integer coefficients. The following theorem gives an even stronger result, namely, an explicit combinatorial interpretation of the coefficients.

6.6 Theorem. *Let $p_i(m, n)$ denote the number of elements of $L(m, n)$ of rank i. Then*

$$\sum_{i \geq 0} p_i(m, n) q^i = \begin{bmatrix} m+n \\ m \end{bmatrix}. \tag{27}$$

(NOTE. The sum on the left-hand side is really a *finite* sum, since $p_i(m,n) = 0$ if $i > mn$.)

Proof. Let $P(m,n)$ denote the left-hand side of (27). We will show that

$$P(0,0) = 1, \text{ and } P(m,n) = 0 \text{ if } m < 0 \text{ or } n < 0 \tag{28}$$

$$P(m,n) = P(m,n-1) + q^n P(m-1,n). \tag{29}$$

Note that equations (28) and (29) completely determine $P(m,n)$. On the other hand, substituting $k = m+n$ and $j = m$ in (26) shows that $\left[{m+n \atop m} \right]$ also satisfies (29). Moreover, the initial conditions of Lemma 6.5 show that $\left[{m+n \atop m} \right]$ also satisfies (28). Hence (28) and (29) imply that $P(m,n) = \left[{m+n \atop m} \right]$, so to complete the proof we need only establish (28) and (29).

Equation (28) is clear, since $L(0,n)$ consists of a single point (the empty partition $\emptyset$), so $\sum_{i\geq 0} p_i(0,n)q^i = 1$; while $L(m,n)$ is empty (or undefined, if you prefer) if $m < 0$ or $n < 0$,

The crux of the proof is to show (29). Taking the coefficient of q^i of both sides of (29), we see [why?] that (29) is equivalent to

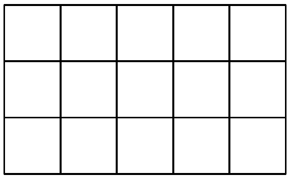
$$p_i(m,n) = p_i(m,n-1) + p_{i-n}(m-1,n). \tag{30}$$

Consider a partition $\lambda \vdash i$ whose Young diagram D fits in an $m \times n$ rectangle R. If D does not contain the upper right-hand corner of R, then D fits in an $m \times (n-1)$ rectangle, so there are $p_i(m,n-1)$ such partitions λ. If on the other hand D does contain the upper right-hand corner of R, then D contains the whole first row of R. When we remove the first row of R, we have left a Young diagram of size $i-n$ which fits in an $(m-1) \times n$ rectangle. Hence there are $p_{i-n}(m-1,n)$ such λ, and the proof follows [why?]. □

Note that if we set $q = 1$ in (27), then the left-hand side becomes $|L(m,n)|$ and the right-hand side $\binom{m+n}{m}$, agreeing with Proposition 6.3.

NOTE: There is another well-known interpretation of $\left[{k \atop j} \right]$, this time not of its coefficients (regarded as a polynomial in q), but rather at its *values* for certain q. Namely, suppose q is the power of a prime. Recall that there is a field $\mathbb{F}_q$ (unique up to isomorphism) with q elements. Then one can show that $\left[{k \atop j}\right]$ is equal to the number of j-dimensional subspaces of a k-dimensional vector space over the field $\mathbb{F}_q$. We will not discuss the proof here since it is not relevant for our purposes.

As the reader may have guessed by now, the poset $L(m,n)$ is isomorphic to a quotient poset B_s/G for a suitable integer $s > 0$ and finite group G acting on B_s. Actually, it is clear that we must have $s = mn$ since $L(m,n)$ has rank mn and in general B_s/G has rank s. What is not so clear is the right choice of G. To this end, let $R = R_{mn}$ denote an $m \times n$ rectangle of squares. For instance, R_{35} is given by the 15 squares of the diagram



We now define the group $G = G_{mn}$ as follows. It is a subgroup of the group $\mathfrak{S}_R$ of all permutations of the squares of R. A permutation π in G is allowed to permute the elements in each row of R in any way, and then to permute the rows themselves of R in any way. The elements of each row can be permuted in $n!$ ways, so since there are m rows there are a total of $n!^m$ permutations preserving the rows. Then the m rows can be permuted in $m!$ ways, so it follows that the order of G_{mn} is given by $m!n!^m$. (The group G_{mn} is called the *wreath product* of $\mathfrak{S}_n$ and $\mathfrak{S}_m$, denoted $\mathfrak{S}_n \wr \mathfrak{S}_m$ or $\mathfrak{S}_n$ wr $\mathfrak{S}_m$. However, we will not discuss the general theory of wreath products here.)

6.7 Example. Suppose $m = 4$ and $n = 5$, with the boxes of X labelled as follows.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

Then a typical permutation π in $G(4,5)$ looks like

16	20	17	19	18
4	1	5	2	3
12	13	15	14	11
7	9	6	10	8

,

i.e., $\pi(16) = 1$, $\pi(20) = 2$, etc.

We have just defined a group G_{mn} of permutations of the set R_{mn} of squares of an $m \times n$ rectangle. Hence G_{mn} acts on the boolean algebra B_R of all subsets of the set R. The next lemma describes the orbits of this action.

6.8 Lemma. *Every orbit $\mathcal{O}$ of the action of G_{mn} on B_R contains exactly one Young diagram D (i.e., exactly one subset $D \subseteq R$ such that D is left-justified, and if λ_i is the number of elements of D in row i of R, then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$).*

Proof. Let S be a subset of R, and suppose that S has α_i elements in row i. If $\pi \in G_{mn}$ and $\pi \cdot S$ has β_i elements in row i, then $\beta_1, \ldots, \beta_m$ is just some permutation of $\alpha_1, \ldots, \alpha_m$ [why?]. There is a unique permutation $\lambda_1, \ldots, \lambda_m$ of $\alpha_1, \ldots, \alpha_m$ satisfying $\lambda_1 \geq \cdots \geq \lambda_m$, so the only possible Young diagram D in the orbit $\pi \cdot S$ is the one of shape $\lambda = (\lambda_1, \ldots, \lambda_m)$. It's easy to see that the Young diagram D_λ of shape λ is indeed in the orbit $\pi \cdot S$. For by permuting the elements in the rows of R we can left-justify the rows of S, and then by permuting the rows of R themselves we can arrange the row sizes of S to be in weakly decreasing order. Thus we obtain the Young diagram D_λ as claimed. $\square$

We are now ready for the main result of this section.

6.9 Theorem. *The quotient poset $B_{R_{mn}}/G_{mn}$ is isomorphic to $L(m,n)$.*

Proof. Each element of B_R/G_{mn} contains a unique Young diagram D_λ by Lemma 6.8. Moreover, two different orbits cannot contain the same Young diagram D since orbits are disjoint. Thus the map $\varphi : B_R/G_{mn} \to L(m,n)$

defined by $\varphi(D_\lambda) = \lambda$ is a bijection (one-to-one and onto). We claim that in fact φ is an isomorphism of partially ordered sets. We need to show the following: Let $\mathcal{O}$ and $\mathcal{O}^*$ be orbits of G_{mn} (i.e., elements of B_R/G_{mn}). Let D_λ and D_{λ^*} be the unique Young diagrams in $\mathcal{O}$ and $\mathcal{O}^*$, respectively. Then there exist $D \in \mathcal{O}$ and $D^* \in \mathcal{O}^*$ satisfying $D \subseteq D^*$ if and only if $\lambda \leq \lambda^*$ in $L(m, n)$.

The "if" part of the previous sentence is clear, for if $\lambda \leq \lambda^*$ then $D_\lambda \subseteq D_{\lambda^*}$. So assume there exist $D \in \mathcal{O}$ and $D^* \in \mathcal{O}^*$ satisfying $D \subseteq D^*$. The lengths of the rows of D, written in decreasing order, are $\lambda_1, \ldots, \lambda_m$, and similarly for D^*. Since each row of D is contained in a row of D^*, it follows that for each $1 \leq j \leq m$, D^* has at least j rows of size at least λ_j. Thus the length λ_j^* of the jth largest row of D^* is at least as large as λ_j. In other words, $\lambda_j \leq \lambda_j^*$, as was to be proved. □

Combining the previous theorem with Theorem 5.9 yields:

6.10 Corollary. *The posets $L(m, n)$ are rank-symmetric, rank-unimodal, and Sperner.*

Note that the rank-symmetry and rank-unimodality of $L(m, n)$ can be rephrased as follows: The q-binomial coefficient $\left[{m+n \atop m} \right]$ has symmetric and unimodal coefficients. While rank-symmetry is easy to prove (see Proposition 6.2), the unimodality of the coefficients of $\left[{m+n \atop m} \right]$ is by no means apparent. It was first proved by J. Sylvester in 1878 by a proof similar to the one above, though stated in the language of the invariant theory of binary forms. For a long time it was an open problem to find a combinatorial proof that the coefficients of $\left[{m+n \atop m} \right]$ are unimodal. Such a proof would give an explicit injection (one-to-one function) $\mu : L(m, n)_i \to L(m, n)_{i+1}$ for $i < \frac{1}{2}mn$. (One difficulty in finding such maps μ is to make use of the hypothesis that $i < \frac{1}{2}mn$.) Finally around 1989 such a proof was found by Kathy O'Hara. However, O'Hara's proof has the defect that the maps μ are not order-matchings. Thus her proof does not prove that $L(m, n)$ is Sperner, but only that it's rank-unimodal. It is an outstanding open problem in algebraic combinatorics to find an explicit order-matching $\mu : L(m, n)_i \to L(m, n)_{i+1}$ for $i < \frac{1}{2}mn$.

Note that the Sperner property of $L(m, n)$ (together with the fact that the largest level is in the middle) can be stated in the following simple terms: The largest possible collection $\mathcal{C}$ of Young diagrams fitting in an $m \times n$ rectangle such that no diagram in $\mathcal{C}$ is contained in another diagram in $\mathcal{C}$ is obtained by taking all the diagrams of size $\frac{1}{2}mn$. Although the statement of this fact requires almost no mathematics to understand, there is no known proof that doesn't use algebraic machinery. (The several known algebraic proofs are all closely related, and the one we have given is the simplest.) Corollary 6.10 is a good example of the efficacy of algebraic combinatorics.

An application to number theory. There is an interesting application of Corollary 6.10 to a number-theoretic problem. Fix a positive integer k. For a finite subset S of $\mathbb{R}^+ = \{\alpha \in \mathbb{R} : \alpha > 0\}$, and for a real number $\alpha > 0$, define

$$f_k(S, \alpha) = \# \left\{ T \in \binom{S}{k} : \sum_{t \in T} t = \alpha \right\}$$

In other words, $f_k(S, \alpha)$ is the number of k-element subsets of S whose elements sum to α. For instance, $f_3(\{1, 3, 4, 6, 7\}, 11) = 2$, since $1 + 3 + 7 = 1 + 4 + 6 = 11$.

Given positive integers $k < n$, our object is to maximize $f_k(S, \alpha)$ subject to the condition that $\#S = n$. We are free to choose both S and α, but k and n are fixed. Call this maximum value $h_k(n)$. Thus

$$h_k(n) = \max_{\substack{\alpha \in \mathbb{R}^+ \\ S \subset \mathbb{R}^+ \\ \#S = n}} f_k(S, \alpha).$$

What sort of behavior can we expect of the maximizing set S? If the elements of S are "spread out," say $S = \{1, 2, 4, 8, \dots, 2^{n-1}\}$, then all the subset sums of S are distinct. Hence for any $\alpha \in \mathbb{R}^+$ we have $f_k(S, \alpha) = 0$ or 1. Similarly, if the elements of S are "unrelated" (e.g., linearly independent over the rationals, such as $S = \{1, \sqrt{2}, \sqrt{3}, \pi, \pi^2\}$), then again all subset sums are distinct and $f_k(S, \alpha) = 0$ or 1. These considerations make it plausible that we should take $S = [n] = \{1, 2, \dots, n\}$ and then choose α appropriately. In other words, we are led to the conjecture that for any $S \in \binom{\mathbb{R}^+}{n}$ and $\alpha \in \mathbb{R}^+$, we have

$$f_k(S, \alpha) \leq f_k([n], \beta), \tag{31}$$

for some $\beta \in \mathbb{R}^+$ to be determined.

First let us evaluate $f_k([n], \alpha)$ for any α. This will enable us to determine the value of β in (31). Let $S = \{i_1, \ldots, i_k\} \subseteq [n]$ with

$$1 \le i_1 < i_2 < \cdots < i_k \le n, \quad i_1 + \cdots + i_k = \alpha. \tag{32}$$

Let $j_r = i_r - r$. Then (since $1 + 2 + \cdots + k = \binom{k+1}{2}$)

$$n - k \ge j_k \ge j_{k-1} \ge \cdots \ge j_1 \ge 0, \quad j_1 + \cdots + j_k = \alpha - \binom{k+1}{2}. \tag{33}$$

Conversely, given $j_1, \ldots, j_k$ satisfying (33) we can recover $i_1, \ldots, i_k$ satisfying (32). Hence $f_k([n], \alpha)$ is equal to the number of sequences $j_1, \ldots, j_k$ satisfying (33). Now let

$$\lambda(S) = (j_k, j_{k-1}, \ldots, j_1).$$

Note that $\lambda(S)$ is a partition of the integer $\alpha - \binom{k+1}{2}$ with at most k parts and with largest part at most $n - k$. Thus

$$f_k([n], \alpha) = p_{\alpha - \binom{k+1}{2}}(k, n - k), \tag{34}$$

or equivalently,

$$\sum_{\alpha \ge \binom{k+1}{2}} f_k([n], \alpha) q^{\alpha - \binom{k+1}{2}} = \begin{bmatrix} n \\ k \end{bmatrix}.$$

By the rank-unimodality (and rank-symmetry) of $L(n-k, k)$ (Corollary 6.10), the largest coefficient of $\begin{bmatrix} n \\ k \end{bmatrix}$ is the middle one, that is, the coefficient of $\lfloor k(n-k)/2 \rfloor$. It follows that for fixed k and n, $f_k([n], \alpha)$ is maximized for $\alpha = \lfloor k(n-k)/2 \rfloor + \binom{k+1}{2} = \lfloor k(n+1)/2 \rfloor$. Hence the following result is plausible.

6.11 Theorem. *Let $S \in \binom{\mathbb{R}^+}{n}$, $\alpha \in \mathbb{R}^+$, and $k \in \mathbb{P}$. Then*

$$f_k(S, \alpha) \le f_k([n], \lfloor k(n+1)/2 \rfloor).$$

Proof. Let $S = \{a_1, \ldots, a_n\}$ with $0 < a_1 < \cdots < a_n$. Let T and U be distinct k-element subsets of S with the same element sums, say $T = \{a_{i_1}, \ldots, a_{i_k}\}$ and $U = \{a_{j_1}, \ldots, a_{j_k}\}$ with $i_1 < i_2 < \cdots < i_k$ and $j_1 < j_2 < \cdots < j_k$. Define $T^* = \{i_1, \ldots, i_k\}$ and $U^* = \{j_1, \ldots, j_k\}$, so $T^*, U^* \in \binom{[n]}{k}$. The crucial observation is the following:

Claim. The elements $\lambda(T^*)$ and $\lambda(U^*)$ are incomparable in $L(k, n-k)$, i.e., neither $\lambda(T^*) \leq \lambda(U^*)$ nor $\lambda(U^*) \leq \lambda(T^*)$.

Proof of claim. Suppose not, say $\lambda(T^*) \leq \lambda(U)^*$ to be definite. Thus by definition of $L(k, n-k)$ we have $i_r - r \leq j_r - r$ for $1 \leq r \leq k$. Hence $i_r \leq j_r$ for $1 \leq r \leq k$, so also $a_{i_r} \leq a_{j_r}$ (since $a_1 < \cdots < a_n$). But $a_{i_1} + \cdots + a_{i_k} = a_{j_1} + \cdots + a_{j_k}$ by assumption, so $a_{i_r} = a_{j_r}$ for all r. This contradicts the assumption that T and U are distinct and proves the claim.

It is now easy to complete the proof of Theorem 6.11. Suppose that $S_1, \ldots, S_r$ are distinct k-element subsets of S with the same element sums. By the claim, $\{\lambda(S_1^*), \ldots, \lambda(S_r^*)\}$ is an antichain in $L(k, n-k)$. Hence r cannot exceed the size of the largest antichain in $L(k, n-k)$. By Theorem 6.6 and Corollary 6.10, the size of the largest antichain in $L(k, n-k)$ is given by $p_{\lfloor k(n-k)/2 \rfloor}(k, n-k)$. By equation (34) this number is equal to $f_k([n], \lfloor k(n+1)/2 \rfloor)$. In other words,

$$r \leq f_k([n], \lfloor k(n+1)/2 \rfloor),$$

which is what we wanted to prove. $\square$

Note that an equivalent statement of Theorem 6.11 is that $h_k(n)$ is equal to the coefficient of $q^{\lfloor k(n-k)/2 \rfloor}$ in $\left[{n \atop k}\right]$ [why?].

Variation on a theme. Suppose that in Theorem 6.11 we do not want to specify the cardinality of the subsets of S. In other words, for any $\alpha \in \mathbb{R}$ and any finite subset $S \subset \mathbb{R}^+$, define

$$f(S, \alpha) = \#\{T \subseteq S \, : \, \sum_{t \in T} t = \alpha\}.$$

How large can $f(S, \alpha)$ be if we require $\#S = n$? Call this maximum value $h(n)$. Thus

$$h(n) = \max_{\substack{\alpha \in \mathbb{R}^+ \\ S \subset \mathbb{R}^+ \\ \#S = n}} f(S, \alpha). \tag{35}$$

For instance, if $S = \{1, 2, 3\}$ then $f(S, 3) = 2$ (coming from the subsets $\{1, 2\}$ and $\{3\}$). This is easily seen to be best possible, i.e., $h(3) = 2$.

We will find $h(n)$ in a manner analogous to the proof of Theorem 6.11. The big difference is that the relevant poset $M(n)$ is *not* of the form B_n/G,

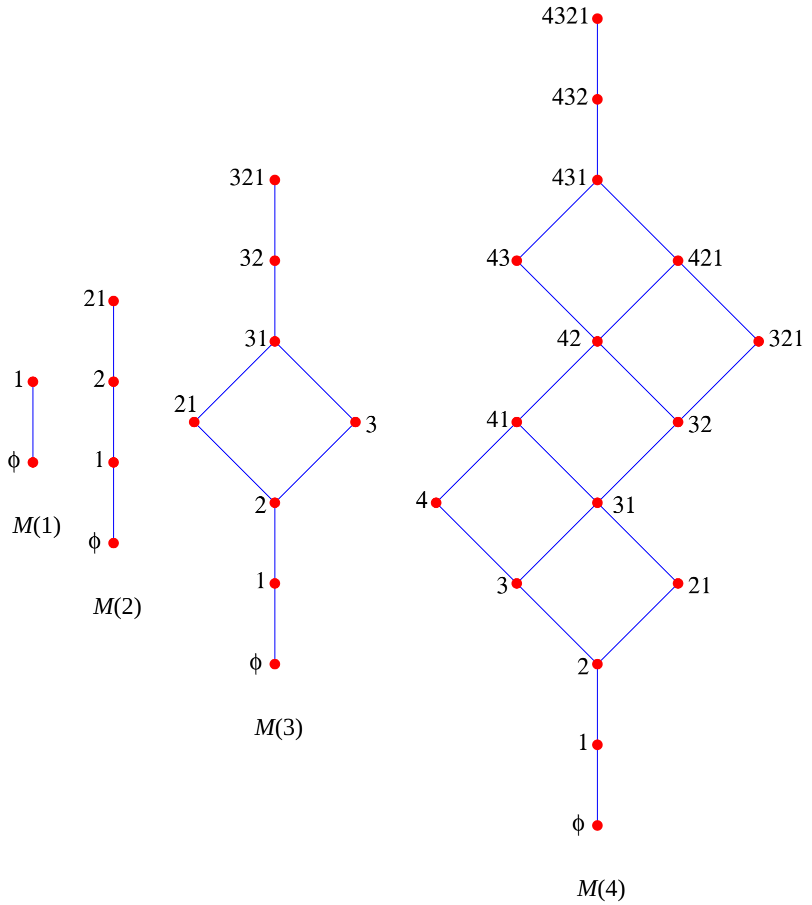


Figure 1: The posets $M(1)$, $M(2)$, $M(3)$ and $M(4)$

so we will have to prove the injectivity of the order-raising operator U_i from scratch. Our proofs will be somewhat sketchy; it shouldn't be difficult for the reader who has come this far to fill in the details.

Let $M(n)$ be the set of all subsets of $[n]$, with the ordering $A \leq B$ if the elements of A are $a_1 > a_2 > \cdots > a_j$ and the elements of B are $b_1 > b_2 > \cdots > b_k$, where $j \leq k$ and $a_i \leq b_i$ for $1 \leq i \leq j$. (The empty set $\emptyset$ is the bottom element of $M(n)$.) Figure 1 shows $M(1)$, $M(2)$, $M(3)$, and $M(4)$.

It is easy to see that $M(n)$ is graded of rank $\binom{n+1}{2}$. The rank of the subset

$T = \{a_1, \ldots, a_k\}$ is

$$\operatorname{rank}(T) = a_1 + \cdots + a_k. \tag{36}$$

It follows [why?] that the rank-generating function of $M(n)$ is given by

$$F(M(n), q) = \sum_{i=0}^{\binom{n+1}{2}} (\#M(n)_i) q^i = (1+q)(1+q^2)\cdots(1+q^n).$$

Define linear transformations

$$U_i : \mathbb{R}M(n)_i \to \mathbb{R}M(n)_{i+1}, \quad D_i : \mathbb{R}M(n)_i \to \mathbb{R}M(n)_{i-1}$$

by

$$\begin{aligned} U_i(x) &= \sum_{\substack{y \in M(n)_{i+1} \\ x < y}} y, \quad x \in M(n)_i \\ D_i(x) &= \sum_{\substack{v \in M(n)_{i-1} \\ v < x}} c(v,x) v, \;\; x \in M(n)_i, \end{aligned}$$

where the coefficient $c(v,x)$ is defined as follows. Let the elements of v be $a_1 > \cdots > a_j > 0$ and the elements of x be $b_1 > \cdots > b_k > 0$. Since x covers v, there is a unique r for which $a_r = b_r - 1$ (and $a_k = b_k$ for all other k). In the case $b_r = 1$ we set $a_r = 0$. (E.g., if x is given by $5 > 4 > 1$ and v by $5 > 4$, then $r = 3$ and $a_3 = 0$.) Set

$$c(v,x) = \begin{cases} \binom{n+1}{2}, & \text{if } a_r = 0 \\ (n - a_r)(n + a_r + 1), & \text{if } a_r > 0. \end{cases}$$

It is a straightforward computation (proof omitted) to obtain the commutation relation

$$D_{i+1}U_i - U_{i-1}D_i = \left(\binom{n+1}{2} - 2i \right) I_i, \tag{37}$$

where I_i denotes the identity linear transformation on $\mathbb{R}M(n)_i$. Clearly by definition U_i is order-raising. We want to show that U_i is injective (one-to-one) for $i < \frac{1}{2}\binom{n+1}{2}$. We can't argue as in the proof of Lemma 4.6 that $U_{i-1}D_i$

is semidefinite since the matrices of U_{i-1} and D_i are no longer transposes of one another. Instead we use the following result from linear algebra. For two proofs, see pp. 331-333 of *Selected Papers on Algebra* (S. Montgomery, *et al.*, eds.), Mathematical Association of America, 1977.

6.12 Lemma. Let V and W be finite-dimensional vector spaces over a field. Let $A : V \to W$ and $B : W \to V$ be linear transformations. Then

$$x^{\dim V} \det(AB - xI) = x^{\dim W} \det(BA - xI).$$

In other words, AB and BA have the same nonzero eigenvalues.

We can now prove the key linear algebraic result.

6.13 Lemma. The linear transformation U_i is injective for $i < \frac{1}{2}\binom{n+1}{2}$ and surjective (onto) for $i \geq \frac{1}{2}\binom{n+1}{2}$.

Proof. We prove by induction on i that $D_{i+1}U_i$ has positive real eigenvalues for $i < \frac{1}{2}\binom{n+1}{2}$. For $i = 0$ this is easy to check since $\dim \mathbb{R}M(n)_0 = 1$. Assume for $i < \frac{1}{2}\binom{n+1}{2} - 1$, i.e., assume that D_iU_{i-1} has positive eigenvalues. By Lemma 6.12, $U_{i-1}D_i$ has nonnegative eigenvalues. By (37), we have

$$D_{i+1}U_i = U_{i-1}D_i + \left(\binom{n+1}{2} - 2i\right) I_i.$$

Thus the eigenvalues of $D_{i+1}U_i$ are $\binom{n+1}{2} - 2i$ more than those of $U_{i-1}D_i$. Since $\binom{n+1}{2} - 2i > 0$, it follows that $D_{i+1}U_i$ has positive eigenvalues. Hence it is invertible, so U_i is injective. Similarly (or by "symmetry") U_i is surjective for $i \geq \frac{1}{2}\binom{n+1}{2}$. □

The main result on the posets $M(n)$ now follows by a familiar argument.

6.14 Theorem. The poset $M(n)$ is graded of rank $\binom{n+1}{2}$, rank-symmetric, rank-unimodal, and Sperner.

Proof. We have already seen that $M(n)$ is graded of rank $\binom{n+1}{2}$ and rank-symmetric. By the previous lemma, U_i is injective for $i < \frac{1}{2}\binom{n+1}{2}$ and surjective for $i \geq \frac{1}{2}\binom{n+1}{2}$. The proof follows from Proposition 4.4 and Lemma 4.5. □

Note. As a consequence of Theorem 6.14, the polynomial $F(M(n),q) = (1+q)(1+q^2)\cdots(1+q^n)$ has unimodal coefficients. No combinatorial proof of this fact is known, unlike the situation for $L(m,n)$ (where we mentioned the proof of O'Hara above).

We can now determine $h(n)$ (as defined by equation (35)) by an argument analogous to the proof of Theorem 6.11.

6.15 Theorem. *Let $S \in \binom{\mathbb{R}^+}{n}$ and $\alpha \in \mathbb{R}^+$. Then*

$$f(S,\alpha) \le f\left([n], \left\lfloor \frac{1}{2}\binom{n+1}{2} \right\rfloor\right) = h(n).$$

Proof. Let $S = \{a_1, \ldots, a_n\}$ with $0 < a_1 < \cdots < a_n$. Let T and U be distinct subsets of S with the same element sums, say $T = \{a_{r_1}, \ldots, a_{r_j}\}$ and $U = \{a_{s_1}, \ldots, a_{s_k}\}$ with $r_1 < r_2 < \cdots < r_j$ and $s_1 < s_2 < \cdots < s_k$. Define $T^* = \{r_1, \ldots, r_j\}$ and $U^* = \{s_1, \ldots, s_k\}$, so $T^*, U^* \in M(n)$. The following fact is proved exactly in the same way as the analogous fact for $L(m,n)$ (the claim in the proof of Theorem 6.11) and will be omitted here.

Fact. The elements T^* and U^* are incomparable in $M(n)$, i.e., neither $T^* \le U^*$ nor $U^* \le T^*$.

It is now easy to complete the proof of Theorem 6.15. Suppose that $S_1, \ldots, S_t$ are distinct subsets of S with the same element sums. By the above fact, $\{S_1^*, \ldots, S_t^*\}$ is an antichain in $M(n)$. Hence t cannot exceed the size of the largest antichain in $M(n)$. By Theorem 6.14, the size of the largest antichain in $M(n)$ is the size $p_{\lfloor \frac{1}{2}\binom{n+1}{2}\rfloor}$ of the middle rank. By equation (36) this number is equal to $f([n], \lfloor \frac{1}{2}\binom{n+1}{2}\rfloor)$. In other words,

$$t \le f\left([n], \left\lfloor \frac{1}{2}\binom{n+1}{2} \right\rfloor\right),$$

which is what we wanted to prove. □

Note. Theorem 6.15 is known as the *weak Erdős-Moser conjecture*. The original (strong) Erdős-Moser conjecture deals with the case $S \subset \mathbb{R}$ rather than $S \subset \mathbb{R}^+$. There is a difference between these two cases; for instance, $h(3) = 2$ (corresponding to $S = \{1, 2, 3\}$ and $\alpha = 3$), while the set $\{-1, 0, 1\}$ has *four* subsets whose elements sum to 0 (including the empty set). (Can you see where the proof of Theorem 6.15 breaks down if we allow $S \subset \mathbb{R}$?) The original Erdős-Moser conjecture asserts that if $\#S = 2m + 1$, then

$$f(S, \alpha) \leq f(\{-m, -m+1, \ldots, m\}, 0).$$

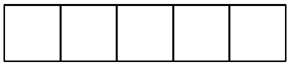
This result can be proved by a somewhat tricky modification of the proof given above for the weak case. No proof of the Erdős-Moser conjecture (weak or strong) is known other than the one indicated here (sometimes given in a more sophisticated context, as explained in the next Note).

Note. The key to the proof of Theorem 6.15 is the definition of U_i and D_i which gives the commutation relation (37). The reader may be wondering how anyone managed to discover these definitions (especially that of D_i). In fact, the original proof of Theorem 6.15 was based on the representation theory of the orthogonal Lie algebra $\mathfrak{o}(2n + 1, \mathbb{C})$. In this context, the definitions of U_i and D_i are built into the theory of the "principal subalgebras" of $\mathfrak{o}(2n + 1, \mathbb{C})$. Robert Proctor was the first to remove the representation theory from the proof and present it solely in terms of linear algebra. See his paper in *Amer. Math. Monthly* **89** (1982), 721–634.

7 Enumeration under group action.

In Sections 5 and 6 we considered the quotient poset B_n/G, where G is a subgroup of the symmetric group $\mathfrak{S}_n$. If p_i is the number of elements of rank i of this poset, then the sequence $p_0, p_1, \ldots, p_n$ is rank-symmetric and rank-unimodal. Thus it is natural to ask whether there is some nice formula for the numbers p_i. For instance, in Theorem 5.10 p_i is the number of nonisomorphic graphs with m vertices (where $n = \binom{m}{2}$) and i edges; is there some nice formula for this number? For the group $G_{mn} = \mathfrak{S}_n \wr \mathfrak{S}_m$ of Theorem 6.6 we obtained a simple generating function for p_i (i.e., a formula for the polynomial $\sum_i p_i q^i$), but this was a very special situation. In this section we will present a general theory for enumerating inequivalent objects subject to a group of symmetries, which will include a formula for the generating function $\sum_i p_i q^i$ as a special case, where p_i is the number of elements of rank i of B_n/G. The chief architect of this theory is G. Pólya (1887–1985) (though much of it was anticipated by J. H. Redfield) and hence is often called *Pólya's theory of enumeration* or just *Pólya theory*.

Pólya theory is most easily understood in terms of "colorings" of some geometric or combinatorial object. For instance, consider a row of five squares:



In how many ways can we color the squares using n colors? Each square can be colored any of the n colors, so there are n^5 ways in all. These colorings can by indicated as

A	B	C	D	E

where A, B, C, D, E are the five colors. Now assume that we are allowed to rotate the five squares 180°, and that two colorings are considered the same if one can be obtained from the other by such a rotation. (We may think that we have cut the row of five squares out of paper and colored them on one side.) We say that two colorings are *equivalent* if they are the same or can be transformed into one another by a 180° rotation. The first naive assumption is that every coloring is equivalent to exactly one other (besides itself), so the number of inequivalent colorings is $n^5/2$. Clearly this reasoning cannot be correct since $n^5/2$ is not always an integer! The problem, of course, is that some colorings stay the same when we rotate 180°. In fact, these are exactly the colorings

A	B	C	B	A

where A, B, C are any three colors. There are n^3 such colorings, so the total number of inequivalent colorings is given by

$$\frac{1}{2}(\text{number of colorings which don't equal their } 180^\circ \text{ rotation})$$

$$+(\text{number of colorings which equal their } 180^\circ \text{ rotation}$$

$$\begin{aligned} &= \frac{1}{2}(n^5 - n^3) + n^3 \\ &= \frac{1}{2}(n^5 + n^3). \end{aligned}$$

Pólya theory gives a systematic method for obtaining formulas of this sort for any underlying symmetry group.

The general setup is the following. Let X be a finite set, and G a subgroup of the symmetric group $\mathfrak{S}_X$. Think of G as a group of symmetries of X. Let C be another set (which may be infinite), which we think of as a set of "colors." A *coloring* of X is a function $f : X \to C$. For instance, X could be the set of four squares of a 2×2 chessboard, labelled as follows:

1	2
3	4

Let $C = \{r, b, y\}$ (the colors red, blue, and yellow). A typical coloring of X would then look like

r	b
y	r

The above diagram thus indicates the function $f : X \to C$ given by $f(1) = r, f(2) = b, f(3) = y, f(4) = r$.

We define two colorings f and g to be *equivalent* (or G-*equivalent*, when it is necessary to specify the group), denoted $f \sim g$ or $f \overset{G}{\sim} g$, if there exists an element $\pi \in G$ such that

$$g(\pi(x)) = f(x) \text{ for all } x \in X.$$

We may write this condition more succinctly as $g\pi = f$, where $g\pi$ denotes the composition of functions (from right to left). It is easy to check, using the fact that G is a group, that $\sim$ is an equivalence relation. One should think that equivalent functions are the same "up to symmetry."

7.1 Example. Let X be the 2×2 chessboard and $C = \{r, b, y\}$ as above. There are many possible choices of a symmetry group G, and this will affect when two colorings are equivalent. For instance, consider the following groups:

- G_1 consists of only the identity permutation $(1)(2)(3)(4)$.
- G_2 is the group generated by a vertical reflection. It consists of the two elements $(1)(2)(3)(4)$ (the identity element) and $(1,2)(3,4)$ (the vertical reflection).
- G_3 is the group generated by a reflection in the main diagonal. It consists of the two elements $(1)(2)(3)(4)$ (the identity element) and $(1)(4)(2,3)$ (the diagonal reflection).
- G_4 is the group of all rotations of X. It is a cyclic group of order four with elements $(1)(2)(3)(4)$, $(1,2,4,3)$, $(1,4)(2,3)$, and $(1,3,4,2)$.
- G_5 is the dihedral group of all rotations and reflections of X. It has eight elements, namely, the four elements of G_4 and the four reflections $(1,2)(3,4)$, $(1,3)(2,4)$, $(1)(4)(2,3)$, and $(2)(3)(1,4)$.
- G_6 is the symmetric group of *all* 24 permutations of X. Although this is a perfectly valid group of symmetries, it no longer has any connection with the geometric representation of X as the squares of a 2×2 chessboard.

Consider the inequivalent colorings of X with two red squares, one blue square, and one yellow square, in each of the six cases above.

(G_1) There are twelve colorings in all with two red squares, one blue square, and one yellow square, and all are inequivalent under the trivial group (the group with one element). In general, whenever G is the trivial group then two colorings are equivalent if and only if they are the same [why?].

(G_2) There are now six inequivalent colorings, represented by

$$\begin{array}{|c|c|}\hline r & r \\ \hline b & y \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & b \\ \hline r & y \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & y \\ \hline r & b \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline b & y \\ \hline r & r \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & b \\ \hline y & r \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & y \\ \hline b & r \\ \hline\end{array}$$

Each equivalence class contains two elements.

(G_3) Now there are seven classes, represented by

$$\begin{array}{|c|c|}\hline r & r \\ \hline b & y \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & r \\ \hline y & b \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline b & y \\ \hline r & r \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline y & b \\ \hline r & r \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & b \\ \hline y & r \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline b & r \\ \hline r & y \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline y & r \\ \hline r & b \\ \hline\end{array}$$

The first five classes contain two elements each and the last two classes only one element. Although G_2 and G_3 are isomorphic as abstract groups, as permutation groups they have a different structure. Specifically, the generator $(1,2)(3,4)$ of G_2 has two cycles of length two, while the generator $(1)(4)(2,3)$ has two cycles of length one and one of length two. As we will see below, it is the lengths of the cycles of the elements of G that determine the sizes of the equivalence classes. This explains why the number of classes for G_2 and G_3 are different.

(G_4) There are three classes, each with four elements. The size of each class is equal to the order of the group because none of the colorings have any symmetry with respect to the group, i.e., for any coloring f, the only group element π that fixes f (so $f\pi = f$) is the identity ($\pi = (1)(2)(3)(4)$).

$$\begin{array}{|c|c|}\hline r & r \\ \hline y & b \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & r \\ \hline b & y \\ \hline\end{array}\quad \begin{array}{|c|c|}\hline r & b \\ \hline y & r \\ \hline\end{array}$$

(G_5) Under the full dihedral group there are now two classes.

r	r
b	y

r	b
y	r

The first class has eight elements and the second four elements. In general, the size of a class is the index in G of the subgroup fixing some fixed coloring in that class [why?]. For instance, the subgroup fixing the second coloring above is $\{(1)(2)(3)(4), (1,4)(2)(3)\}$, which has index four in the dihedral group of order eight.

(G_6) Under the group $\mathfrak{S}_4$ of all permutations of the squares there is clearly only one class, with all twelve colorings. In general, for any set X if the group is the symmetric group $\mathfrak{S}_X$ then two colorings are equivalent if and only if each color appears the same number of times [why?].

Our object in general is to count the number of equivalence classes of colorings which use each color a specified number of times. We will put the information into a *generating function* — a polynomial whose coefficients are the numbers we seek. Consider for example the set X, the group $G = G_5$ (the dihedral group), and the set $C = \{r, b, y\}$ of colors in Example 7.1 above. Let $\kappa(i, j, k)$ be the number of inequivalent colorings using red i times, blue j times, and yellow k times. Think of the colors r, b, y as *variables*, and form the polynomial

$$F_G(r, b, y) = \sum_{i+j+k=4} \kappa(i, j, k) r^i b^j y^k.$$

Note that we sum only over i, j, k satisfying $i + j + k = 4$ since a total of four colors will be used to color the four-element set X. The reader should check that

$$\begin{aligned} F_G(r, b, y) &= (r^4 + b^4 + y^4) + (r^3 b + r b^3 + r^3 y + r y^3 + b^3 y + b y^3) \\ &\quad +2(r^2 b^2 + r^2 y^2 + b^2 y^2) + 2(r^2 b y + r b^2 y + r b y^2). \end{aligned}$$

For instance, the coefficient of $r^2 by$ is two because, as we have seen above, there are two inequivalent colorings using the colors r, r, b, y. Note that $F_G(r, b, y)$ is a *symmetric function* of the variables r, b, y (i.e., it stays the same if we permute the variables in any way), because insofar as counting inequivalent colorings goes, it makes no difference what *names* we give the colors. As a special case we may ask for the *total* number of inequivalent colorings with four colors. This obtained by setting $r = b = y = 1$ in $F_G(r, b, y)$ [why?], yielding $F_G(1, 1, 1) = 3 + 6 + 2 \cdot 3 + 2 \cdot 3 = 21$.

What happens to the generating function F_G in the above example when we use the n colors $r_1, r_2, \ldots, r_n$ (which can be thought of as different shades of red)? Clearly all that matters are the *multiplicities* of the colors, without regard for their order. In other words, there are five cases: (a) all four colors the same, (b) one color used three times and another used once, (c) two colors used twice each, (d) one color used twice and two others once each, and (e) four colors used once each. These five cases correspond to the five partitions of 4, i.e., the five ways of writing 4 as a sum of positive integers without regard to order: 4, 3+1, 2+2, 2+1+1, 1+1+1+1. Our generating function becomes

$$F_G(r_1, r_2, \ldots, r_n) = \sum_i r_i^4 + \sum_{i \neq j} r_i^3 r_j + 2 \sum_{i<j} r_i^2 r_j^2 + 2 \sum_{\substack{i \neq j \\ i \neq k \\ j<k}} r_i^2 r_j r_k + 3 \sum_{i<j<k<l} r_i r_j r_k r_l,$$

where the indices in each sum lie between 1 and n. If we set all variables equal to one (obtaining the total number of colorings with n colors), then simple combinatorial reasoning yields

$$\begin{aligned} F_G(1, 1, \ldots, 1) &= n + n(n-1) + 2\binom{n}{2} + 2n\binom{n-1}{2} + 3\binom{n}{4} \\ &= \frac{1}{8}(n^4 + 2n^3 + 3n^2 + 2n). \qquad (38) \end{aligned}$$

Note that the polynomial (38) has the following description: The denominator 8 is the order of the group G_5, and the coefficient of n^i in the numerator is just the number of permutations in G_5 with i cycles! For instance, the coefficient of n^2 is 3, and G_5 has the three elements $(1, 2)(3, 4)$, $(1, 3)(2, 4)$, and $(1, 4)(2, 3)$ with two cycles. We want to prove a general result of this nature.

The basic tool which we will use is a simple result from the theory of permutation groups known as *Burnside's lemma*. It was actually first proved by Cauchy when G is transitive (i.e., $|Y/G| = 1$) and by Frobenius in the general case, and is sometimes called the *Cauchy-Frobenius lemma*.

7.2 Lemma. (Burnside's lemma) *Let Y be a finite set and G a subgroup of $\mathfrak{S}_Y$. For each $\pi \in G$, let*

$$\mathrm{Fix}(\pi) = \{y \in Y : \pi(y) = y\},$$

so $|\mathrm{Fix}(\pi)|$ is the number of cycles of length one in the permutation π. Let Y/G be the set of orbits of G. Then

$$|Y/G| = \frac{1}{|G|} \sum_{\pi \in G} |\mathrm{Fix}(\pi)|.$$

An equivalent form of Burnside's lemma is that statement that the average number of elements of Y fixed by an element of G is equal to the number of orbits. Before proceeding to the proof, let us consider an example.

7.3 Example. Let $Y = \{a, b, c, d\}$, $G = \{(a)(b)(c)(d), (a, b)(c, d), (a, c), (b, d), (a, d)(b, c)\}$, and $G' = \{(a)(b)(c)(d), (a, b)(c)(d), (a)(b)(c, d), (a, b)(c, d)\}$. Both groups are isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$ (compare Example 5.1(c) and (d)). By Burnside's lemma the number of orbits of G is $\frac{1}{4}(4 + 0 + 0 + 0) = 1$. Indeed, given any two elements $i, j \in Y$, it is clear by inspection that there is a $\pi \in G$ (which happens to be unique) such that $\pi(i) = j$. On the other hand, the number of orbits of G' is $\frac{1}{4}(4 + 2 + 2 + 0) = 2$. Indeed, the two orbits are $\{a, b\}$ and $\{c, d\}$.

Proof of Burnside's lemma. For $y \in Y$ let $G_y = \{\pi \in G : \pi \cdot y = y\}$ (the set of permutations fixing y). Then

$$\begin{aligned}
\frac{1}{|G|} \sum_{\pi \in G} |\mathrm{Fix}(\pi)| &= \frac{1}{|G|} \sum_{\pi \in G} \sum_{\substack{y \in Y \\ \pi \cdot y = y}} 1 \\
&= \frac{1}{|G|} \sum_{y \in Y} \sum_{\substack{\pi \in G \\ \pi \cdot y = y}} 1 \\
&= \frac{1}{|G|} \sum_{y \in Y} |G_y|.
\end{aligned}$$

Now (as in the proof of Lemma 5.7) the multiset of elements $\pi \cdot y$, $\pi \in G$, contains every element in the orbit Gy the same number of times, namely

$|G|/|Gy|$ times. Thus y occurs $|G|/|Gy|$ times among the $\pi \cdot y$, so

$$\frac{|G|}{|Gy|} = |G_y|.$$

Thus

$$\begin{aligned} \frac{1}{|G|}\sum_{\pi\in G}|\mathrm{Fix}(\pi)| &= \frac{1}{|G|}\sum_{y\in Y}\frac{|G|}{|Gy|} \\ &= \sum_{y\in Y}\frac{1}{|Gy|}. \end{aligned}$$

How many times does a term $1/|\mathcal{O}|$ appear in the above sum, where $\mathcal{O}$ is a fixed orbit? We are asking for the number of y such that $Gy = \mathcal{O}$. But $Gy = \mathcal{O}$ if and only if $y \in \mathcal{O}$, so $1/|\mathcal{O}|$ appears $|\mathcal{O}|$ times. Thus each orbit gets counted exactly once, so the above sum is equal to the number of orbits. □

7.4 Example. How many inequivalent colorings of the vertices of a regular hexagon H are there using n colors, under cyclic symmetry? Let $\mathcal{C}_n$ be the set of all n-colorings of H. Let G be the group of all permutations of $\mathcal{C}_n$ which permute the colors cyclically, so $G \cong \mathbb{Z}_6$. We are asking for the number of orbits of G [why?]. We want to apply Burnside's lemma, so for each of the six elements σ of G we need to compute the number of colorings fixed by that element. Let π be a generator of G.

- $\sigma = 1$ (the identity): All n^6 colorings are fixed by σ.
- $\sigma = \pi, \pi^{-1}$: Only the n colorings with all colors equal are fixed.
- $\sigma = \pi^2, \pi^4$: Any coloring of the form *ababab* is fixed (writing the colors linearly in the order they appear around the hexagon, starting at any fixed vertex). There are n choices for a and n for b, so n^2 colorings in all.
- $\sigma = \pi^3$: The fixed colorings are of the form *abcabc*, so n^3 in all.

Hence by Burnside's lemma, we have

$$\text{number of orbits } = \frac{1}{6}(n^6 + n^3 + 2n^2 + 2n).$$

The reader who has followed the preceding example will have no trouble understanding the following result.

7.5 Theorem. *Let G be a group of permutations of a finite set X. Then the number of inequivalent (with respect to G) n-colorings of X is equal to*

$$\frac{1}{|G|}\sum_{\pi\in G} n^{c(\pi)},$$

where $c(\pi)$ denotes the number of cycles of π.

Proof. Let π_n denote the action of $\pi \in G$ on the set $\mathcal{C}_n$ of n-colorings of X. We want to determine the set $\mathrm{Fix}(\pi_n)$, so that we can apply Burnside's lemma. Let C be the set of n colors. If $f : X \to C$ is a coloring fixed by π, then for all $x \in X$ we have

$$f(x) = \pi_n \cdot f(x) = f(\pi \cdot x).$$

Thus $f \in \mathrm{Fix}(\pi_n)$ if and only if $f(x) = f(y)$ whenever $\pi(x) = y$. In other words, we must have $f(x) = f(\pi(x))$. Hence $f(x) = f(\pi^k(x))$ for any $k \geq 1$ [why?]. The elements y of X of the form $\pi^k(x)$ for $k \geq 1$ are just the elements of the cycle of π containing x. Thus to obtain $f \in \mathrm{Fix}(\pi_n)$, we should take the cycles $\sigma_1, \dots, \sigma_{c(\pi)}$ of π and color each element of σ_i the same color. There are n choices for each σ_i, so $n^{c(\pi)}$ colorings in all fixed by π. In other words, $|\mathrm{Fix}(\pi_n)| = n^{c(\pi)}$, and the proof follows by Burnside's lemma. □

We would now like not just to count the *total* number of inequivalent colorings with n-colors, but more strongly to specify the number of occurences of each color. We will need to use not just the number $c(\pi)$ of cycles of each $\pi \in G$, but rather the lengths of each of the cycles of π. Thus given a permutation π of an n-element set X, define the *type* of π to be

$$\mathrm{type}(\pi) = (c_1, c_2, \dots, c_n),$$

where π has c_i i-cycles. For instance, if $\pi = 4, 7, 3, 8, 2, 10, 11, 1, 6, 9, 5$, then

$$\begin{aligned} \mathrm{type}(\pi) &= \mathrm{type}\ (1,4,8)(2,7,11,5)(3)(6,10,9) \\ &= (1,0,2,1,0,0,0,0,0,0,0). \end{aligned}$$

Note that we always have $\sum_i ic_i = n$ [why?]. Define the *cycle indicator* of π to be the monomial

$$Z_\pi = z_1^{c_1} z_2^{c_2} \cdots z_n^{c_n}.$$

(Many other notations are used for the cycle indicator. The use of Z_π comes from the German word *Zyklus* for cycle. The original paper of Pólya was written in German.) Thus for the example above, we have $Z_\pi = z_1 z_3^2 z_4$.

Now given a subgroup G of $\mathfrak{S}_X$, the *cycle indicator* (or *cycle index polynomial*) of G is defined by

$$Z_G = Z_G(z_1, \dots, z_n) = \frac{1}{|G|} \sum_{\pi \in G} Z_\pi.$$

Thus Z_G (also denoted P_G, $\mathrm{Cyc}(G)$, etc.) is a polynomial in the variables $z_1, \dots, z_n$.

7.6 Example. If X consists of the vertices of a square and G is the group of rotations of X (a cyclic group of order 4), then

$$Z_G = \frac{1}{4}(z_1^4 + z_2^2 + 2z_4).$$

If reflections are also allowed (so G is the dihedral group of order 8), then

$$Z_G = \frac{1}{8}(z_1^4 + 3z_2^2 + 2z_1^2 z_2 + 2z_4).$$

We are now ready to state the main result of this section.

7.7 Theorem. (Pólya's theorem, 1937) *Let G be a group of permutations of the n-element set X. Let $C = \{r_1, r_2, \dots\}$ be a set of colors. Let $\kappa(i_1, i_2, \dots)$ be the number of inequivalent (under the action of G) colorings $f : X \to C$ such that color r_j is used i_j times. Define*

$$F_G(r_1, r_2, \dots) = \sum_{i_1, i_2, \dots} \kappa(i_1, i_2, \dots) r_1^{i_1} r_2^{i_2} \cdots.$$

(Thus F_G is a polynomial or a power series in the variables $r_1, r_2, \dots$, depending on whether or not C is finite or infinite.) Then

$$F_G(r_1, r_2, \dots) = Z_G(r_1 + r_2 + r_3 + \cdots, r_1^2 + r_2^2 + r_3^2 + \cdots, \dots, r_1^j + r_2^j + r_3^j + \cdots).$$

(In other words, substitute $\sum_i r_i^j$ for z_j in Z_G.)

Before giving the proof let us consider an example.

7.8 Example. Suppose that in Example 7.6 our set of colors is $C = \{a, b, c, d\}$, and that we take G to be the group of cyclic symmetries. Then

$$\begin{aligned} F_G(a,b,c,d) &= \frac{1}{4}\left((a+b+c+d)^4 + (a^2+b^2+c^2+d^2)^2 + 2(a^4+b^4+c^4+d^4)\right) \\ &= (a^4+\cdots) + (a^3b+\cdots) + 2(a^2b^2+\cdots) + 3(a^2bc+\cdots) + 6abcd. \end{aligned}$$

An expression such as $(a^2b^2+\cdots)$ stands for the sum of all monomials in the variables a, b, c, d with exponents $2, 2, 0, 0$ (in some order). The coefficient of all such monomials is 2, indicating two inequivalent colorings using one color twice and another color twice. If instead G were the full dihedral group, we would get

$$\begin{aligned} F_G(a,b,c,d) &= \frac{1}{8}\big((a+b+c+d)^4 + 3(a^2+b^2+c^2+d^2)^2 \\ &\qquad + 2(a+b+c+d)^2(a^2+b^2+c^2+d^2) + 2(a^4+b^4+c^4+d^4)\big) \\ &= (a^4+\cdots) + (a^3b+\cdots) + 2(a^2b^2+\cdots) + 2(a^2bc+\cdots) + 3abcd. \end{aligned}$$

Proof of Pólya's theorem. Let $|X| = t$ and $i_1 + i_2 + \cdots = t$, where each $i_j \geq 0$. Let $i = (i_1, i_2, ...)$, and let $\mathcal{C}_{\mathbf{i}}$ denote the set of all colorings of X with color r_j used i_j times. The group G acts on $\mathcal{C}_i$, since if $f \in \mathcal{C}_i$ and $\pi \in G$, then $\pi \cdot f \in \mathcal{C}_i$. ("Rotating" a colored object does not change how many times each color appears.) Let π_i denote the action of π on $\mathcal{C}_i$. We want to apply Burnside's lemma to compute the number of orbits, so we need to find $|\mathrm{Fix}(\pi_i)|$.

In order for $f \in \mathrm{Fix}(\pi_i)$, we must color X so that (a) in any cycle of π, all the elements get the same color, and (b) the color r_j appears i_j times. Consider the product

$$H_\pi = \prod_j (r_1^j + r_2^j + \cdots)^{c_j(\pi)},$$

where $c_j(\pi)$ is the number of j-cycles (cycles of length j) of π. When we expand this product as a sum of monomials $r_1^{j_1} r_2^{j_2} \cdots$, we get one of these monomials by choosing a term r_k^j from each factor of H_π and multiplying these terms together. Choosing r_k^j corresponds to coloring all the elements of some j-cycle with r_k. Since a factor $r_1^j + r_2^j + \cdots$ occurs precisely $c_j(\pi)$ times in H_π, choosing a term r_k^j from every factor corresponds to coloring X so that every cycle is monochromatic (i.e., all the elements of that cycle get the same color). The product of these terms r_k^j will be the monomial $r_1^{j_1} r_2^{j_2} \cdots$, where we have used color r_k a total of j_k times. It follows that the coefficient of $r_i^{i_1} r_2^{i_2} \cdots$ in H_π is equal to $|\mathrm{Fix}(\pi_i)|$. Thus

$$H_\pi = \sum_i |\mathrm{Fix}(\pi_i)| r_1^{i_1} r_2^{i_2} \cdots. \tag{39}$$

Now sum both sides of (39) over all $\pi \in G$ and divide by $|G|$. The left-hand side becomes

$$\frac{1}{|G|} \sum_{\pi \in G} \prod_j (r_1^j + r_2^j + \cdots)^{c_j(\pi)} = Z_G(r_1 + r_2 + \cdots, r_1^2 + r_2^2 + \cdots, \ldots).$$

On the other hand, the right-hand side becomes

$$\sum_i \left[\frac{1}{|G|} \sum_{\pi \in G} |\mathrm{Fix}(\pi_i)| \right] r_1^{i_1} r_2^{i_2} \cdots.$$

By Burnside's lemma, the expression in brackets is just the number of orbits of π_i acting on $\mathcal{C}_i$, i.e., the number of inequivalent colorings using color r_j a total of i_j times, as was to be proved. □

7.9 Example. (Necklaces) A *necklace* of length ℓ is a circular arrangement of ℓ (colored) beads. Two necklaces are considered the same if they are cyclic rotations of one another. Let X be a set of ℓ (uncolored) beads, say $X = \{1, 2, \ldots, \ell\}$. Regarding the beads as being placed equidistantly on a circle in the order $1, 2 \ldots, \ell$, let G be the cyclic group of rotations of X. Thus if π is the cycle $(1, 2, \ldots, \ell)$, then $G = \{1, \pi, \pi^2, \ldots, \pi^{\ell-1}\}$. For example, if $\ell = 6$ then the elements of G are

$$\begin{aligned} \pi^0 &= (1)(2)(3)(4)(5)(6) \\ \pi &= (1,2,3,4,5,6) \\ \pi^2 &= (1,3,5)(2,4,6) \\ \pi^3 &= (1,4)(2,5)(3,6) \\ \pi^4 &= (1,5,3)(2,6,4) \\ \pi^5 &= (1,6,5,4,3,2). \end{aligned}$$

In general, if d is the greatest common divisor of m and ℓ (denoted $d = \gcd(m, \ell)$), then π^m has d cycles of length ℓ/d. An integer m satisfies $1 \leq m \leq \ell$ and $\gcd(m, \ell) = d$ if and only if $1 \leq m/d \leq \ell/d$ and $\gcd(m/d, \ell/d) = 1$. Hence the number of such integers m is given by the Euler phi-function (or totient function) $\phi(\ell/d)$, which by definition is equal to the number of integers $1 \leq i \leq \ell/d$ such that $\gcd(i, \ell/d) = 1$. Recall that $\phi(k)$ can be computed by the formula

$$\phi(k) = k \prod_{\substack{p|k \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right).$$

For instance, $\phi(1000) = 1000(1 - \frac{1}{2})(1 - \frac{1}{5}) = 400$. Putting all this together gives the following formula for the cycle enumerator $Z_G(z_1, \ldots, z_\ell)$:

$$Z_G(z_1, \ldots, z_\ell) = \frac{1}{\ell} \sum_{d|\ell} \phi(\ell/d) z_{\ell/d}^d,$$

or (substituting ℓ/d for d),

$$Z_G(z_1, \ldots, z_\ell) = \frac{1}{\ell} \sum_{d|\ell} \phi(d) z_d^{\ell/d}.$$

There follows from Pólya's theorem the following result (originally proved by P. A. MacMahon (1854–1929) before Pólya discovered his general result).

7.10 Theorem.

(a) *The number $N_\ell(n)$ of n-colored necklaces of length ℓ is given by*

$$N_\ell(n) = \frac{1}{\ell} \sum_{d|\ell} \phi(\ell/d) n^d. \tag{40}$$

(b) *We have*

$$F_G(r_1, r_2, \ldots) = \frac{1}{\ell} \sum_{d|\ell} \phi(d)(r_1^d + r_2^d + \cdots)^{\ell/d}.$$

NOTE: (b) reduces to (a) if $r_1 = r_2 = \cdots = 1$. Moreover, since clearly $N_\ell(1) = 1$, putting $n = 1$ in (40) yields the famous identity

$$\sum_{d|\ell} \phi(\ell/d) = \ell.$$

What if we are allowed to flip necklaces over, not just rotate them? Now the group becomes the dihedral group of order 2ℓ, and the corresponding inequivalent colorings are called *dihedral necklaces*. We leave to the reader to work out the cycle enumerators

$$\frac{1}{2\ell}\left(\sum_{d|\ell}\phi(d)z_d^{\ell/d}+mz_1^2z_2^{m-1}+mz_2^m\right),\quad \text{if } \ell=2m$$

$$\frac{1}{2\ell}\left(\sum_{d|\ell}\phi(d)z_d^{\ell/d}+\ell z_1z_2^m\right),\quad \text{if } \ell=2m+1.$$

7.11 Example. Let $G=\mathfrak{S}_\ell$, the group of all permutations of $\{1,2,\ldots,\ell\}=X$. Thus for instance

$$\begin{aligned} Z_{\mathfrak{S}_3}(z_1,z_2,z_3) &= \frac{1}{6}(z_1^3+3z_1z_2+2z_3) \\ Z_{\mathfrak{S}_4}(z_1,z_2,z_3,z_4) &= \frac{1}{24}(z_1^4+6z_1^2z_2+3z_2^2+8z_1z_3+6z_4). \end{aligned}$$

It is easy to count the number of inequivalent colorings in $\mathcal{C}_i$. If two colorings of X use each color the same number of times, then clearly there is *some* permutation of X which sends one of the colorings to the other. Hence $\mathcal{C}_i$ consists of a single orbit. Thus

$$F_{\mathfrak{S}_\ell}(r_1,r_2,\ldots)=\sum_{i_1+i_2+\cdots=\ell} r_1^{i_1}r_2^{i_2}\cdots,$$

the sum of all monomials of degree ℓ.

To count the total number of inequivalent n-colorings, note that

$$\sum_{\ell\geq 0}F_{\mathfrak{S}_\ell}(r_1,r_2,\ldots)x^\ell=\frac{1}{(1-r_1x)(1-r_2x)\cdots}. \qquad (41)$$

since if we expand each factor on the right-hand side into the series $\sum_{j\geq 0}r_i^jx^j$ and multiply, the coefficient of x^ℓ will just be the sum of all monomials of degree ℓ. For fixed n, let $f_n(\ell)$ denote the number of inequivalent n-colorings of X. Since $f_n(\ell) = F_{\mathfrak{S}_\ell}(1,1,\dots,1)$ (n 1's in all), there follows from (41) that

$$\sum_{\ell\geq 0} f_n(\ell)x^\ell = \frac{1}{(1-x)^n}.$$

The right-hand side can be expanded (e.g. by Taylor's theorem) as

$$\frac{1}{(1-x)^n} = \sum_{\ell\geq 0}\binom{n+\ell-1}{\ell}x^\ell.$$

Hence

$$f_n(\ell) = \binom{n+\ell-1}{\ell}.$$

It is natural to ask whether there might be a more direct proof of such a simple result. This is actually a standard result in elementary enumerative combinatorics. For fixed ℓ and n we want the number of solutions to $i_1 + i_2 + \cdots + i_n = \ell$ in nonnegative integers. Setting $k_j = i_j + 1$, this is the same as the number of solutions to $k_1 + k_2 + \cdots + k_n = \ell + n$ in *positive* integers. Place $\ell + n$ dots in a horizontal line. There are $\ell + n - 1$ spaces between the dots. Choose $n-1$ of these spaces and draw a vertical bar in them in $\binom{n+\ell-1}{n-1} = \binom{n+\ell-1}{\ell}$ ways. For example, if $n = 5$ and $\ell = 6$, then one way of drawing the bars is

$$\bullet \quad \bullet \;\Bigg|\; \bullet \quad \bullet \quad \bullet \;\Bigg|\; \bullet \quad \bullet \;\Bigg|\; \bullet \;\Bigg|\; \bullet \quad \bullet \quad \bullet$$

The number of dots in each "compartment," read from left to right, gives the numbers $k_1,\dots,k_n$. For the above example we get $2+3+2+1+3 = 11$, corresponding to the original solution $1+2+1+0+2 = 6$ (i.e., one element of X colored r_1, two elements colored r_2, one colored r_3, and two colored r_5). Since this correspondence between solutions to $i_1 + i_2 + \cdots + i_n = \ell$ and sets of bars is clearly a bijection, we get $\binom{n+\ell-1}{\ell}$ solutions as claimed.

Recall (Theorem 7.5) that the number of inequivalent n-colorings of X (with respect to any group G of permutations of X) is given by

$$\frac{1}{|G|}\sum_{\pi\in G} n^{c(\pi)},$$

where $c(\pi)$ denotes the number of cycles of π. Hence for $G = \mathfrak{S}_\ell$ we get the identity

$$\begin{aligned} \frac{1}{\ell!}\sum_{\pi\in\mathfrak{S}_\ell} n^{c(\pi)} &= \binom{n+\ell-1}{\ell} \\ &= \frac{1}{\ell!}n(n+1)(n+2)\cdots(n+\ell-1). \end{aligned}$$

Multiplying by $\ell!$ yields

$$\sum_{\pi\in\mathfrak{S}_\ell} n^{c(\pi)} = n(n+1)(n+2)\cdots(n+\ell-1). \tag{42}$$

Equivalently [why?], if we define $c(\ell,k)$ to be the number of permutations in $\mathfrak{S}_\ell$ with k cycles (called a *signless Stirling number of the first kind*), then

$$\sum_{k=1}^{\ell} c(\ell,k)x^k = x(x+1)(x+2)\cdots(x+\ell-1).$$

For instance, $x(x+1)(x+2)(x+3) = x^4+6x^3+11x^2+6x$, so (taking the coefficient of x^2) eleven permutations in $\mathfrak{S}_4$ have two cycles, namely, (123)(4), (132)(4), (124)(3), (142)(3), (134)(2), (143)(2), (234)(1), (243)(1), (12)(34), (13)(24), (14)(23).

Although it was easy to compute the generating function $F_{\mathfrak{S}_\ell}(r_1,r_2,\ldots)$ directly without the necessity of computing the cycle indicator $Z_{\mathfrak{S}_\ell}(z_1,\ldots,z_\ell)$, we can still ask whether there is a formula of some kind for this polynomial. First we determine explicitly its coefficients.

7.12 Theorem. *Let $\sum ic_i = \ell$. The number of permutations $\pi \in \mathfrak{S}_\ell$ with c_i cycles of length i (or equivalently, the coefficient of $z_1^{c_1}z_2^{c_2}\cdots$ in $\ell!Z_{\mathfrak{S}_\ell}(z_1,\ldots,z_\ell)$) is equal to $\ell!/1^{c_1}c_1!2^{c_2}c_2!\cdots$.*

Example. The number of permutations in $\mathfrak{S}_{15}$ with three 1-cycles, two 2-cycles, and two 4-cycles is $15!/1^3\cdot 3!\cdot 2^2\cdot 2!\cdot 4^2\cdot 2! = 851,350,500$.

Proof of Theorem 7.12. Fix $c = (c_1,c_2,\ldots)$ and let X_c be the set of all permutations $\pi\in\mathfrak{S}_\ell$ with c_i cycles of length i. Given a permutation $\sigma = a_1a_2\cdots a_\ell$ in $\mathfrak{S}_\ell$, construct a permutation $f(\sigma)\in X_c$ as follows. Let the 1-cycles of $f(\sigma)$ be $(a_1),(a_2),\ldots,(a_{c_1})$. Then let the 2-cycles of $f(\sigma)$ be $(a_{c_1+1},a_{c_1+2}),(a_{c_1+3},a_{c_1+4}),\ldots$, $(a_{c_1+2c_2-1},a_{c_1+2c_2})$. Then let the 3-cycles of $f(\sigma)$ be $(a_{c_1+2c_2+1},a_{c_1+2c_2+2},a_{c_1+2c_2+3})$, $(a_{c_1+2c_2+4},a_{c_1+2c_2+5},a_{c_1+2c_2+6})$, $\ldots$, $(a_{c_1+2c_2+3c_3-2},a_{c_1+2c_2+3c_3-1},a_{c_1+2c_2+3c_3})$, etc., continuing until we reach a_ℓ and have produced a permutation in X_c. For instance, if $\ell=11, c_1=3, c_2=2, c_4=1$, and $\sigma=4,9,6,11,7,1,3,8,10,2,5$, then

$$f(\sigma)=(4)(9)(6)(11,7)(1,3)(8,10,2,5).$$

We have defined a function $f:\mathfrak{S}_\ell\to X_c$. Given $\pi\in X_c$, what is $\#f^{-1}(\pi)$, the number of permutations sent to π by f? A cycle of length i can be written in i ways, namely,

$$(b_1,b_2,\ldots,b_i)=(b_2,b_3,\ldots,b_i,b_1)=\cdots=(b_i,b_1,b_2,\ldots,b_{i-1}).$$

Moreover, there are $c_i!$ ways to order the c_i cycles of length i. Hence

$$\#f^{-1}(\pi)=c_1!c_2!c_3!\cdots 1^{c_1}2^{c_2}3^{c_3}\cdots,$$

the same number for any $\pi\in X_c$. It follows that

$$\begin{aligned}\#X_c &= \frac{\#\mathfrak{S}_\ell}{c_1!c_2!\cdots 1^{c_1}2^{c_2}\cdots}\\ &= \frac{\ell!}{c_1!c_2!\cdots 1^{c_1}2^{c_2}\cdots},\end{aligned}$$

as was to be proved. □

As for the polynomial $Z_{\mathfrak{S}_\ell}$ itself, we have the following result. Write $\exp y=e^y$.

7.13 Theorem. *We have*

$$\sum_{\ell\geq 0} Z_{\mathfrak{S}_\ell}(z_1,z_2,\ldots)x^\ell=\exp\left(z_1x+z_2\frac{x^2}{2}+z_3\frac{x^3}{3}+\cdots\right).$$

Proof. There are some sophisticated ways to prove this theorem which "explain" why the exponential function appears, but we will be content here with a "naive" proof. Write

$$\begin{aligned} e^{z_1x+z_2\frac{x^2}{2}+z_3\frac{x^3}{3}+\cdots} &= e^{zx}\cdot e^{z_2\frac{x^2}{2}}\cdot e^{z_3\frac{x^3}{3}}\cdots \\ &= \left(\sum_{n\geq 0}\frac{z_1^n x^n}{n!}\right)\left(\sum_{n\geq 0}\frac{z_2^n x^{2n}}{2^n n!}\right)\left(\sum_{n\geq 0}\frac{z_3^n x^{3n}}{3^n n!}\right)\cdots. \end{aligned}$$

When we multiply this product out, the coefficient of $z_1^{c_1}z_2^{c_2}\cdots x^\ell$, where $\ell = c_1+2c_2+\cdots$, is given by

$$\frac{1}{1^{c_1}c_1!2^{c_2}c_2!\cdots} = \frac{1}{\ell!}\left(\frac{\ell!}{1^{c_1}c_1!2^{c_2}c_2!\cdots}\right).$$

By Theorem 7.12 this is just the coefficient of $z_1^{c_1}z_2^{c_2}\cdots$ in $Z_{\mathfrak{S}_\ell}(z_1,z_2,\ldots)$, as was to be proved. □

As a check of Theorem 7.13, set each $z_i = n$ to obtain

$$\begin{aligned} \sum_{\ell\geq 0} Z_{\mathfrak{S}_\ell}(n,n,\ldots)x^\ell &= e^{nx+n\frac{x^2}{2}+n\frac{x^3}{3}+\cdots} \\ &= e^{n(x+\frac{x^2}{2}+\frac{x^3}{3}+\cdots)} \\ &= e^{n\log(1-x)^{-1}} \\ &= \frac{1}{(1-x)^n} \\ &= \sum_{\ell\geq 0}\binom{-n}{\ell}(-x)^\ell \\ &= \sum_{\ell\geq 0}\binom{n+\ell-1}{\ell}x^\ell, \end{aligned}$$

the last step following from the easily checked equality $\binom{-n}{\ell} = (-1)^\ell\binom{n+\ell-1}{\ell}$. Equating coefficients of x^ℓ in the first and last terms of the above string of equalities gives

$$\begin{aligned} Z_{\mathfrak{S}_\ell}(n,n,\ldots) &= \binom{n+\ell-1}{\ell} \\ &= \frac{n(n+1)\cdots(n+\ell-1)}{\ell!}, \end{aligned}$$

agreeing with Theorem 7.5 and equation (42).

Theorem 7.13 has many enumerative applications. We give one such result here as an example.

7.14 Proposition. Let $f(n)$ be the number of permutations $w \in \mathfrak{S}_n$ of odd order. Equivalently, $w^k = 1$ for some odd k. Then

$$f(n) = \begin{cases} 1^2 \cdot 3^2 \cdot 5^2 \cdots (n-1)^2, & n \text{ even} \\ 1^2 \cdot 3^2 \cdot 5^2 \cdots (n-2)^2 \cdot n, & n \text{ odd.} \end{cases}$$

Proof. A permutation has odd order if and only if all its cycle lengths are odd. Hence [why?]

$$f(n) = n! Z_{\mathfrak{S}_n}(z_i = 1,\ i \text{ odd}; z_i = 0,\ i \text{ even}).$$

Making this substitution in Theorem 7.13 gives

$$\sum_{n\geq 0} f(n)\frac{x^n}{n!} = \exp\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots\right).$$

Since $-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$, we get [why?]

$$\begin{aligned} \sum_{n\geq 0} f(n)\frac{x^n}{n!} &= \exp\left(\frac{1}{2}\left(-\log(1-x) + \log(1+x)\right)\right) \\ &= \exp\frac{1}{2}\log\left(\frac{1+x}{1-x}\right) \\ &= \sqrt{\frac{1+x}{1-x}}. \end{aligned}$$

We therefore need to find the coefficients in the power series expansion of $\sqrt{(1+x)/(1-x)}$ at $x = 0$. There is a simple trick for doing so:

$$\begin{aligned} \sqrt{\frac{1+x}{1-x}} &= (1+x)(1-x^2)^{-1/2} \\ &= (1+x)\sum_{m\geq 0}\binom{-1/2}{m}(-x^2)^m \\ &= \sum_{m\geq 0}(-1)^m\binom{-1/2}{m}(x^{2m} + x^{2m+1}), \end{aligned}$$

where by definition

$$\binom{-1/2}{m} = \frac{1}{m!}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2m-1}{2}\right).$$

It is now a routine computation to check that the coefficient of $x^n/n!$ in $\sqrt{(1+x)/(1-x)}$ agrees with the desired value of $f(n)$. $\square$

Quotients of boolean algebra. We will show how to apply Pólya theory to the problem of counting the number of elements of given rank in a quotient poset B_X/G. Here X is a finite set, B_X is the boolean algebra of all subsets of X, and G is a group of permutations of X (with an induced action on B_X). What do colorings of X have to do with subsets? The answer is very simple: A 2-coloring $f : X \to \{0,1\}$ corresponds to a subset S_f of X by the usual rule

$$s \in S_f \Longleftrightarrow f(s) = 1.$$

Note that two 2-colorings f and g are G-equivalent if and only if S_f and S_g are in the same orbit of G (acting on B_X). Thus the number of inequivalent 2-colorings f of X with i values equal to 1 is just $\#(B_X/G)_i$, the number of elements of B_X/G of rank i. As an immediate application of Pólya's theorem (Theorem 7.7) we obtain the following result.

7.15 Corollary. *We have*

$$\sum_i \#(B_X/G)_i q^i = Z_G(1+q, 1+q^2, 1+q^3, \ldots).$$

Proof. If $\kappa(i,j)$ denotes the number of inequivalent 2-colorings of X with the colors 0 and 1 such that 0 is used j times and 1 is used i times (so $i + j = \#X$), then by Pólya's theorem we have

$$\sum_{i,j} \kappa(i,j) x^i y^j = Z_G(x+y, x^2+y^2, x^3+y^3, \ldots).$$

Setting $x = q$ and $y = 1$ yields the desired result [why?]. $\square$

Combining Corollary 7.15 with the rank-unimodality of B_X/G (Theorem 5.9) yields the following corollary.

7.16 Corollary. *For any finite group G of permutations of a finite set X, the polynomial $Z_G(1+q, 1+q^2, 1+q^3, \ldots)$ has symmetric, unimodal, integer coefficients.*

7.17 Example. (a) For the poset P of Example 5.5(a) we have $G = \{(1)(2)(3), (1,2)(3)\}$, so $Z_G(z_1, z_2, z_3) = \frac{1}{2}(z_1^3 + z_1 z_2)$. Hence

$$\begin{aligned} \sum_{i=0}^{3} (\#P_i) q^i &= \frac{1}{2}\left((1+q)^3 + (1+q)(1+q^2)\right) \\ &= 1 + 2q + 2q^2 + q^3. \end{aligned}$$

(b) For the poset P of Example 5.5(b) we have $G = \{(1)(2)(3)(4)(5), (1,2,3,4,5), (1,3,5,2,4), (1,4,2,5,3), (1,5,4,3,2)\}$, so $Z_G(z_1, z_2, z_3, z_4, z_5) = \frac{1}{5}(z_1^5 + 4z_5)$. Hence

$$\begin{aligned} \sum_{i=0}^{5} (\#P_i) q^i &= \frac{1}{5}\left((1+q)^5 + 4(1+q^5)\right) \\ &= 1 + q + 2q^2 + 2q^3 + q^4 + q^5. \end{aligned}$$

(c) Let X be the squares of a 2×2 chessboard, labelled as follows:

1	2
3	4

Let G be the wreath product $\mathfrak{S}_2 \wr \mathfrak{S}_2$, as defined in Section 6. Then

$$\begin{aligned} G = \{&(1)(2)(3)(4), (1,2)(3)(4), (1)(2)(3,4), (1,2)(3,4), \\ &(1,3)(2,4), (1,4)(2,3), (1,3,2,4), (1,4,2,3)\}, \end{aligned}$$

so

$$Z_G(z_1, z_2, z_3, z_4) = \frac{1}{8}(z_1^4 + 2z_1^2 z_2 + 3z_2^2 + 2z_4).$$

Hence

$$\begin{aligned} \sum_{i=0}^{4} (\#P_i) q^i &= \frac{1}{4}\left((1+q)^4 + 2(1+q)^2(1+q^2) + 3(1+q^2)^2 + 2(1+q^4)\right) \\ &= 1 + q + 2q^2 + q^3 + q^4 \\ &= \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \end{aligned}$$

agreeing with Theorem 6.6.

Using more sophisticated methods (such as the representation theory of the symmetric group), the following generalization of Corollary 7.16 can be proved: Let $P(q)$ be any polynomial with symmetric, unimodal, nonnegative, integer coefficients, such as $1+q+3q^2+3q^3+8q^4+3q^5+3q^6+q^7+q^8$ or q^5+q^6 $(=0+0q+\cdots+0q^4+q^5+q^6+0q^7+\cdots+0q^{11})$. Then the polynomial $Z_G(P(q),P(q^2),P(q^3),\ldots)$ has symmetric, unimodal, nonnegative, integer coefficients.

Graphs. A standard application of Pólya theory is to the enumeration of nonisomorphic graphs. We saw at the end of Section 5 that if M is an m-element vertex set, $X=\binom{M}{2}$, and $\mathfrak{S}_m^{(2)}$ is the group of permutations of X induced by permutations of M, then an orbit of i-element subsets of X may be regarded as an isomorphism class of graphs on the vertex set M with i-edges. Thus $\#(B_X/\mathfrak{S}_m^{(2)})_i$ is the number of nonisomorphic graphs (without loops or multiple edges) on the vertex set M with i edges. It follows from Corollary 7.15 that if $g_i(m)$ denotes the number of nonisomorphic graphs with m vertices and i edges, then

$$\sum_{i=0}^{\binom{m}{2}} g_i(m)q^i = Z_{\mathfrak{S}_m^{(2)}}(1+q,1+q^2,1+q^3,\ldots).$$

Thus we would like to compute the cycle enumerator $Z_{\mathfrak{S}_m^{(2)}}(z_1,z_2,\ldots)$. If two permutations π and σ of M have the same cycle type (number of cycles of each length), then their actions on X also have the same cycle type [why?]. Thus for each possible cycle type of a permutation of M (i.e., for each partition of m) we need to compute the induced cycle type on X. We also know from Theorem 7.12 the number of permutations of M of each type. For small values of m we can pick some permutation π of each type and compute directly its action on X in order to determine the induced cycle type. For $m=4$ we have:

CYCLE			INDUCED	CYCLE
LENGTHS			PERMUTATION	LENGTHS
OF π	NUMBER	π	π'	OF π'
$1,1,1,1$	1	$(1)(2)(3)(4)$	$(12)(13)(14)(23)24)(34)$	$1,1,1,1,1,1$
$2,1,1$	6	$(1,2)(3)(4)$	$(12)(12,23)(14,24)(34)$	$2,2,1,1$
$3,1$	8	$(1,2,3)(4)$	$(12,23,13)(14,24,34)$	$3,3$
$2,2$	3	$(1,2)(3,4)$	$(12)(13,24)(14,23)(34)$	$2,2,1,1$
4	6	$(1,2,3,4)$	$(12,23,34,14)(13,24)$	$4,2$

It follows that

$$Z_{\mathfrak{S}_4^{(2)}}(z_1,z_2,z_3,z_4,z_5,z_6)=\frac{1}{24}(z_1^6+9z_1^2z_2^2+8z_3^2+6z_2z_4).$$

If we set $z_i = 1+q^i$ and simplify, we obtain the polynomial

$$\sum_{i=0}^{6} g_i(4)q^i = 1+q+2q^2+3q^3+2q^4+q^5+q^6.$$

Suppose that we instead wanted to count the number $h_i(4)$ of nonisomorphic graphs with four vertices and i edges, where now we allow at most *two* edges between any two vertices. We can take M, X, and $G=\mathfrak{S}_4^{(2)}$ as before, but now we have three colors: red for no edges, blue for one edge, and yellow for two edges. A monomial $r^ib^jy^k$ corresponds to a coloring with i pairs of vertices having no edges between them, j pairs having one edge, and k pairs having two edges. The total number e of edges is $j+2k$. Hence if we let $r=1, b=q, y=q^2$, then the monomial $r^ib^jy^k$ becomes $q^{j+2k}=q^e$. It follows that

$$\begin{aligned}\sum_{i=0}^{i(i-1)} h_i(4)q^i &= Z_{\mathfrak{S}_4^{(2)}}(1+q+q^2, 1+q^2+q^4, 1+q^3+q^6,\ldots)\\ &= \frac{1}{24}\big((1+q+q^2)^6+9(1+q+q^2)^2(1+q^2+q^4)^2\\ &\qquad +8(1+q^3+q^6)^2+6(1+q^2+q^4)(1+q^4+q^8)\big)\\ &= 1+q+3q^2+5q^3+8q^4+9q^5+12q^6+9q^7+8q^8+5q^9\\ &\qquad +3q^{10}+q^{11}+q^{12}.\end{aligned}$$

The total number of nonisomorphic graphs on four vertices with edge multiplicities at most two is $\sum_i h_i(4)=66$.

It should now be clear that if we restrict the edge multiplicity to be r, then the corresponding generating function is $Z_{\mathfrak{S}_4^{(2)}}(1+q+q^2+\cdots+q^{r-1}, 1+q^2+q^4+\cdots+q^{2r-2}, \ldots)$. In particular, to obtain the *total* number $N(r,4)$ of nonisomorphic graphs on four vertices with edge multiplicity at most r, we simply set each $z_i = r$, obtaining

$$\begin{aligned} N(r,4) &= Z_{\mathfrak{S}_4^{(2)}}(r,r,r,r,r,r) \\ &= \frac{1}{24}(r^6+9r^4+14r^2). \end{aligned}$$

This is the same as number of inequivalent r-colorings of the set $X = \binom{M}{2}$ (where $\#M = 4$) [why?].

Of course the same sort of reasoning can be applied to any number of vertices. For five vertices our table becomes the following (using such notation as 1^5 to denote a sequence of five 1's).

CYCLE			INDUCED	CYCLE
LENGTHS			PERMUTATION	LENGTHS
OF π	NO.	π	π'	OF π'
1^5	1	$(1)(2)(3)(4)(5)$	$(12)(13)\cdots(45)$	1^{10}
$2,1^3$	10	$(1,2)(3)(4)(5)$	$(12)(13,23)(14,25)(15,25)(34)(35)(45)$	$2^3,1^4$
$3,1^2$	20	$(1,2,3)(4)(5)$	$(12,23,13)(14,24,34)(15,25,35)(45)$	$3^3,1$
$2^2,1$	15	$(1,2)(3,4)(5)$	$(12)(13,24)(14,23)(15,25)(34)(35,45)$	$2^4,1^2$
$4,1$	30	$(1,2,3,4)(5)$	$(12,23,34,14)(13,24)(15,25,35,45)$	$4^2,2$
$3,2$	20	$(1,2,3)(4,5)$	$(12,23,13)(14,25,34,15,24,35)(45)$	$6,3,1$
5	24	$(1,2,3,4,5)$	$(12,23,34,45,15)(13,24,35,14,25)$	5^2

Thus

$$Z_{\mathfrak{S}_5^{(2)}}(z_1,\ldots,z_{10}) = \frac{1}{120}(z_1^{10}+10z_1^4z_2^3+20z_1z_3^3+15z_1^2z_2^4+30z_2z_4^2+20z_1z_3z_6+24z_5^2),$$

from which we compute

$$\begin{aligned} \sum_{i=0}^{10} g_i(5)q^i &= Z_{\mathfrak{S}_5^{(2)}}(1+q, 1+q^2, \ldots, 1+q^{10}) \\ &= 1+q+2q^2+4q^3+6q^4+6q^5+6q^6+4q^7+2q^8+q^9+q^{10}. \end{aligned}$$

For an arbitrary number $m = \#M$ of vertices there exist explicit formulas for the cycle indicator of the induced action of $\pi \in \mathfrak{S}_M$ on $\binom{M}{2}$, thereby obviating the need to compute π' explicitly as we did in the above tables, but the overall expression for $Z_{\mathfrak{S}_m^{(2)}}$ cannot be simplified significantly or put into a simple generating function as we did in Theorem 7.13. For reference we record

$$\begin{aligned} Z_{\mathfrak{S}_6^{(2)}} &= \frac{1}{6!}(z_1^{15} + 15z_1^7z_2^4 + 40z_1^3z_3^4 + 45z_1^3z_2^6 + 90z_1z_2z_4^3 + 120z_1z_2z_3^2z_6 \\ &\qquad +144z_5^3 + 15z_1^3z_2^6 + 90z_1z_2z_4^3 + 40z_3^5 + 120z_3z_6^2) \end{aligned}$$

$$(g_0(6), g_1(6), \ldots, g_{15}(6)) = (1, 1, 2, 5, 9, 15, 21, 24, 24, 21, 15, 9, 5, 2, 1, 1).$$

Moreover if $u(n)$ denotes the number of nonisomorphic simple graphs with n vertices, then

$$\begin{gathered} (u(0), u(1), \ldots, u(11)) \\ = (1, 1, 2, 4, 11, 34, 156, 1044, 12346, 274668, 12005168, 1018997864). \end{gathered}$$

A table of $u(n)$ for $n \leq 75$ is given at

`http://www.research.att.com/∼njas/sequences/b000088.txt`

In particular,

$$\begin{aligned} u(75) = \;& 91965776790545918117055311393231179873443957239 \\ & 0555232344598910500368551136102062542965342147 \\ & 8723210428876893185920222186100317580740213865 \\ & 7140377683043095632048495393006440764501648363 \\ & 4760490012493552274952950606265577383468983364 \\ & 6883724923654397496226869104105041619919159586 \\ & 8518775275216748149124234654756641508154401414 \\ & 8480274454866344981385848105320672784068407907 \\ & 1134767688676890584660201791139593590722767979 \\ & 8617445756819562952590259920801220117529208077 \\ & 0705444809177422214784902579514964768094933848 \\ & 3173060596932480677345855848701061537676603425 \\ & 1254842843718829212212327337499413913712750831 \end{aligned}$$

$$\begin{gathered}0550986833980707875560051306072520155744624852\\0263616216031346723897074759199703968653839368\\77636080643275926566803872596099072,\end{gathered}$$

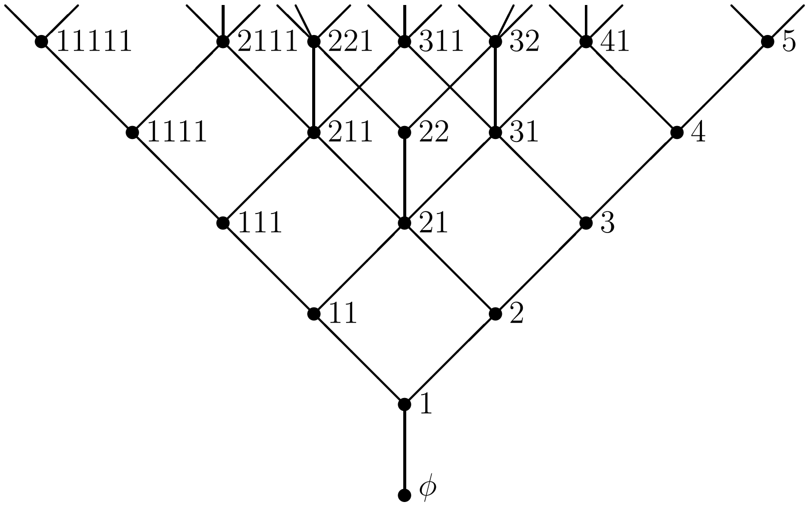
a number of 726 digits! Compare

$$\frac{2^{\binom{75}{2}}}{75!} = .91965776790545918 09 \times 10^{726},$$

which agrees with $u(75)$ to 17 significant digits [why?].

8 A glimpse of Young tableaux.

We defined in Section 6 Young's lattice Y, the poset of all partitions of all nonnegative integers, ordered by containment of their Young diagrams.



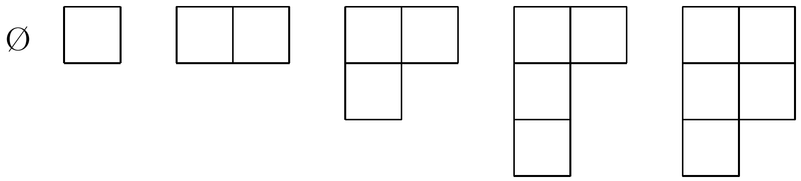
Young's lattice

Here we will be concerned with the counting of certain walks in the Hasse diagram (considered as a graph) of Y. Note that since Y is infinite, we cannot talk about its eigenvalues and eigenvectors. We need different techniques for counting walks. (It will be convenient to denote the length of a walk by n, rather than by ℓ as in previous sections.)

Note that Y is a graded poset (of infinite rank), with Y_i consisting of all partitions of i. In other words, we have $Y = Y_0 \cup Y_1 \cup \cdots$ (disjoint union), where every maximal chain intersects each level Y_i exactly once. We call Y_i the ith *level* of Y.

Since the Hasse diagram of Y is a simple graph (no loops or multiple edges), a walk of length n is specified by a sequence $\lambda^0, \lambda^1, \ldots, \lambda^n$ of vertices of Y. We will call a walk in the Hasse diagram of a poset a *Hasse walk.* Each λ^i is a partition of some integer, and we have either (a) $\lambda^i < \lambda^{i+1}$ and $|\lambda^i| = |\lambda^{i+1}| - 1$, or (b) $\lambda^i > \lambda^{i+1}$ and $|\lambda^i| = |\lambda^{i+1}| + 1$. A step of type (a) is denoted by U (for "up," since we move up in the Hasse diagram), while a step of type (b) is denoted by D (for "down"). If the walk W has steps of types $A_1, A_2, \dots, A_n$, respectively, where each A_i is either U or D, then we say that W is of *type* $A_n A_{n-1} \cdots A_2 A_1$. Note that the type of a walk is written in the *opposite* order to that of the walk. This is because we will soon regard U and D as linear transformations, and we multiply linear transformations *right-to-left* (opposite to the usual left-to-right reading order). For instance (abbreviating a partition $(\lambda_1, \dots, \lambda_m)$ as $\lambda_1 \cdots \lambda_m$), the walk $\varnothing, 1, 2, 1, 11, 111, 211, 221, 22, 21, 31, 41$ is of type $UUDDUUUUDUU = U^2D^2U^4DU^2$.

There is a nice combinatorial interpretation of walks of type U^n which begin at $\varnothing$. Such walks are of course just saturated chains $\varnothing = \lambda^0 < \lambda^1 < \cdots < \lambda^n$. In other words, they may be regarded as sequences of Young diagrams, beginning with the empty diagram and adding one new square at each step. An example of a walk of type U^5 is given by



We can specify this walk by taking the final diagram and inserting an i into square s if s was added at the ith step. Thus the above walk is encoded by the "tableau"

1	2
3	5
4	

Such an object τ is called a *standard Young tableaux* (or SYT). It consists of the Young diagram D of some partition λ of an integer n, together with the numbers $1, 2, \dots, n$ inserted into the squares of D, so that each number appears exactly once, and every row and column is *increasing.* We call λ the *shape* of the SYT τ, denoted $\lambda = \mathrm{sh}(\tau)$. For instance, there are five SYT of shape $(2,2,1)$, given by

$$\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 \\ \cline{1-1}\end{array}\qquad\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 5 \\ \hline 4 \\ \cline{1-1}\end{array}\qquad\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 4 \\ \hline 5 \\ \cline{1-1}\end{array}\qquad\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 5 \\ \hline 4 \\ \cline{1-1}\end{array}\qquad\begin{array}{|c|c|}\hline 1 & 4 \\ \hline 2 & 5 \\ \hline 3 \\ \cline{1-1}\end{array}$$

Let f^λ denote the number of SYT of shape λ, so for instance $f^{(2,2,1)} = 5$. The numbers f^λ have many interesting properties; for instance, there is a famous explicit formula for them known as the Frame-Robinson-Thrall hook formula. For the sake of completeness we state this formula without proof, though it is not needed in what follows.

Let u be a square of the Young diagram of the partition λ. Define the *hook* $H(u)$ of u (or at u) to be the set of all squares directly to the right of u or directly below u, including u itself. The size (number of squares) of $H(u)$ is called the *hook length* of u (or at u), denoted $h(u)$. In the diagram of the partition $(4,2,2)$ below, we have inserted the hook length $h(u)$ inside each square u.

$$\begin{array}{|c|c|c|c|}\hline 6 & 5 & 2 & 1 \\ \hline 3 & 2 \\ \cline{1-2} 2 & 1 \\ \cline{1-2}\end{array}$$

8.1 Theorem. (hook formula) *Let $\lambda \vdash n$. Then*

$$f^\lambda = \frac{n!}{\prod_{u\in\lambda} h(u)}.$$

Here the notation $u \in \lambda$ means that u ranges over all squares of the Young diagram of λ.

For instance, the diagram of the hook lengths of $\lambda = (4,2,2)$ above gives

$$f^\lambda = \frac{8!}{6\cdot 5\cdot 2\cdot 1\cdot 3\cdot 2\cdot 2\cdot 1} = 56.$$

In this section we will be concerned with the connection between SYT and counting walks in Young's lattice. If $w = A_n A_{n-1} \cdots A_1$ is some word in U and D and $\lambda \vdash n$, then let us write $\alpha(w, \lambda)$ for the number of Hasse walks in Y of type w which start at the empty partition $\emptyset$ and end at λ. For instance, $\alpha(UDUU, 11) = 2$, the corresponding walks being $\emptyset, 1, 2, 1, 11$ and $\emptyset, 1, 11, 1, 11$. Thus in particular $\alpha(U^n, \lambda) = f^\lambda$ [why?]. In a similar fashion, since the number of Hasse walks of type $D^n U^n$ which begin at $\emptyset$, go up to a partition $\lambda \vdash n$, and then back down to $\emptyset$ is given by $(f^\lambda)^2$, we have

$$\alpha(D^n U^n, \emptyset) = \sum_{\lambda \vdash n} (f^\lambda)^2. \tag{43}$$

Our object is to find an explicit formula for $\alpha(w, \lambda)$ of the form $f^\lambda c_w$, where c_w does not depend on λ. (It is by no means *a priori* obvious that such a formula should exist.) In particular, since $f^\emptyset = 1$, we will obtain by setting $\lambda = \emptyset$ a simple formula for the number of (closed) Hasse walks of type w from $\emptyset$ to $\emptyset$ (thus including a simple formula for (43)).

There is an easy condition for the existence of *any* Hasse walks of type w from $\emptyset$ to λ, given by the next lemma.

8.2 Lemma. *Suppose $w = D^{s_k} U^{r_k} \cdots D^{s_2} U^{r_2} D^{s_1} U^{r_1}$, where $r_i \geq 0$ and $s_i \geq 0$. Let $\lambda \vdash n$. Then there exists a Hasse walk of type w from $\emptyset$ to λ if and only if:*

$$\sum_{i=1}^{k} (r_i - s_i) = n$$

$$\sum_{i=1}^{j} (r_i - s_i) \geq 0 \text{ for } 1 \leq j \leq k.$$

Proof. Since each U moves up one level and each D moves down one level, we see that $\sum_{i=1}^{k}(r_i - s_i)$ is the level at which a walk of type w beginning at $\emptyset$ ends. Hence $\sum_{i=1}^{k}(r_i - s_i) = |\lambda| = n$.

After $\sum_{i=1}^{j}(r_i + s_i)$ steps we will be at level $\sum_{i=1}^{j}(r_i - s_i)$. Since the lowest level is level 0, we must have $\sum_{i=1}^{j}(r_i - s_i) \geq 0$ for $1 \leq j \leq k$.

The easy proof that the two conditions of the lemma are *sufficient* for the existence of a Hasse walk of type w from $\emptyset$ to λ is left to the reader. $\Box$

If w is a word in U and D satisfying the conditions of Lemma 8.2, then we say that w is a *valid* λ*-word.* (Note that the condition of being a valid λ-word depends only on $|\lambda|$.)

The proof of our formula for $\alpha(w, \lambda)$ will be based on linear transformations analogous to those defined by (18) and (19). As in Section 4 let $\mathbb{R}Y_j$ be the real vector space with basis Y_j. Define two linear transformations $U_i : \mathbb{R}Y_i \to \mathbb{R}Y_{i+1}$ and $D_i : \mathbb{R}Y_i \to \mathbb{R}Y_{i-1}$ by

$$U_i(\lambda) = \sum_{\substack{\mu \vdash i+1 \\ \lambda < \mu}} \mu$$

$$D_i(\lambda) = \sum_{\substack{\nu \vdash i-1 \\ \nu < \lambda}} \nu,$$

for all $\lambda \vdash i$. For instance (using abbreviated notation for partitions)

$$U_{21}(54422211) = 64422211 + 55422211 + 54432211 + 54422221 + 544222111$$

$$D_{21}(54422211) = 44422211 + 54322211 + 54422111 + 5442221.$$

It is clear [why?] that if r is the number of *distinct* (i.e., unequal) parts of λ, then $U_i(\lambda)$ is a sum of $r + 1$ terms and $D_i(\lambda)$ is a sum of r terms. The next lemma is an analogue for Y of the corresponding result for B_n (Lemma 4.6).

8.3 Lemma. *For any $i \geq 0$ we have*

$$D_{i+1}U_i - U_{i-1}D_i = I_i, \tag{44}$$

the identity linear transformation on $\mathbb{R}Y_i$.

Proof. Apply the left-hand side of (44) to a partition λ of i, expand in terms of the basis Y_i, and consider the coefficient of a partition μ. If $\mu \neq \lambda$ and μ can be obtained from λ by adding one square s to (the Young diagram of) λ and then removing a (necessarily different) square t, then there is exactly one choice of s and t. Hence the coefficient of μ in $D_{i+1}U_i(\lambda)$ is equal to 1. But then there is exactly one way to remove a square from λ and then add a square to get μ, namely, remove t and add s. Hence the coefficient of μ in $U_{i-1}D_i(\lambda)$ is also 1, so the coefficient of μ when the left-hand side of (44) is applied to λ is 0.

If now $\mu \neq \lambda$ and we cannot obtain μ by adding a square and then deleting a square from λ (i.e., μ and λ differ in more than two rows), then clearly when we apply the left-hand side of (44) to λ, the coefficient of μ will be 0.

Finally consider the case $\lambda = \mu$. Let r be the number of distinct (unequal) parts of λ. Then the coefficient of λ in $D_{i+1}U_i(\lambda)$ is $r+1$, while the coefficient of λ in $U_{i-1}D_i(\lambda)$ is r, since there are $r+1$ ways to add a square to λ and then remove it, while there are r ways to remove a square and then add it back in. Hence when we apply the left-hand side of (44) to λ, the coefficient of λ is equal to 1.

Combining the conclusions of the three cases just considered shows that the left-hand side of (44) is just I_i, as was to be proved. □

We come to one of the main results of this section.

8.4 Theorem. *Let λ be a partition and $w = A_nA_{n-1}\cdots A_1$ a valid λ-word. Let $S_w = \{i : A_i = D\}$. For each $i \in S_w$, let a_i be the number of D's in w to the right of A_i, and let b_i be the number of U's in w to the right of A_i. Then*

$$\alpha(w,\lambda) = f^{\lambda} \prod_{i\in S_w} (b_i - a_i).$$

Before proving Theorem 8.4, let us give an example. Suppose $w = U^3D^2U^2DU^3 = UUUDDUUDUUU$ and $\lambda = (2,2,1)$. Then $S_w = \{4,7,8\}$ and $a_4 = 0$, $b_4 = 3$, $a_7 = 1$, $b_7 = 5$, $a_8 = 2$, $b_8 = 5$. We have also seen earlier that $f^{221} = 5$. Thus

$$\alpha(w,\lambda) = 5(3-0)(5-1)(5-2) = 180.$$

Proof of Theorem 8.4. For notational simplicity we will omit the subscripts from the linear transformations U_i and D_i. This should cause no confusion since the subscripts will be uniquely determined by the elements on which U and D act. For instance, the expression $UDUU(\lambda)$ where $\lambda \vdash i$ must mean $U_{i+1}D_{i+2}U_{i+1}U_i(\lambda)$; otherwise it would be undefined since U_j and D_j can only act on elements of $\mathbb{R}Y_j$, and moreover U_j raises the level by one while D_j lowers it by one.

By (44) we can replace DU in any word y in the letters U and D by $UD + I$. This replaces y by a sum of two words, one with one fewer D and the other with one D moved one space to the right. For instance, replacing the first DU in $UUDUDDU$ by $UD + I$ yields $UUUDDDU + UUDDU$. If we begin with the word w and iterate this procedure, replacing a DU in any word with $UD+I$, eventually there will be no U's to the right of any D's and the procedure will come to an end. At this point we will have expressed w as a linear combination (with integer coefficients) of words of the form U^iD^j. Since the operation of replacing DU with $UD + I$ preserves the difference between the number of U's and D's in each word, all the words U^iD^j which appear will have $i - j$ equal to some constant n (namely, the number of U's minus the number of D's in w). Specifically, say we have

$$w = \sum_{i-j=n} r_{ij}(w)U^iD^j, \tag{45}$$

where each $r_{ij}(w) \in \mathbb{Z}$. (We also define $r_{ij}(w) = 0$ if $i < 0$ or $j < 0$.) We claim that the $r_{ij}(w)$'s are uniquely determined by w. Equivalently [why?], if we have

$$\sum_{i-j=n} d_{ij}U^iD^j = 0 \tag{46}$$

(as an identity of linear transformations acting on the space $\mathbb{R}Y_k$ for *any* k), where each $d_{ij} \in \mathbb{Z}$ (or $d_{ij} \in \mathbb{R}$, if you prefer), then each $d_{ij} = 0$. Let j' be the least integer for which $d_{j'+n,j'} \neq 0$. Let $\mu \vdash j'$, and apply both sides of (46) to μ. The left-hand side has exactly one nonzero term, namely, the term with $j = j'$ [why?]. The right-hand side, on the other hand[1], is 0, a contradiction. Thus the $r_{ij}(w)$'s are unique.

[1] The phrase "the right-hand side, on the other hand" does not mean the left-hand side!

Now apply U on the left to (45). We get

$$Uw = \sum_{i,j} r_{ij}(w) U^{i+1} D^j.$$

Hence (using uniqueness of the r_{ij}'s) there follows [why?]

$$r_{ij}(Uw) = r_{i-1,j}(w). \tag{47}$$

We next want to apply D on the left to (45). It is easily proved by induction on i (left as an exercise) that

$$DU^i = U^i D + iU^{i-1}. \tag{48}$$

(We interpret U^{-1} as being 0, so that (48) is true for $i = 0$.) Hence

$$\begin{aligned} Dw &= \sum_{i,j} r_{ij}(w) DU^i D^j \\ &= \sum_{i,j} r_{ij}(w)(U^i D + iU^{i-1}) D^j, \end{aligned}$$

from which it follows [why?] that

$$r_{ij}(Dw) = r_{i,j-1}(w) + (i+1) r_{i+1,j}(w). \tag{49}$$

Setting $j = 0$ in (47) and (49) yields

$$r_{i0}(Uw) = r_{i-1,0}(w) \tag{50}$$

$$r_{i0}(Dw) = (i+1) r_{i+1,0}(w). \tag{51}$$

Now let (45) operate on $\emptyset$. Since $D^j(\emptyset) = 0$ for all $j > 0$, we get $w(\emptyset) = r_{n0}(w) U^n(\emptyset)$. Thus the coefficient of λ in $w(\emptyset)$ is given by

$$\alpha(w, \lambda) = r_{n0}(w) \alpha(U^n, \lambda) = r_{n0} f^\lambda,$$

where as usual $\lambda \vdash n$. It is easy to see from (50) and (51) that

$$r_{n0}(w) = \prod_{j \in S_w} (b_j - a_j),$$

and the proof follows. □

An interesting special case of the previous theorem allows us to evaluate equation (43).

8.5 Corollary. *We have*

$$\alpha(D^nU^n, \emptyset) = \sum_{\lambda \vdash n} (f^\lambda)^2 = n!$$

Proof. When $w = D^nU^n$ in Theorem 8.4 we have $S_w = \{n+1, n+2, \ldots, 2n\}$, $a_i = n - i$, and $b_i = n$, from which the proof is immediate. □

Note (for those familiar with the representation theory of finite groups). It can be shown that the numbers f^λ, for $\lambda \vdash n$, are the degrees of the irreducible representations of the symmetric group $\mathcal{S}_n$. Given this, Corollary 8.5 is a special case of the result that the sum of the squares of the degrees of the irreducible representations of a finite group G is equal to the order $|G|$ of G. There are many other intimate connections between the representation theory of $\mathcal{S}_n$, on the one hand, and the combinatorics of Young's lattice and Young tableaux, on the other. There is also an elegant combinatorial proof of Corollary 8.5, based on the *RSK algorithm* (after Gilbert de Beauregard Robinson, Craige Schensted, and Donald Knuth) or *Robinson-Schensted correspondence*, with many fascinating properties and with deep connections with representation theory. In the Appendix to this section we give the definition of the RSK algorithm.

We now consider a variation of Theorem 8.4 in which we are not concerned with the type w of a Hasse walk from $\emptyset$ to w, but only with the number of steps. For instance, there are three Hasse walks of length three from $\emptyset$ to the partition 1, given by $\emptyset, 1, \emptyset, 1$; $\emptyset, 1, 2, 1$; and $\emptyset, 1, 11, 1$. Let $\beta(\ell, \lambda)$ denote the number of Hasse walks of length ℓ from $\emptyset$ to λ. Note the two following easy facts:

(F1) $\beta(\ell, \lambda) = 0$ unless $\ell \equiv |\lambda| \pmod 2$.

(F2) $\beta(\ell, \lambda)$ is the coefficient of λ in the expansion of $(D+U)^\ell(\emptyset)$ as a linear combination of partitions.

Because of (F2) it is important to write $(D+U)^\ell$ as a linear combination of terms U^iD^j, just as in the proof of Theorem 8.4 we wrote a word w in U and D in this form. Thus define integers $b_{ij}(\ell)$ by

$$(D+U)^\ell = \sum_{i,j} b_{ij}(\ell)U^iD^j. \tag{52}$$

Just as in the proof of Theorem 8.4, the numbers $b_{ij}(\ell)$ exist and are well-defined.

8.6 Lemma. *We have $b_{ij}(\ell)=0$ if $\ell-i-j$ is odd. If $\ell-i-j=2m$ then*

$$b_{ij}(\ell) = \frac{\ell!}{2^m\, i!\, j!\, m!}. \tag{53}$$

Proof. The assertion for $\ell-i-j$ odd is equivalent to (F1) above, so assume $\ell-i-j$ is even. The proof is by induction on ℓ. It's easy to check that (53) holds for $\ell=1$. Now assume true for some fixed $\ell\geq 1$. Using (52) we obtain

$$\begin{aligned}\sum_{i,j} b_{ij}(\ell+1)U^iD^j &= (D+U)^{\ell+1}\\ &= (D+U)\sum_{i,j} b_{ij}(\ell)U^iD^j\\ &= \sum_{i,j} b_{ij}(\ell)(DU^iD^j+U^{i+1}D^j).\end{aligned}$$

In the proof of Theorem 8.4 we saw that $DU^i=U^iD+iU^{i-1}$ (see equation (48)). Hence we get

$$\sum_{i,j} b_{ij}(\ell+1)U^iD^j = \sum_{i,j} b_{ij}(\ell)(U^iD^{j+1}+iU^{i-1}D^j+U^{i+1}D^j). \tag{54}$$

As mentioned after (52), the expansion of $(D+U)^{\ell+1}$ in terms of U^iD^j is unique. Hence equating coefficients of U^iD^j on both sides of (54) yields the recurrence

$$b_{ij}(\ell+1) = b_{i,j-1}(\ell)+(i+1)b_{i+1,j}(\ell)+b_{i-1,j}(\ell). \tag{55}$$

It is a routine matter to check that the function $\ell!/2^m i!j!m!$ satisfies the same recurrence (55) as $b_{ij}(\ell)$, with the same intial condition $b_{00}(0) = 1$. From this the proof follows by induction. □

From Lemma 8.6 it is easy to prove the following result.

8.7 Theorem. *Let $\ell \geq n$ and $\lambda \vdash n$, with $\ell - n$ even. Then*

$$\beta(\ell, \lambda) = \binom{\ell}{n}(1 \cdot 3 \cdot 5 \cdots (\ell - n - 1))f^\lambda.$$

Proof. Apply both sides of (52) to Ø. Since $U^i D^j(\text{Ø}) = 0$ unless $j = 0$, we get

$$\begin{aligned} (D+U)^\ell(\text{Ø}) &= \sum_i b_{i0}(\ell)U^i(\text{Ø}) \\ &= \sum_i b_{i0}(\ell)\sum_{\lambda \vdash i} f^\lambda \lambda. \end{aligned}$$

Since by Lemma 8.6 we have $b_{i0}(\ell) = \binom{\ell}{i}(1 \cdot 3 \cdot 5 \cdots (\ell - i - 1))$ when $\ell - i$ is even, the proof follows from (F2). □

Note. The proof of Theorem 8.7 only required knowing the value of $b_{i0}(\ell)$. However, in Lemma 8.6 we computed $b_{ij}(\ell)$ for all j. We could have carried out the proof so as only to compute $b_{i0}(\ell)$, but the general value of $b_{ij}(\ell)$ is so simple that we have included it too.

8.8 Corollary. *The total number of Hasse walks in Y of length $2m$ from Ø to Ø is given by*

$$\beta(2m, \text{Ø}) = 1 \cdot 3 \cdot 5 \cdots (2m - 1).$$

Proof. Simply substitute $\lambda = \text{Ø}$ (so $n = 0$) and $\ell = 2m$ in Theorem 8.7. □

The fact that we can count various kinds of Hasse walks in Y suggests that there may be some finite graphs related to Y whose eigenvalues we can also compute. This is indeed the case, and we will discuss the simplest case here. Let $Y_{j-1,j}$ denote the restriction of Young's lattice Y to ranks $j-1$ and j. Identify $Y_{j-1,j}$ with its Hasse diagram, regarded as a (bipartite) graph. Let $p(i) = |Y_i|$, the number of partitions of i. (The function $p(i)$ has been extensively studied, beginning with Euler, though we will not discuss its fascinating properties here.)

8.9 Theorem. *The eigenvalues of $Y_{j-1,j}$ are given as follows: 0 is an eigenvalue of multiplicity $p(j) - p(j-1)$; and for $1 \leq s \leq j$, the numbers $\pm\sqrt{s}$ are eigenvalues of multiplicity $p(j-s) - p(j-s-1)$.*

Proof. Let $\boldsymbol{A}$ denote the adjacency matrix of $Y_{j-1,j}$. Since $\mathbb{R}Y_{j-1,j} = \mathbb{R}Y_{j-1} \oplus \mathbb{R}Y_j$ (vector space direct sum), any vector $v \in \mathbb{R}Y_{j-1,j}$ can be written uniquely as $v = v_{j-1} + v_j$, where $v_i \in \mathbb{R}Y_i$. The matrix $\boldsymbol{A}$ acts on the vector space $\mathbb{R}Y_{j-1,j}$ as follows [why?]:

$$\boldsymbol{A}(v) = D(v_j) + U(v_{j-1}). \tag{56}$$

Just as Theorem 4.7 followed from Lemma 4.6, we deduce from Lemma 8.3 that for any i we have that $U_i : \mathbb{R}Y_i \to \mathbb{R}Y_{i+1}$ is one-to-one and $D_i : \mathbb{R}Y_i \to \mathbb{R}Y_{i-1}$ is onto. It follows in particular that

$$\begin{aligned} \dim(\ker(D_i)) &= \dim \mathbb{R}Y_i - \dim \mathbb{R}Y_{i-1} \\ &= p(i) - p(i-1), \end{aligned}$$

where ker denotes kernel.

Case 1. Let $v \in \ker(D_j)$, so $v = v_j$. Then $\boldsymbol{A}v = Dv = 0$. Thus $\ker(D_j)$ is an eigenspace of $\boldsymbol{A}$ for the eigenvalue 0, so 0 is an eigenvalue of multiplicity at least $p(j) - p(j-1)$.

Case 2. Let $v \in \ker(D_s)$ for some $0 \leq s \leq j-1$. Let

$$v^* = \pm\sqrt{j-s}U^{j-1-s}(v) + U^{j-s}(v).$$

Note that $v^* \in \mathbb{R}Y_{j-1,j}$, with $v^*_{j-1} = \pm\sqrt{j-s}U^{j-1-s}(v)$ and $v^*_j = U^{j-s}(v)$. Using equation (48), we compute

$$\boldsymbol{A}(v^*) \;=\; U(v^*_{j-1}) + D(v^*_j)$$

$$\begin{aligned}
&= \pm\sqrt{j-s}U^{j-s}(v) + DU^{j-s}(v) \\
&= \pm\sqrt{j-s}U^{j-s}(v) + U^{j-s}D(v) + (j-s)U^{j-s-1}(v) \\
&= \pm\sqrt{j-s}U^{j-s}(v) + (j-s)U^{j-s-1}(v) \\
&= \pm\left(\sqrt{j-s}\right)v^*. \qquad (57)
\end{aligned}$$

It's easy to verify (using the fact that U is one-to-one) that if $v(1), \dots, v(t)$ is a basis for $\ker(D_s)$, then $v(1)^*, \dots, v(t)^*$ are linearly independent. Hence by (57) we have that $\pm\sqrt{j-s}$ is an eigenvalue of $\boldsymbol{A}$ of multiplicity at least $t = \dim(\ker(D_s)) = p(s) - p(s-1)$.

We have found a total of

$$p(j) - p(j-1) + 2\sum_{s=0}^{j-1}(p(s) - p(s-1)) = p(j-1) + p(j)$$

eigenvalues of $\boldsymbol{A}$. (The factor 2 above arises from the fact that both $+\sqrt{j-s}$ and $-\sqrt{j-s}$ are eigenvalues.) Since the graph $Y_{j-1,j}$ has $p(j-1)+p(j)$ vertices, we have found all its eigenvalues. □

An elegant combinatorial consequence of Theorem 8.9 is the following.

8.10 Corollary. *Fix $j \geq 1$. The number of ways to choose a partition λ of j, then delete a square from λ (keeping it a partition), then insert a square, then delete a square, etc., for a total of m insertions and m deletions, ending back at λ, is given by*

$$\sum_{s=1}^{j}[p(j-s) - p(j-s-1)]s^m, \ m > 0. \qquad (58)$$

Proof. Exactly half the closed walks in $Y_{j-1,j}$ of length $2m$ begin at an element of Y_j [why?]. Hence if $Y_{j-1,j}$ has eigenvalues $\theta_1, \dots, \theta_r$, then by Corollary 1.3 the desired number of walks is given by $\frac{1}{2}(\theta_1^{2m} + \cdots + \theta_r^{2m})$. Using the values of $\theta_1, \dots, \theta_r$ given by Theorem 8.9 yields (58). □

For instance, when $j = 7$, equation (58) becomes $4 + 2 \cdot 2^m + 2 \cdot 3^m +$

$4^m + 5^m + 7^m$. When $m = 1$ we get 30, the number of edges of the graph $Y_{6,7}$ [why?].

APPENDIX: THE RSK ALGORITHM

We will describe a bijection between permutations $w \in \mathfrak{S}_n$ and pairs (P, Q) of SYT of the same shape $\lambda \vdash n$. Define a *near Young tableau* (NYT) to be the same as an SYT, except that the entries can be any distinct integers, not necessarily the integers $1, 2, \ldots, n$. Let P_{ij} denote the entry in row i and column j of P. The basic operation of the RSK-algorithm consists of the *row insertion* $P \leftarrow k$ of a positive integer k into an NYT $P = (P_{ij})$. The operation $P \leftarrow k$ is defined as follows: Let r be the largest integer such that $P_{1,r-1} < k$. (If $P_{11} > k$ then let $r = 1$.) If P_{1r} doesn't exist (i.e., P has $r-1$ columns), then simply place k at the end of the first row. The insertion process stops, and the resulting NYT is $P \leftarrow k$. If, on the other hand, P has at least r columns so that P_{1r} exists, then replace P_{1r} by k. The element k then "bumps" $P_{1r} := k'$ into the second row, i.e., insert k' into the second row of P by the insertion rule just described. Continue until an element is inserted at the end of a row (possibly as the first element of a new row). The resulting array is $P \leftarrow k$.

8.11 Example. Let

$$P = \begin{array}{cccc} 3 & 7 & 9 & 14 \\ 6 & 11 & 12 & \\ 10 & 16 & & \\ 13 & & & \\ 15 & & & \end{array}$$

Then $P \leftarrow 8$ is shown below, with the elements inserted into each row (either by bumping or by the final insertion in the fourth row) in boldface. Thus the 8 bumps the 9, the 9 bumps the 11, the 11 bumps the 16, and the 16 is inserted at the end of a row. Hence

$$(P \leftarrow 8) = \begin{array}{cccc} 3 & 7 & \mathbf{8} & 14 \\ 6 & \mathbf{9} & 12 & \\ 10 & \mathbf{11} & & \\ 13 & \mathbf{16} & & \\ 15. & & & \end{array}$$

We omit the proof, which is fairly straightforward, that if P is an NYT, then so is $P \leftarrow k$. We can now describe the RSK algorithm. Let $w =$

$a_1a_2\cdots a_n \in \mathfrak{S}_n$. We will inductively construct a sequence (P_0, Q_0), (P_1, Q_1), $\dots, (P_n, Q_n)$ of pairs (P_i, Q_i) of NYT of the same shape, where P_i and Q_i each have i squares. First, define $(P_0, Q_0) = (\emptyset, \emptyset)$. If (P_{i-1}, Q_{i-1}) have been defined, then set $P_i = P_{i-1} \leftarrow a_i$. In other words, P_i is obtained from P_{i-1} by row inserting a_i. Now define Q_i to be the NYT obtained from Q_{i-1} by inserting i so that Q_i and P_i have the same shape. (The entries of Q_{i-1} don't change; we are simply placing i into a certain new square and not row-inserting it into Q_{i-1}.) Finally let $(P, Q) = (P_n, Q_n)$.

8.12 Example. Let $w = 4273615 \in \mathfrak{S}_7$. The pairs $(P_1, Q_1), \dots, (P_7, Q_7) = (P, Q)$ are as follows:

$\underline{P_i}$	$\underline{Q_i}$
4	1
2 4	1 2
2 7 4	1 3 2
2 3 4 7	1 3 2 4
2 3 6 4 7	1 3 5 2 4

1 3 6
2 7
4
	1 3 5
	2 4
	6

1 3 5
2 6
4 7
	1 3 5
	2 4
	6 7

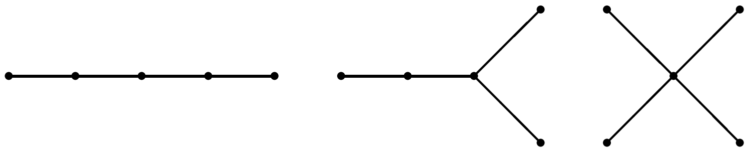
8.13 Theorem. The RSK algorithm defines a bijection between the symmetric group $\mathfrak{S}_n$ and the set of all pairs (P, Q) of SYT of the same shape, where the shape λ is a partition of n.

Proof (sketch). The key step is to define the inverse of RSK. In other words, if $w \mapsto (P, Q)$, then how can we recover w uniquely from (P, Q)? Moreover, we need to find w for *any* (P, Q). Observe that the position occupied by n in Q is the last position to be occupied in the insertion process. Suppose that k occupies this position in P. It was bumped into this position by some element j in the row above k that is currently the largest element of its row less than k. Hence we can "inverse bump" k into the position occupied by j, and now inverse bump j into the row above it by the same procedure. Eventually an element will be placed in the first row, inverse bumping another element t out of the tableau altogether. Thus t was the last element of w to be inserted, i.e., if $w = a_1 a_2 \cdots a_n$ then $a_n = t$. Now locate the position occupied by $n-1$ in Q and repeat the procedure, obtaining a_{n-1}. Continuing in this way, we uniquely construct w one element at a time from right-to-left, such that $w \mapsto (P, Q)$. $\square$

THE MATRIX-TREE THEOREM AND RELATED RESULTS

9 The Matrix-Tree Theorem.

The Matrix-Tree Theorem is a formula for the number of spanning trees of a graph in terms of the determinant of a certain matrix. We begin with the necessary graph-theoretical background. Let G be a finite graph, allowing multiple edges but not loops. (Loops could be allowed, but they turn out to be completely irrelevant.) We say that G is *connected* if there exists a walk between any two vertices of G. A *cycle* is a closed walk with no repeated vertices or edges, except for the the first and last vertex. A *tree* is a connected graph with no cycles. In particular, a tree cannot have multiple edges, since a double edge is equivalent to a cycle of length two. The three nonisomorphic trees with five vertices are given by:



A basic theorem of graph theory (whose easy proof we leave as an exercise) is the following.

9.1 Proposition. *Let G be a graph with p vertices. The following conditions are equivalent.*

(a) *G is a tree.*

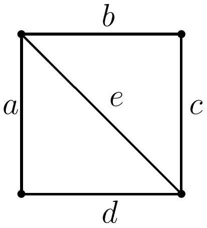
(b) *G is connected and has $p-1$ edges.*

(c) *G is has no cycles and has $p-1$ edges.*

(d) *There is a unique path (= walk with no repeated vertices) between any two vertices.*

A *spanning subgraph* of a graph G is a graph H with the same vertex set as G, and such that every edge of H is an edge of G. If G has q edges, then the number of spanning subgraphs of G is equal to 2^q, since we can choose any subset of the edges of G to be the set of edges of H. (Note that multiple edges between the same two vertices are regarded as *distinguishable*, in accordance with the definition of a graph in Section 1.) A spanning subgraph which is a tree is called a *spanning tree*. Clearly G has a spanning tree if and only if it is connected [why?]. An important invariant of a graph G is its number of spanning trees, called the *complexity* of G and denoted $\kappa(G)$.

9.2 Example. Let G be the graph illustrated below, with edges a, b, c, d, e.



Then G has eight spanning trees, namely, *abc*, *abd*, *acd*, *bcd*, *abe*, *ace*, *bde*, and *cde* (where, e.g., *abc* denotes the spanning subgraph with edge set $\{a, b, c\}$).

9.3 Example. Let $G = K_5$, the complete graph on five vertices. A simple counting argument shows that K_5 has 60 spanning trees isomorphic to the first tree in the above illustration of all nonisomorphic trees with five vertices, 60 isomorphic to the second tree, and 5 isomorphic to the third tree. Hence $\kappa(K_5) = 125$. It is even easier to verify that $\kappa(K_1) = 1$, $\kappa(K_2) = 1$, $\kappa(K_3) = 3$, and $\kappa(K_4) = 16$. Can the reader make a conjecture about the value of $\kappa(K_p)$ for any $p \geq 1$?

Our object is to obtain a "determinantal formula" for $\kappa(G)$. For this we need an important result from matrix theory which is often omitted from a beginning linear algebra course. (Later (Theorem 10.4) we will prove a more general determinantal formula without the use of the Binet-Cauchy theorem. However, the use of the Binet-Cauchy theorem does afford some additional algebraic insight.) This result, known as the Binet-Cauchy theorem (or sometimes as the Cauchy-Binet theorem), is a generalization of the familiar fact that if A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$ (where det denotes determinant). We want to extend this formula to the case where A and B are rectangular matrices whose product is a square matrix (so that $\det(AB)$ is defined). In other words, A will be an $m \times n$ matrix and B an $n \times m$ matrix, for some $m, n \geq 1$.

We will use the following notation involving submatrices. Suppose $A = (a_{ij})$ is an $m \times n$ matrix, with $1 \leq i \leq m$, $1 \leq j \leq n$, and $m \leq n$. Given an m-element subset S of $\{1, 2, \ldots, n\}$, let $A[S]$ denote the $m \times m$ submatrix of A obtained by taking the columns indexed by the elements of S. In other words, if the elements of S are given by $j_1 < j_2 < \cdots < j_m$, then $A[S] = (a_{i,j_k})$, where $1 \leq i \leq m$ and $1 \leq k \leq m$. For instance, if

$$A = \left[\begin{array}{rrrrr} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{array} \right]$$

and $S = \{2, 3, 5\}$, then

$$A[S] = \left[\begin{array}{rrr} 2 & 3 & 5 \\ 7 & 8 & 10 \\ 12 & 13 & 15 \end{array} \right].$$

Similarly, let $B = (b_{ij})$ be an $n \times m$ matrix with $1 \leq i \leq n$, $1 \leq j \leq m$ and $m \leq n$. Let S be an m-element subset of $\{1, 2, \ldots, n\}$ as above. Then $B[S]$ denotes the $m \times m$ matrix obtained by taking the *rows* of B indexed by S. Note that $A^t[S] = A[S]^t$, where ^{t} denotes transpose.

9.4 Theorem. (the Binet-Cauchy Theorem) *Let $A = (a_{ij})$ be an $m \times n$ matrix, with $1 \leq i \leq m$ and $1 \leq j \leq n$. Let $B = (b_{ij})$ be an $n \times m$ matrix with $1 \leq i \leq n$ and $1 \leq j \leq m$. (Thus AB is an $m \times m$ matrix.) If $m > n$, then* $\det(AB) = 0$. *If $m \leq n$, then*

$$\det(AB) = \sum_S (\det A[S])(\det B[S]),$$

where S ranges over all m-element subsets of $\{1, 2, \ldots, n\}$.

Before proceeding to the proof, let us give an example. We write $|a_{ij}|$ for the determinant of the matrix (a_{ij}). Suppose

$$A = \left[\begin{array}{ccc} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{array} \right], \quad B = \left[\begin{array}{cc} c_1 & d_1 \\ c_2 & d_2 \\ c_3 & d_3 \end{array} \right].$$

Then

$$\det(AB) = \left| \begin{array}{cc} a_1 & a_2 \\ b_1 & b_2 \end{array} \right| \cdot \left| \begin{array}{cc} c_1 & d_1 \\ c_2 & d_2 \end{array} \right| + \left| \begin{array}{cc} a_1 & a_3 \\ b_1 & b_3 \end{array} \right| \cdot \left| \begin{array}{cc} c_1 & d_1 \\ c_3 & d_3 \end{array} \right| + \left| \begin{array}{cc} a_2 & a_3 \\ b_2 & b_3 \end{array} \right| \cdot \left| \begin{array}{cc} c_2 & d_2 \\ c_3 & d_3 \end{array} \right|.$$

Proof of Theorem 9.4 (sketch). First suppose $m > n$. Since from linear algebra we know that $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$ and that the rank of an $m \times n$ matrix cannot exceed n (or m), we have that $\operatorname{rank}(AB) \leq n < m$. But AB is an $m \times m$ matrix, so $\det(AB) = 0$, as claimed.

Now assume $m \leq n$. We use notation such as M_{rs} to denote an $r \times s$ matrix M. It is an immediate consequence of the definition of matrix multiplication (which the reader should check) that

$$\left[\begin{array}{cc} R_{mm} & S_{mn} \\ T_{nm} & U_{nn} \end{array} \right] \left[\begin{array}{cc} V_{mn} & W_{mm} \\ X_{nn} & Y_{nm} \end{array} \right] = \left[\begin{array}{cc} RV + SX & RW + SY \\ TV + UX & TW + UY \end{array} \right]. \tag{59}$$

In other words, we can multiply "block" matrices of suitable dimensions as if their entries were numbers. Note that the entries of the right-hand side of (59) all have well-defined dimensions (sizes), e.g., $RV + SX$ is an $m \times n$ matrix since both RV and SX are $m \times n$ matrices.

Now in equation (59) let $R = I_m$ (the $m \times m$ identity matrix), $S = A$, $T = O_{nm}$ (the $n \times m$ matrix of 0's), $U = I_n$, $V = A$, $W = O_{mm}$, $X = -I_n$, and $Y = B$. We get

$$\left[\begin{array}{cc} I_m & A \\ O_{nm} & I_n \end{array} \right] \left[\begin{array}{cc} A & O_{mm} \\ -I_n & B \end{array} \right] = \left[\begin{array}{cc} O_{mn} & AB \\ -I_n & B \end{array} \right]. \tag{60}$$

Take the determinant of both sides of (60). The first matrix on the left-hand side is an upper triangular matrix with 1's on the main diagonal. Hence its determinant is one. Since the determinant of a product of square matrices is the product of the determinants of the factors, we get

$$\left|\begin{array}{cc} A & O_{mm} \\ -I_n & B \end{array}\right| = \left|\begin{array}{cc} O_{mn} & AB \\ -I_n & B \end{array}\right|. \tag{61}$$

It is easy to see [why?] that the determinant on the right-hand side of (61) is equal to $\pm\det(AB)$. So consider the left-hand side. A nonzero term in the expansion of the determinant on the left-hand side is obtained by taking the product (with a certain sign) of $m+n$ nonzero entries, no two in the same row and column (so one in each row and each column). In particular, we must choose m entries from the last m columns. These entries belong to m of the bottom n rows [why?], say rows $m+s_1, m+s_2, \ldots, m+s_m$. Let $S = \{s_1, s_2, \ldots, s_m\} \subseteq \{1, 2, \ldots, n\}$. We must choose $n-m$ further entries from the last n rows, and we have no choice but to choose the -1's in those rows $m+i$ for which $i \notin S$. Thus every term in the expansion of the left-hand side of (61) uses exactly $n-m$ of the -1's in the bottom left block $-I_n$.

What is the contribution to the expansion of the left-hand side of (61) from those terms which use exactly the -1's from rows $m+i$ where $i \notin S$? We obtain this contribution by deleting all rows and columns to which these -1's belong (in other words, delete row $m+i$ and column i whenever $i \in \{1, 2, \ldots, n\} - S$), taking the determinant of the $2m \times 2m$ matrix M_S that remains, and multiplying by an appropriate sign [why?]. But the matrix M_S is in block-diagonal form, with the first block just the matrix $A[S]$ and the second block just $B[S]$. Hence $\det M_S = (\det A[S])(\det B[S])$ [why?]. Taking all possible subsets S gives

$$\det AB = \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ |S|=m}} \pm(\det A[S])(\det B[S]).$$

It is straightforward but somewhat tedious to verify that all the signs are +; we omit the details. This completes the proof. □

In Section 1 we defined the adjacency matrix $\boldsymbol{A}(G)$ of a graph G with vertex set $V = \{v_1, \ldots, v_p\}$ and edge set $E = \{e_1, \ldots, e_q\}$. We now define two related matrices. Assume for simplicity that G has no loops. (This assumption is harmless since loops have no effect on $\kappa(G)$.)

9.5 Definition. Let G be as above. Give G an *orientation* $\mathfrak{o}$, i.e, for every edge e with vertices u, v, choose one of the ordered pairs (u, v) or (v, u). (If we choose (u, v), say, then we think of putting an arrow on e pointing from u to v; and we say that e is directed from u to v, that u is the *initial vertex* and v the *final vertex* of e, etc.)

(a) The *incidence matrix* $\boldsymbol{M}(G)$ of G (with respect to the orientation $\mathfrak{o}$) is the $p \times q$ matrix whose (i, j)-entry $\boldsymbol{M}_{ij}$ is given by

$$\boldsymbol{M}_{ij} = \begin{cases} 1, & \text{if the edge } e_j \text{ has initial vertex } v_i \\ -1, & \text{if the edge } e_j \text{ has final vertex } v_i \\ 0, & \text{otherwise.} \end{cases}$$

(b) The *laplacian matrix* $\boldsymbol{L}(G)$ of G is the $p \times p$ matrix whose (i, j)-entry $\boldsymbol{L}_{ij}$ is given by

$$\boldsymbol{L}_{ij} = \begin{cases} -m_{ij}, & \text{if } i \neq j \text{ and there are } m_{ij} \text{ edges between } v_i \text{ and } v_j \\ \deg(v_i), & \text{if } i = j, \end{cases}$$

where $\deg(v_i)$ is the number of edges incident to v_i. (Thus $\boldsymbol{L}(G)$ is symmetric and does not depend on the orientation $\mathfrak{o}$.)

Note that every column of $\boldsymbol{M}(G)$ contains one 1, one -1, and $q - 2$ 0's; and hence the sum of the entries in each column is 0. Thus all the rows sum to the 0 vector, a linear dependence relation which shows that $\operatorname{rank}(\boldsymbol{M}(G)) < p$. Two further properties of $\boldsymbol{M}(G)$ and $\boldsymbol{L}(G)$ are given by the following lemma.

9.6 Lemma. (a) *We have $\boldsymbol{M}\boldsymbol{M}^t = \boldsymbol{L}$.*

(b) *If G is regular of degree d, then $\boldsymbol{L}(G) = dI - \boldsymbol{A}(G)$, where $\boldsymbol{A}(G)$ denotes the adjacency matrix of G. Hence if G (or $\boldsymbol{A}(G)$) has eigenvalues $\lambda_1, \ldots, \lambda_p$, then $\boldsymbol{L}(G)$ has eigenvalues $d - \lambda_1, \ldots, d - \lambda_p$.*

Proof. (a) This is immediate from the definition of matrix multiplication. Specifically, for $v_i, v_j \in V(G)$ we have

$$(\boldsymbol{M}\boldsymbol{M}^t)_{ij} = \sum_{e_k \in E(G)} \boldsymbol{M}_{ik}\boldsymbol{M}_{jk}.$$

If $i \neq j$, then in order for $\boldsymbol{M}_{ik}\boldsymbol{M}_{jk} \neq 0$, we must have that the edge e_k connects the vertices v_i and v_j. If this is the case, then one of $\boldsymbol{M}_{ik}$ and $\boldsymbol{M}_{jk}$ will be 1 and the other -1 [why?], so their product is always -1. Hence $(\boldsymbol{M}\boldsymbol{M}^t)_{ij} = -m_{ij}$, as claimed.

There remains the case $i = j$. Then $\boldsymbol{M}_{ik}\boldsymbol{M}_{ik}$ will be 1 if e_k is an edge with v_i as one of its vertices and will be 0 otherwise [why?]. So now we get $(\boldsymbol{M}\boldsymbol{M}^t)_{ii} = \deg(v_i)$, as claimed. This proves (a).

(b) Clear by (a), since the diagonal elements of $\boldsymbol{M}\boldsymbol{M}^t$ are all equal to d. □

Now assume that G is connected, and let $\boldsymbol{M}_0(G)$ be $\boldsymbol{M}(G)$ with its last row removed. Thus $\boldsymbol{M}_0(G)$ has $p-1$ rows and q columns. Note that the number of rows is equal to the number of edges in a spanning tree of G. We call $\boldsymbol{M}_0(G)$ the *reduced incidence matrix* of G. The next result tells us the determinants (up to sign) of all $(p-1)\times(p-1)$ submatrices N of $\boldsymbol{M}_0$. Such submatrices are obtained by choosing a set S of $p-1$ edges of G, and taking all columns of $\boldsymbol{M}_0$ indexed by the edges in S. Thus this submatrix is just $\boldsymbol{M}_0[S]$.

9.7 Lemma. *Let S be a set of $p-1$ edges of G. If S does not form the set of edges of a spanning tree, then* $\det \boldsymbol{M}_0[S] = 0$. *If, on the other hand, S is the set of edges of a spanning tree of G, then* $\det \boldsymbol{M}_0[S] = \pm 1$.

Proof. If S is not the set of edges of a spanning tree, then some subset R of S forms the edges of a cycle C in G. Suppose that the cycle C defined by R has edges $f_1, \dots, f_s$ in that order. Multiply the column of $\boldsymbol{M}_0[S]$ indexed by f_j by 1 if in going around C we traverse f_i in the direction of its arrow; otherwise multiply the column by -1. Then add these modified columns. It is easy to see (check a few small examples to convince yourself) that we get the 0 column. Hence the columns of $\boldsymbol{M}_0[S]$ are linearly dependent, so $\det \boldsymbol{M}_0[S] = 0$, as claimed.

Now suppose that S is the set of edges of a spanning tree T. Let e be an edge of T which is connected to v_p (the vertex which indexed the bottom row of $\boldsymbol{M}$, i.e., the row removed to get $\boldsymbol{M}_0$). The column of $\boldsymbol{M}_0[S]$ indexed by e contains exactly one nonzero entry [why?], which is ± 1. Remove from $\boldsymbol{M}_0[S]$

the row and column containing the nonzero entry of column e, obtaining a $(p-2)\times(p-2)$ matrix $\boldsymbol{M}'_0$. Note that $\det(\boldsymbol{M}_0[S]) = \pm\det(\boldsymbol{M}'_0)$ [why?]. Let T' be the tree obtained from T by contracting the edge e to a single vertex (so that v_p and the remaining vertex of e are merged into a single vertex u). Then $\boldsymbol{M}'_0$ is just the matrix obtained from the incidence matrix $\boldsymbol{M}(T')$ by removing the row indexed by u [why?]. Hence by induction on the number p of vertices (the case $p=1$ being trivial), we have $\det(\boldsymbol{M}'_0) = \pm 1$. Thus $\det(\boldsymbol{M}_0[S]) = \pm 1$, and the proof follows. □

NOTE. An alternative way of seeing that $\det(\boldsymbol{M}_0 S) = \pm 1$ when S is the set of of edges of a spanning tree T is as follows. Let $u_1, u_2, \dots, u_{p-1}$ be an ordering of the vertices $v_1, \dots, v_{p-1}$ such that u_i is an endpoint of the tree obtained from T by removing vertices $u_1, \dots, u_{i-1}$. (It is easy to see that such an ordering is possible.) Permute the rows of $\boldsymbol{M}_0[S]$ so that the ith row is indexed by u_i. Then permute the columns in the order $e_1, \dots, e_{p-1}$ so that e_i is the unique edge adjacent to u_i after $u_1, \dots, u_{i-1}$ have been removed. Then we obtain a lower triangular matrix with ±1's on the main diagonal, so the determinant is ±1.

We have now assembled all the ingredients for the main result of this section (due originally to Borchardt). Recall that $\kappa(G)$ denotes the number of spanning trees of G.

9.8 Theorem. (the Matrix-Tree Theorem) *Let G be a finite connected graph without loops, with laplacian matrix $\boldsymbol{L} = \boldsymbol{L}(G)$. Let $\boldsymbol{L}_0$ denote $\boldsymbol{L}$ with the last row and column removed (or with the ith row and column removed for any i). Then*

$$\det(\boldsymbol{L}_0) = \kappa(G).$$

Proof. Since $\boldsymbol{L} = \boldsymbol{M}\boldsymbol{M}^t$ (Lemma 9.6(a)), it follows immediately that $\boldsymbol{L}_0 = \boldsymbol{M}_0\boldsymbol{M}_0^t$. Hence by the Binet-Cauchy theorem (Theorem 9.4), we have

$$\det(\boldsymbol{L}_0) = \sum_S (\det \boldsymbol{M}_0[S])(\det \boldsymbol{M}_0^t[S]), \tag{62}$$

where S ranges over all $(p-1)$-element subsets of $\{1, 2\dots, q\}$ (or equivalently, over all $(p-1)$-element subsets of the set of edges of G). Since in general $A^t[S] = A[S]^t$, equation (62) becomes

$$\det(\boldsymbol{L}_0) = \sum_S (\det \boldsymbol{M}_0[S])^2. \tag{63}$$

According to Lemma 9.7, $\det(\boldsymbol{M}_0[S])$ is ± 1 if S forms the set of edges of a spanning tree of G, and is 0 otherwise. Therefore the term indexed by S in the sum on the right-hand side of (63) is 1 if S forms the set of edges of a spanning tree of G, and is 0 otherwise. Hence the sum is equal to $\kappa(G)$, as desired. □

The operation of removing a row and column from $\boldsymbol{L}(G)$ may seem somewhat contrived. We would prefer a description of $\kappa(G)$ directly in terms of $\boldsymbol{L}(G)$. Such a description will follow from the next lemma.

9.9 Lemma. *Let M be a $p \times p$ matrix (with entries in a field) such that the sum of the entries in every row and column is 0. Let M_0 be the matrix obtained from M by removing the last row and last column (or more generally, any row and any column). Then the coefficient of x in the characteristic polynomial $\det(M - xI)$ of M is equal to $-p \cdot \det(M_0)$. (Moreover, the constant term of $\det(M - xI)$ is 0.)*

Proof. The constant term of $\det(M - xI)$ is $\det(M)$, which is 0 since the rows of M sum to 0.

For simplicity we prove the rest of the lemma only for removing the last row and column, though the proof works just as well for any row and column. Add all the rows of $M - xI$ except the last row to the last row. This doesn't effect the determinant, and will change the entries of the last row all to $-x$ (since the rows of M sum to 0). Factor out $-x$ from the last row, yielding a matrix $N(x)$ satisfying $\det(M - xI) = -x\det(N(x))$. Hence the coefficient of x in $\det(M - xI)$ is given by $-\det(N(0))$. Now add all the columns of $N(0)$ except the last column to the last column. This does not effect $\det(N(0))$. Because the columns of M sum to 0, the last column of $N(0)$ becomes the column vector $[0, 0, \ldots, 0, p]^t$. Expanding the determinant by the last column shows that $\det(N(0)) = p \cdot \det(M_0)$, and the proof follows. □

9.10 Corollary. (a) *Let G be a connected (loopless) graph with p vertices. Suppose that the eigenvalues of $\boldsymbol{L}(G)$ are $\mu_1, \ldots, \mu_{p-1}, \mu_p$, with $\mu_p = 0$. Then*

$$\kappa(G) = \frac{1}{p}\mu_1\mu_2\cdots\mu_{p-1}.$$

(b) *Suppose that G is also regular of degree d, and that the eigenvalues of*

$\boldsymbol{A}(G)$ *are* $\lambda_1, \ldots, \lambda_{p-1}, \lambda_p$, *with* $\lambda_p = d$. *Then*

$$\kappa(G) = \frac{1}{p}(d-\lambda_1)(d-\lambda_2)\cdots(d-\lambda_{p-1}).$$

Proof. (a) We have

$$\begin{aligned} \det(\boldsymbol{L} - xI) &= (\mu_1 - x)\cdots(\mu_{p-1}-x)(\mu_p - x) \\ &= -(\mu_1 - x)(\mu_2 - x)\cdots(\mu_{p-1}-x)x. \end{aligned}$$

Hence the coefficient of x is $-\mu_1\mu_2\cdots\mu_{p-1}$. By Lemma 9.9, we get $-\mu_1\mu_2\cdots\mu_{p-1} = p\cdot\det(\boldsymbol{L}_0)$. By Theorem 9.8 we have $\det(\boldsymbol{L}_0) = \kappa(G)$, and the proof follows.

(b) Immediate from (a) and Lemma 9.6(b). □

Let us look at a couple of examples of the use of the Matrix-Tree Theorem.

9.11 Example. Let $G = K_p$, the complete graph on p vertices. Now K_p is regular of degree $d = p-1$, and by Proposition 1.5 its eigenvalues are -1 ($p-1$ times) and $p-1 = d$. Hence from Corollary 8.10(b) there follows

$$\kappa(K_p) = \frac{1}{p}((p-1)-(-1))^{p-1} = p^{p-2}.$$

This surprising result is often attributed to Cayley, who stated it without proof in 1889 (and even cited Borchardt explicitly). However, it was in fact stated by Sylvester in 1857, while a proof was published by Borchardt in 1860. It is clear that Cayley and Sylvester could have produced a proof if asked to do so. There are many other proofs known, including elegant combinatorial arguments due to Prüfer, Joyal, and others.

9.12 Example. Let $G = C_n$, the n-cube discussed in Section 2. Now C_n is regular of degree n, and by Corollary 2.5 its eigenvalues are $n-2i$ with multiplicity $\binom{n}{i}$ for $0 \leq i \leq n$. Hence from Corollary 8.10(b) there follows the amazing result

$$\begin{aligned} \kappa(C_n) &= \frac{1}{2^n}\prod_{i=1}^{n}(2i)^{\binom{n}{i}} \\ &= 2^{2^n-n-1}\prod_{i=1}^{n} i^{\binom{n}{i}}. \end{aligned}$$

To my knowledge a direct combinatorial proof is not known.

10 Eulerian digraphs and oriented trees.

A famous problem which goes back to Euler asks for what graphs G is there a closed walk which uses every edge exactly once. (There is also a version for non-closed walks.) Such a walk is called an *Eulerian tour* (also known as an *Eulerian cycle*). A graph which has an Eulerian tour is called an *Eulerian graph.* Euler's famous theorem (the first real theorem of graph theory) states that G is Eulerian if and only if it is connected and every vertex has even degree. Here we will be concerned with the analogous theorem for directed graphs. We want to know not just whether an Eulerian tour exists, but how many there are. We will prove an elegant determinantal formula for this number closely related to the Matrix-Tree Theorem. For the case of undirected graphs no analogous formula is known, explaining why we consider only the directed case.

A (finite) *directed graph* or *digraph* D consists of a *vertex set* $V = \{v_1, \dots, v_p\}$ and edge set $E = \{e_1, \dots, e_q\}$, together with a function $\varphi : E \to V \times V$ (the set of ordered pairs (u, v) of elements of V). If $\varphi(e) = (u, v)$, then we think of e as an arrow from u to v. We then call u the *initial vertex* and v the *final vertex* of e. (These concepts arose in the definition of an orientation in Definition 8.5.) A *tour* in D is a sequence $e_1, e_2, \dots, e_r$ of *distinct* edges such that the final vertex of e_i is the initial vertex of e_{i+1} for all $1 \leq i \leq r - 1$, and the final vertex of e_r is the initial vertex of e_1. A tour is *Eulerian* if every edge of D occurs at least once (and hence exactly once). A digraph which has no isolated vertices and contains an Eulerian tour is called an *Eulerian digraph.* Clearly an Eulerian digraph is connected. The *outdegree* of a vertex v, denoted $\text{outdeg}(v)$, is the number of edges of G with initial vertex v. Similarly the *indegree* of v, denoted $\text{indeg}(v)$, is the number of edges of D with final vertex v. A loop (edge of the form (v, v)) contributes one to both the indegree and outdegree. A digraph is *balanced* if $\text{indeg}(v) = \text{outdeg}(v)$ for all vertices v.

10.1 Theorem. *A digraph D is Eulerian if and only if it is connected and balanced.*

Proof. Assume D is Eulerian, and let $e_1, \dots, e_q$ be an Eulerian tour. As we move along the tour, whenever we enter a vertex v we must exit it, except at the very end we enter the final vertex v of e_q without exiting it. However, at the beginning we exited v without having entered it. Hence every vertex is entered as often as it is exited and so must have the same outdegree as indegree. Therefore D is balanced, and as noted above D is clearly connected.

Now assume that D is balanced and connected. We may assume that D has at least one edge. We first claim that for any edge e of D, D has a tour for which $e = e_1$. If e_1 is a loop we are done. Otherwise we have entered the vertex $\mathrm{fin}(e_1)$ for the first time, so since D is balanced there is some exit edge e_2. Either $\mathrm{fin}(e_2) = \mathrm{init}(e_1)$ and we are done, or else we have entered the vertex $\mathrm{fin}(e_2)$ once more than we have exited it. Since D is balanced there is new edge e_3 with $\mathrm{fin}(e_2) = \mathrm{init}(e_3)$. Continuing in this way, either we complete a tour or else we have entered the current vertex once more than we have exited it, in which case we can exit along a new edge. Since D has finitely many edges, eventually we must complete a tour. Thus D does have a tour which uses e_1.

Now let $e_1, \dots, e_r$ be a tour C of maximum length. We must show that $r = q$, the number of edges of D. Assume to the contrary that $r < q$. Since in moving along C every vertex is entered as often as it is exited (with $\mathrm{init}(e_1)$ exited at the beginning and entered at the end), when we remove the edges of C from D we obtain a digraph H which is still balanced, though it need not be connected. However, since D is connected, at least one connected component H_1 of H contains at least one edge and has a vertex v in common with C [why?]. Since H_1 is balanced, there is an edge e of H_1 with initial vertex v. The argument of the previous paragraph shows that H_1 has a tour C' of positive length beginning with the edge e. But then when moving along C, when we reach v we can take the "detour" C' before continuing with C. This gives a tour of length longer than r, a contradiction. Hence $r = q$, and the theorem is proved. □

Our primary goal is to count the number of Eulerian tours of a connected balanced digraph. A key concept in doing so is that of an oriented tree. An *oriented tree* with root v is a (finite) digraph T with v as one of its vertices, such that there is a unique directed path from any vertex u to v. In other words, there is a unique sequence of edges $e_1, \dots, e_r$ such that (a) $\mathrm{init}(e_1) = u$, (b) $\mathrm{fin}(e_r) = v$, and (c) $\mathrm{fin}(e_i) = \mathrm{init}(e_{i+1})$ for $1 \le i \le r - 1$.

It's easy to see that this means that the underlying undirected graph (i.e., "erase" all the arrows from the edges of T) is a tree, and that all arrows in T "point toward" v. There is a surprising connection between Eulerian tours and oriented trees, given by the next result (due to de Bruijn and van Aardenne-Ehrenfest). This result is sometimes called the BEST Theorem, after de **B**ruijn, van Aardenne-**E**hrenfest, **S**mith, and **T**utte. However, Smith and Tutte were not involved in the original discovery.

10.2 Theorem. *Let D be a connected balanced digraph with vertex set V. Fix an edge e of D, and let $v = \mathrm{init}(e)$. Let $\tau(D, v)$ denote the number of oriented (spanning) subtrees of D with root v, and let $\epsilon(D, e)$ denote the number of Eulerian tours of D starting with the edge e. Then*

$$\epsilon(D, e) = \tau(D, v) \prod_{u \in V} (\mathrm{outdeg}(u) - 1)!. \tag{64}$$

Proof. Let $e = e_1, e_2, \ldots, e_q$ be an Eulerian tour E in D. For each vertex $u \neq v$, let $e(u)$ be the "last exit" from u in the tour, i.e., let $e(u) = e_j$ where $\mathrm{init}(e(u)) = u$ and $\mathrm{init}(e_k) \neq u$ for any $k > j$.

Claim #1. The vertices of D, together with the edges $e(u)$ for all vertices $u \neq v$, form an oriented subtree of D with root v.

Proof of Claim #1. This is a straightforward verification. Let T be the spanning subgraph of D with edges $e(u)$, $u \neq v$. Thus if $|V| = p$, then T has p vertices and $p - 1$ edges [why?]. There are three items to check to insure that T is an oriented tree with root v:

(a) T does not have two edges f and f' satisfying $\mathrm{init}(f) = \mathrm{init}(f')$. This is clear since both f and f' can't be last exits from the same vertex.

(b) T does not have an edge f with $\mathrm{init}(f) = v$. This is clear since by definition the edges of T consist only of last exits from vertices other than v, so no edge of T can exit from v.

(c) T does not have a (directed) cycle C. For suppose C were such a cycle. Let f be that edge of C which occurs after all the other edges of C in

> the Eulerian tour E. Let f' be the edge of C satisfying $\text{fin}(f) = \text{init}(f')$ ($= u$, say). We can't have $u = v$ by (b). Thus when we enter u *via* f, we must exit u. We can't exit u *via* f' since f occurs after f' in E. Hence f' is not the last exit from u, contradicting the definition of T.

It's easy to see that conditions (a)–(c) imply that T is an oriented tree with root v, proving the claim.

Claim #2. We claim that the following converse to Claim #1 is true. Given a connected balanced digraph D and a vertex v, let T be an oriented (spanning) subtree of D with root v. Then we can construct an Eulerian tour E as follows. Choose an edge e_1 with $\text{init}(e_1) = v$. Then continue to choose any edge possible to continue the tour, except we never choose an edge f of E unless we have to, i.e., unless it's the only remaining edge exiting the vertex at which we stand. Then we never get stuck until all edges are used, so we have constructed an Eulerian tour E. Moreover, the set of last exits of E from vertices $u \neq v$ of D coincides with the set of edges of the oriented tree T.

Proof of Claim #2. Since D is balanced, the only way to get stuck is to end up at v with no further exits available, but with an edge still unused. Suppose this is the case. At least one unused edge must be a last exit edge, i.e., an edge of T [why?]. Let u be a vertex of T closest to v in T such that the unique edge f of T with $\text{init}(f) = u$ is not in the tour. Let $y = \text{fin}(f)$. Suppose $y \neq v$. Since we enter y as often as we leave it, we don't use the last exit from y. Thus $y = v$. But then we can leave v, a contradiction. This proves Claim #2.

We have shown that every Eulerian tour E beginning with the edge e has associated with it a "last exit" oriented subtree $T = T(E)$ with root $v = \text{init}(e)$. Conversely, given an oriented subtree T with root v, we can obtain all Eulerian tours E beginning with e and satisfying $T = T(E)$ by choosing for each vertex $u \neq v$ the order in which the edges from u, except the edge of T, appear in E; as well as choosing the order in which all the edges from v except for e appear in E. Thus for each vertex u we have $(\text{outdeg}(u) - 1)!$ choices, so for each T we have $\prod_u (\text{outdeg}(u) - 1)!$ choices. Since there are $\tau(G, v)$ choices for T, the proof is complete. $\square$

10.3 Corollary. *Let D be a connected balanced digraph, and let v be a vertex of D. Then the number $\tau(D,v)$ of oriented subtrees with root v is independent of v.*

Proof. Let e be an edge with initial vertex v. By equation (64), we need to show that the number $\epsilon(G,e)$ of Eulerian tours beginning with e is independent of e. But $e_1e_2\cdots e_q$ is an Eulerian tour if and only if $e_ie_{i+1}\cdots e_qe_1e_2\cdots e_{i-1}$ is also an Eulerian tour, and the proof follows [why?]. □

What we obviously need to do next is find a formula for $\tau(G,v)$. Such a formula is due to W. Tutte in 1948. This result is very similar to the Matrix-Tree Theorem, and indeed we will show (Example 10.6) that the Matrix-Tree Theorem is a simple corollary to Theorem 10.4.

10.4 Theorem. *Let D be a loopless connected digraph with vertex set $V=\{v_1,\dots,v_p\}$. Let $\boldsymbol{L}(D)$ be the $p\times p$ matrix defined by*

$$\boldsymbol{L}_{ij}=\begin{cases}-m_{ij}, & \textit{if } i\neq j \textit{ and there are } m_{ij} \textit{ edges with} \\ & \quad \textit{initial vertex } v_i \textit{ and final vertex } v_j \\ \operatorname{outdeg}(v_i), & \textit{if } i=j.\end{cases}$$

(Thus $\boldsymbol{L}$ is the directed analogue of the laplacian matrix of an undirected graph.) Let $\boldsymbol{L}_0$ denote $\boldsymbol{L}$ with the last row and column deleted. Then

$$\det \boldsymbol{L}_0=\tau(D,v_p). \tag{65}$$

Note. If we remove the ith row and column from $\boldsymbol{L}$ instead of the last row and column, then equation (65) still holds with v_p replaced with v_i.

Proof (sketch). Induction on q, the number of edges of D. The fewest number of edges which D can have is $p-1$ (since D is connected). Suppose then that D has $p-1$ edges, so that as an undirected graph D is a tree. If D is not an oriented tree with root v_p, then some vertex $v_i\neq v_p$ of D has outdegree 0 [why?]. Then $\boldsymbol{L}_0$ has a zero row, so $\det\boldsymbol{L}_0=0=\tau(D,v_p)$. If on the other hand D is an oriented tree with root v_p, then an argument like that used to prove Lemma 9.7 (in the case when S is the set of edges of a spanning tree) shows that $\det\boldsymbol{L}_0=1=\tau(D,v_p)$.

Now assume that D has $q > p-1$ edges, and assume the theorem for digraphs with at most $q-1$ edges. We may assume that no edge f of D has initial vertex v, since such an edge belongs to no oriented tree with root v and also makes no contribution to $\boldsymbol{L}_0$. It then follows, since D has at least p edges, that there exists a vertex $u \neq v$ of D of outdegree at least two. Let e be an edge with $\text{init}(e) = u$. Let D_1 be D with the edge e removed. Let D_2 be D with all edges e' removed such that $\text{init}(e) = \text{init}(e')$ and $e' \neq e$. (Note that D_2 is strictly smaller than D since $\text{outdeg}(u) \geq 2$.) By induction, we have $\det \boldsymbol{L}_0(D_1) = \tau(D_1, v_p)$ and $\det \boldsymbol{L}_0(D_2) = \tau(D_2, v_p)$. Clearly $\tau(D, v_p) = \tau(D_1, v_p) + \tau(D_2, v_p)$, since in an oriented tree T with root v_p, there is exactly one edge whose initial vertex coincides with that of e. On the other hand, it follows immediately from the multilinearity of the determinant [why?] that

$$\det \boldsymbol{L}_0(D) = \det \boldsymbol{L}_0(D_1) + \det \boldsymbol{L}_0(D_2).$$

From this the proof follows by induction. □

10.5 Corollary. *Let D be a connected balanced digraph with vertex set $V = \{v_1, \dots, v_p\}$. Let e be an edge of D. Then the number $\epsilon(D, e)$ of Eulerian tours of D with first edge e is given by*

$$\epsilon(D, e) = (\det \boldsymbol{L}_0(D)) \prod_{u \in V} (\text{outdeg}(u) - 1)!.$$

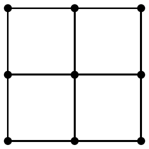
Equivalently (since D is balanced, so Lemma 9.9 applies), if $\boldsymbol{L}(D)$ has eigenvalues $\mu_1, \dots, \mu_p$ with $\mu_p = 0$, then

$$\epsilon(D, e) = \frac{1}{p} \mu_1 \cdots \mu_{p-1} \prod_{u \in V} (\text{outdeg}(u) - 1)!.$$

Proof. Combine Theorems 10.2 and 10.4. □

10.6 Example. (the Matrix-Tree Theorem revisited) Let G be a connected loopless undirected graph. Let $\hat{G}$ be the digraph obtained from G by replacing each edge $e = uv$ of G with a pair of directed edges $u \to v$ and $v \to u$. Clearly $\hat{G}$ is balanced and connected. Choose a vertex v of G. There is an obvious one-to-one correspondence between spanning trees T of G and oriented spanning trees $\hat{T}$ of $\hat{G}$ with root v, namely, direct each edge of T toward v. Moreover, $\boldsymbol{L}(G) = \boldsymbol{L}(\hat{G})$ [why?]. Hence the Matrix-Tree Theorem is an immediate consequence of the Theorem 10.4.

10.7 Example. (the efficient mail carrier) A mail carrier[2] has an itinerary of city blocks to which he (or she) must deliver mail. He wants to accomplish this by walking along each block twice, once in each direction, thus passing along houses on each side of the street. The blocks form the edges of a graph G, whose vertices are the intersections. The mail carrier wants simply to walk along an Eulerian tour in the digraph $\hat{G}$ of the previous example. Making the plausible assumption that the graph is connected, not only does an Eulerian tour always exist, but we can tell the mail carrier how many there are. Thus he will know how many different routes he can take to avoid boredom. For instance, suppose G is the 3×3 grid illustrated below.



This graph has 128 spanning trees. Hence the number of mail carrier routes beginning with a fixed edge (in a given direction) is $128 \cdot 1!^4 2!^4 3! = 12288$. The total number of routes is thus 12288 times twice the number of edges [why?], viz., $12288 \times 24 = 294912$. Assuming the mail carrier delivered mail 250 days a year, it would be 1179 years before he would have to repeat a route!

10.8 Example. (binary de Bruijn sequences) A *binary sequence* is just a sequence of 0's and 1's. A *binary de Bruijn sequence* of degree n is a binary sequence $A = a_1 a_2 \cdots a_{2^n}$ such that every binary sequence $b_1 \cdots b_n$ of length n occurs exactly once as a "circular factor" of A, i.e., as a sequence $a_i a_{i+1} \cdots a_{i+n-1}$, where the subscripts are taken modulo n if necessary. For instance, some circular factors of the sequence *abcdefg* are *a*, *bcde*, *fgab*, and *defga*. Note that there are exactly 2^n binary sequences of length n, so the only possible length of a binary de Bruijn sequence of degree n is 2^n [why?]. Clearly any cyclic shift $a_i a_{i+1} \cdots a_{2^n} a_1 a_2 \cdots a_{i-1}$ of a binary de Bruijn

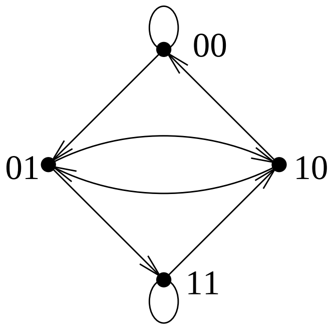
[2] postperson?

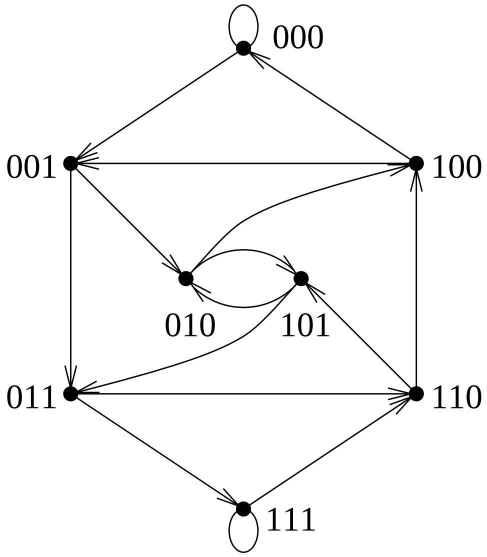
sequence $a_1 a_2 \cdots a_{2^n}$ is also a binary de Bruijn sequence, and we call two such sequences *equivalent.* This relation of equivalence is obviously an equivalence relation, and every equivalence class contains exactly one sequence beginning with n 0's [why?]. Up to equivalence, there is one binary de Bruijn sequence of degree two, namely, 0011. It's easy to check that there are two inequivalent binary de Bruijn sequences of degree three, namely, 00010111 and 00011101. However, it's not clear at this point whether binary de Bruijn sequences exist for all n. By a clever application of Theorems 10.2 and 10.4, we will not only show that such sequences exist for all positive integers n, but we will also count the number of them. It turns out that there are *lots* of them. For instance, the number of inequivalent binary de Bruijn sequences of degree eight is equal to

$$1329227995784915872903807060280344576.$$

The reader with some extra time on his or her hands is invited to write down these sequences. De Bruijn sequences are named after Nicolaas Govert de Bruijn, who published his work on this subject in 1946. However, it was discovered in 1975 that de Bruijn sequences had been earlier created and enumerated by C. Flye Sainte-Marie in 1894. De Bruijn sequences have a number of interesting applications to the design of switching networks and related topics.

Our method of enumerating binary de Bruijn sequences will be to set up a correspondence between them and Eulerian tours in a certain directed graph D_n, the *de Bruijn graph* of degree n. The graph D_n has 2^{n-1} vertices, which we will take to consist of the 2^{n-1} binary sequences of length $n-1$. A pair $(a_1 a_2 \cdots a_{n-1}, b_1 b_2 \cdots b_{n-1})$ of vertices forms an edge of D_n if and only if $a_2 a_3 \cdots a_{n-1} = b_1 b_2 \cdots b_{n-2}$, i.e., e is an edge if the last $n-2$ terms of $\mathrm{init}(e)$ agree with the first $n-2$ terms of $\mathrm{fin}(e)$. Thus every vertex has indegree two and outdegree two [why?], so D_n is balanced. The number of edges of D_n is 2^n. Moreover, it's easy to see that D_n is connected (see Lemma 10.9). The graphs D_3 and D_4 look as follows:





Suppose that $E = e_1 e_2 \cdots e_{2^n}$ is an Eulerian tour in D_n. If $\mathrm{fin}(e_i)$ is the binary sequence $a_{i1}a_{i2}\cdots a_{i,n-1}$, then replace e_i in E by the last bit $a_{i,n-1}$. It is easy to see that the resulting sequence $\beta(E) = a_{1,n-1}a_{2,n-1}\cdots a_{2^n,n-1}$ is a binary de Bruijn sequence, and conversely every binary de Bruijn sequence arises in this way. In particular, since D_n is balanced and connected there exists at least one binary de Bruijn sequence. In order to count the total number of such sequences, we need to compute $\det \boldsymbol{L}(D_n)$. One way to do this is by a clever but messy sequence of elementary row and column operations which transforms the determinant into triangular form. We will give instead an elegant computation of the eigenvalues of $\boldsymbol{L}(D_n)$ based on the following simple lemma.

10.9 Lemma. *Let u and v be any two vertices of D_n. Then there is a unique (directed) walk from u to v of length $n-1$.*

Proof. Suppose $u = a_1 a_2 \cdots a_{n-1}$ and $v = b_1 b_2 \cdots b_{n-1}$. Then the unique path of length $n-1$ from u to v has vertices

$$a_1a_2\cdots a_{n-1}, a_2a_3\cdots a_{n-1}b_1, a_3a_4\cdots a_{n-1}b_1b_2, \ldots,$$

$$a_{n-1}b_1\cdots b_{n-2}, b_1b_2\cdots b_{n-1}. \ \square$$

10.10 Theorem. *The eigenvalues of $\boldsymbol{L}(D_n)$ are* 0 *(with multiplicity* *one) and* 2 *(with multiplicity* $2^{n-1}-1$*).*

Proof. Let $\boldsymbol{A}(D_n)$ denote the directed adjacency matrix of D_n, i.e., the rows and columns are indexed by the vertices, with

$$\boldsymbol{A}_{uv} = \begin{cases} 1, & \text{if } (u,v) \text{ is an edge} \\ 0, & \text{otherwise.} \end{cases}$$

Now Lemma 10.9 is equivalent to the assertion that $\boldsymbol{A}^{n-1} = J$, the $2^{n-1}\times 2^{n-1}$ matrix of all 1's [why?]. If the eigenvalues of $\boldsymbol{A}$ are $\lambda_1, \dots \lambda_{2^{n-1}}$, then the eigenvalues of $J = A^{n-1}$ are $\lambda_1^{n-1}, \dots, \lambda_{2^{n-1}}^{n-1}$. By Lemma 1.4, the eigenvalues of J are 2^{n-1} (once) and 0 ($2^{n-1}-1$ times). Hence the eigenvalues of $\boldsymbol{A}$ are 2ζ (once, where ζ is an $(n-1)$-st root of unity to be determined), and 0 ($2^{n-1}-1$ times). Since the trace of $\boldsymbol{A}$ is 2, it follows that $\zeta = 1$, and we have found all the eigenvalues of A.

Now $\boldsymbol{L}(D_n) = 2I - \boldsymbol{A}(D_n)$ [why?]. Hence the eigenvalues of $\boldsymbol{L}$ are $2 - \lambda_1, \dots, 2 - \lambda_{2^{n-1}}$, and the proof follows from the above determination of $\lambda_1, \dots, \lambda_{2^{n-1}}$. □

10.11 Corollary. *The number* $B_0(n)$ *of binary de Bruijn sequences of degree* n *beginning with* n *0's is equal to* $2^{2^{n-1}-n}$*. The total number* $B(n)$ *of binary de Bruijn sequences of degree* n *is equal to* $2^{2^{n-1}}$*.*

Proof. By the above discussion, $B_0(n)$ is the number of Eulerian tours in D_n whose first edge the loop at vertex $00\cdots 0$. Moreover, the outdegree of every vertex of D_n is two. Hence by Corollary 10.5 and Theorem 10.10 we have

$$B_0(n) = \frac{1}{2^{n-1}} 2^{2^{n-1}-1} = 2^{2^{n-1}-n}.$$

Finally, $B(n)$ is obtained from $B_0(n)$ by multiplying by the number 2^n of edges, and the proof follows. □

Note that the total number of binary sequences of length 2^n is $N = 2^{2^n}$. By the previous corollary, the number of these which are de Bruijn sequences is just $\sqrt{N}$. This suggests the following problem, solved by H. Bidkhori and S. Kishore. Let $\mathcal{A}_n$ be the set of all binary sequences of length 2^n. Let $\mathcal{B}_n$ be the set of binary de Bruijn sequences of degree n. Find an explicit bijection $\varphi : \mathcal{B}_n \times \mathcal{B}_n \to \mathcal{A}_n$, thereby giving a combinatorial proof of Corollary 10.11.

11 Cycles, bonds, and electrical networks.

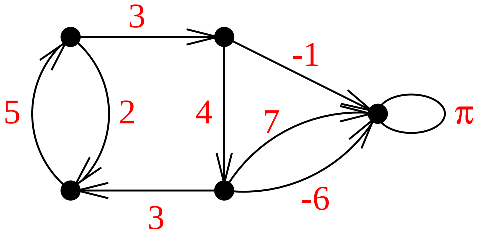
NOTE. This section is in a preliminary form.

11.1 The cycle space and bond space.

In this section we will deal with some interesting linear algebra related to the structure of a directed graph. Let $D = (V, E)$ be a digraph. A function $f : E \to \mathbb{R}$ is called a *circulation* if for every vertex $v \in V$, we have

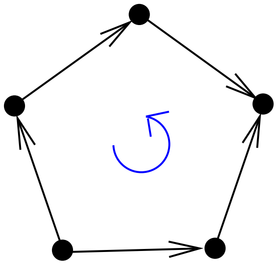
$$\sum_{\substack{e\in E \\ \mathrm{init}(e)=v}} f(e) = \sum_{\substack{e\in E \\ \mathrm{fin}(e)=v}} f(e). \tag{66}$$

Thus if we think of the edges as pipes and f as measuring the flow (quantity per unit of time) of some commodity (such as oil) through the pipe in the specified direction (so that a negative value of $f(e)$ means a flow of $|f(e)|$ in the direction opposite the direction of e), then equation (66) simply says that the amount flowing into each vertex equals the amount flowing out. In other words, the flow is *conservative*. The figure below illustrates a circulation in a digraph D.



Let $\mathcal{C} = \mathcal{C}_D$ denote the set of all circulations on D. Clearly if $f, g \in \mathbb{C}$ and $\alpha, \beta \in \mathbb{R}$ then $\alpha f + \beta g \in \mathcal{C}$. Hence $\mathcal{C}$ is a (real) vector space, called the *cycle space* of D. Thus if $q = |E|$, then $\mathcal{C}_D$ is a subspace of the q-dimensional vector space $\mathbb{R}^E$ of all functions $f : E \to \mathbb{R}$.

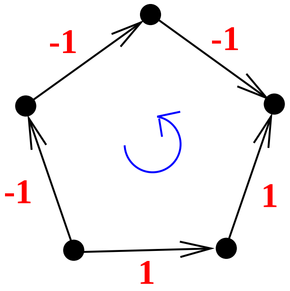
What do circulations have do with something "circulating," and what does the cycle space have to do with actual cycles? To see this, define a *circuit* or *elementary cycle* in D to be a set of edges of a closed walk, *ignoring* *the direction of the arrows*, with no repeated vertices except the first and last. Suppose that C has been assigned an orientation (direction of travel) $\mathfrak{o}$. (Note that this meaning of orientation is not the same as that appearing in Definition 9.5.)



Define a function $f_C : E \to \mathbb{R}$ (which also depends on the orientation $\mathfrak{o}$, though we suppress it from the notation) by

$$f_C(e) = \begin{cases} 1, & \text{if } e \in C \text{ and } e \text{ agrees with } \mathfrak{o} \\ -1, & \text{if } e \in C \text{ and } e \text{ is opposite to } \mathfrak{o} \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that f_C is a circulation. Later we will see that the circulations f_C span the cycle space $\mathcal{C}$, explaining the terminology "circulation" and "cycle space." The figure below shows a circuit C with an orientation $\mathfrak{o}$, and the corresponding circulation f_C.



Given a function $p : V \to \mathbb{R}$, define a new function $\delta p : E \to \mathbb{R}$, called the *coboundary*[3] of p, by

$$\delta p(e) = p(v) - p(u), \quad \text{if } u = \text{init}(e) \text{ and } v = \text{fin}(e).$$

[3] The term "coboundary" arises from algebraic topology, but we will not explain the connection here.

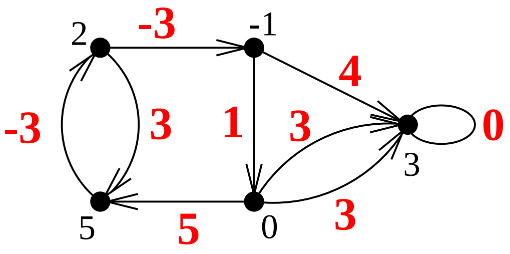


Figure 2: A function and its coboundary

Figure 2 shows a digraph D with the value $p(v)$ of some function $p : V \to \mathbb{R}$ indicated at each vertex v, and the corresponding values $\delta p(e)$ shown at each edge e.

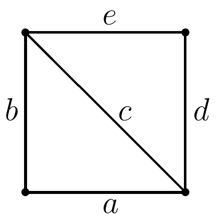
One should regard δ as an operator which takes an element p of the vector space $\mathbb{R}^V$ of all functions $V \to \mathbb{R}$ and produces an element of the vector space $\mathbb{R}^E$ of all functions $E \to \mathbb{R}$. It is immediate from the definition of δ that δ is *linear*, i.e.,

$$\delta(\alpha p + \beta q) = \alpha \cdot \delta p + \beta \cdot \delta q,$$

for all $p, q \in \mathbb{R}^V$ and $\alpha, \beta \in \mathbb{R}$. Thus δ is simply a linear transformation $\delta : \mathbb{R}^V \to \mathbb{R}^E$ between two finite-dimensional vector spaces.

A function $g : E \to \mathbb{R}$ is called a *potential difference* on D if $g = \delta p$ for some $p : V \to \mathbb{R}$. (Later we will see the connection with electrical networks that accounts for the terminology "potential difference.") Let $\mathcal{B} = \mathcal{B}_D$ be the set of all potential differences on D. Thus $\mathcal{B}$ is just the image of the linear transformation δ and is hence a real vector space, called the *bond space* of D.

Let us explain the reason behind the terminology "bond space." A *bond* in a digraph D is a set B of edges such that (a) removing B from D disconnects some (undirected) component of D (that is, removing B creates a digraph which has more connected components, as an undirected graph, than D), and (b) no proper subset of B has this property. A subset of edges satisfying (a) is called a *cutset*, so a bond is just a minimal cutset. Suppose, for example, that D is given as follows (with no arrows drawn since they are irrelevant to the definition of bond):



Then the bonds are the six subsets $ab, de, acd, bce, ace, bcd$.

Let B be a bond. Suppose B disconnects the component (V', E') into two pieces (a bond always disconnects some component into exactly two pieces [why?]) with vertex set S in one piece and $\bar{S}$ in the other. Thus $S \cup \bar{S} = V'$ and $S \cap \bar{S} = \emptyset$. Define

$$[S, \bar{S}] = \{e \in E : \text{exactly one vertex of } e \text{ lies in } S \text{ and one lies in } \bar{S}\}.$$

Clearly $B = [S, \bar{S}]$. It is often convenient to use the notation $[S, \bar{S}]$ for a bond.

Given a bond $B = [S, \bar{S}]$ of D, define a function $g_B : E \to \mathbb{R}$ by

$$g_B(e) = \begin{cases} 1, & \text{if } \mathrm{init}(e) \in \bar{S},\ \mathrm{fin}(e) \in S \\ -1, & \text{if } \mathrm{init}(e) \in S,\ \mathrm{fin}(e) \in \bar{S} \\ 0, & \text{otherwise.} \end{cases}$$

Note that g_B really depends not just on B, but on whether we write B as $[S, \bar{S}]$ or $[\bar{S}, S]$. Writing B in the reverse way simply changes the sign of g_B. Whenever we deal with g_B we will assume that some choice $B = [S, \bar{S}]$ has been made.

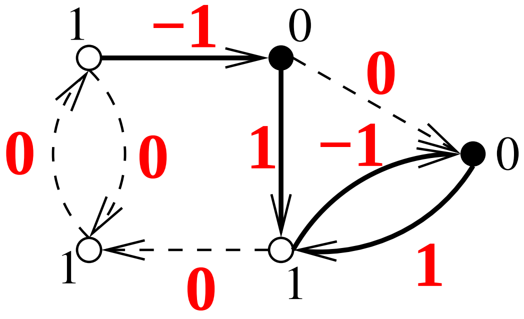
Now note that $g_B = \delta p$, where

$$p(v) = \begin{cases} 1, & \text{if } v \in S \\ 0, & \text{if } v \in \bar{S}. \end{cases}$$

Hence $g_B \in \mathcal{B}$, the bond space of D. We will later see that $\mathcal{B}$ is in fact spanned by the functions g_B, explaining the termininology "bond space."

11.1 Example. In the digraph below, open (white) vertices indicate an element of S and closed (black) vertices an element of $\bar{S}$ for a certain bond

$B = [S, \bar{S}]$. The elements of B are drawn darker than the other edges. The edges are labelled by the values of g_B, and the vertices by the function p for which $g_B = \delta p$.

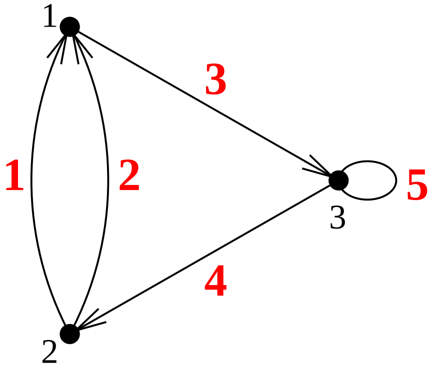


Recall that in Definition 9.5 we defined the incidence matrix $\boldsymbol{M}(G)$ of a loopless undirected graph G with respect to an orientation $\mathfrak{o}$. We may just as well think of G together with its orientation $\mathfrak{o}$ as a directed graph. We also will allow loops. Thus if $D = (V, E)$ is any (finite) digraph, define the *incidence matrix* $\boldsymbol{M} = \boldsymbol{M}(D)$ to be the $p \times q$ matrix whose rows are indexed by V and columns by E, as follows. The entry in row $v \in V$ and column $e \in E$ is denoted $m_v(e)$ and is given by[4]

$$m_v(e) = \begin{cases} -1, & \text{if } v = \operatorname{init}(e) \text{ and } e \text{ is not a loop} \\ 1, & \text{if } v = \operatorname{fin}(e) \text{ and } e \text{ is not a loop} \\ 0, & \text{otherwise.} \end{cases}$$

For instance, if D is given by

[4] Actually, this definition gives the *negative* of the matrix defined in Definition 9.5, though it makes no difference here. We will fix this inconsistency of notation in a later version of these notes.



then

$$\boldsymbol{M}(D) = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{bmatrix}.$$

11.2 Theorem. *The row space of $\boldsymbol{M}(D)$ is the bond space $\mathcal{B}_D$. Equivalently, the functions $m_v : E \to \mathbb{R}$, where v ranges over all vertices of D, span $\mathcal{B}$.*

Proof. Let $g = \delta p$ be a potential difference on D, so

$$\begin{aligned} g(e) &= p(\text{fin}(e)) - p(\text{init}(e)) \\ &= \sum_{v \in V} p(v) m_v(e). \end{aligned}$$

Thus $g = \sum_{v \in V} p(v) m_v$, so g belongs to the row space of $\boldsymbol{M}$.

Conversely, if $g = \sum_{v \in V} q(v) m_v$ is in the row space of $\boldsymbol{M}$, where $q : V \to \mathbb{R}$, then $g = \delta q \in \mathcal{B}$. □

We now define a scalar product (or inner product) on the space $\mathbb{R}^E$ by

$$\langle f, g \rangle = \sum_{e \in E} f(e) g(e),$$

for any $f, g \in \mathbb{R}^E$. If we think of the numbers $f(e)$ and $g(e)$ as the coordinates of f and g with respect to the basis E, then $\langle f, g \rangle$ is just the usual dot product of f and g. Because we have a scalar product, we have a notion of what it means for f and g to be *orthogonal*, viz., $\langle f, g\rangle = 0$. If $\mathcal{V}$ is any subspace of $\mathbb{R}^E$, then define the *orthogonal complement* $\mathcal{V}^\perp$ of $\mathcal{V}$ by

$$\mathcal{V}^\perp = \{f \in \mathbb{R}^E : \langle f, g\rangle = 0 \text{ for all } g \in \mathbb{R}^E\}.$$

Recall from linear algebra that

$$\dim \mathcal{V} + \dim \mathcal{V}^\perp = \dim \mathbb{R}^E = \#E. \qquad (67)$$

Furthermore, $\left(\mathcal{V}^\perp\right)^\perp = \mathcal{V}$.

Intuitively there is a kind of "duality" between elementary cycles and bonds. Cycles "hold vertices together," while bonds "tear them apart." The precise statement of this duality is given by the next result.

11.3 Theorem. *The cycle and bond spaces of D are related by $\mathcal{C} = \mathcal{B}^\perp$. (Equivalently, $\mathcal{B} = \mathcal{C}^\perp$.)*

Proof. Let $f : E \to \mathbb{R}$. Then f is a circulation if and only if

$$\sum_{e\in E} m_v(e) f(e) = 0$$

for all $v \in V$ [why?]. But this is exactly the condition that $f \in \mathcal{B}^\perp$. □

11.2 Bases for the cycle space and bond space.

We want to examine the incidence matrix $\boldsymbol{M}(D)$ in more detail. In particular, we would like to determine which rows and columns of $\boldsymbol{M}(D)$ are linearly independent, and which span the row and column spaces. As a corollary, we will determine the dimension of the spaces $\mathcal{B}$ and $\mathcal{C}$. We begin by defining the *support* $\|f\|$ of $f : E \to \mathbb{R}$ to be the set of edges $e \in E$ for which $f(e) \neq 0$.

11.4 Lemma. *If $0 \neq f \in \mathcal{C}$, then $\|f\|$ contains an undirected circuit.*

Proof. If not, then $\|f\|$ has a vertex of degree one [why?], which is clearly impossible. □

11.5 Lemma. *If $0 \neq g \in \mathcal{B}$, then $\|g\|$ contains a bond.*

Proof. Let $0 \neq g \in \mathcal{B}$, so $g = \delta p$ for some $p : V \to \mathbb{R}$. Choose a vertex v which is incident to an edge of $\|g\|$, and set

$$U = \{u \in V : p(u) = p(v)\}.$$

Let $\bar{U} = V - U$. Note that $\bar{U} \neq \emptyset$, since otherwise p is constant so $g = 0$. Since $g(e) \neq 0$ for all $e \in [U, \bar{U}]$ [why?], we have that $\|g\|$ contains the cutset $[U, \bar{U}]$. Since a bond is by definition a minimal cutset, it follows that $\|g\|$ contains a bond. $\square$

11.6 Definition. A matrix $\boldsymbol{B}$ is called a *basis matrix* of $\mathcal{B}$ if the rows of $\boldsymbol{B}$ form a basis for $\mathcal{B}$. Similary define a basis matrix $\boldsymbol{C}$ of $\mathcal{C}$.

Recall the notation of Theorem 9.4: Let A be a matrix with at least as many columns as rows, whose columns are indexed by the elements of a set T. If $S \subseteq T$, then $A[S]$ denotes the submatrix of A consisting of the columns indexed by the elements of S. In particular, $A[e]$ (short for $A[\{e\}]$) denotes the column of A indexed by e. We come to our first significant result about bases for the vector spaces $\mathcal{B}$ and $\mathcal{C}$.

11.7 Theorem. *Let $\boldsymbol{B}$ be a basis matrix of $\mathcal{B}$, and $\boldsymbol{C}$ a basis matrix of $\mathcal{C}$. (Thus the columns of $\boldsymbol{B}$ and $\boldsymbol{C}$ are indexed by the edges $e \in E$ of D.) Let $S \subseteq E$, Then:*

(i) *The columns of $\boldsymbol{B}[S]$ are linearly independent if and only if S is acyclic (i.e., contains no circuit as an undirected graph).*

(ii) *The columns of $\boldsymbol{C}[S]$ are linearly independent if and only if S contains no bond.*

Proof. The columns of $\boldsymbol{B}[S]$ are linearly dependent if and only if there exists a function $f : E \to \mathbb{R}$ such that

$$f(e) \neq 0 \text{ for some } e \in S$$

$$f(e) = 0 \text{ for all } e \notin S$$

$$\sum_{e\in E} f(e)\boldsymbol{B}[e] = \mathbf{0} \text{ the column vector of 0's.} \tag{68}$$

The last condition is equivalent to $\langle f, m_v\rangle = 0$ for all $v \in V$, i.e., f is a circulation. Thus the columns of $\boldsymbol{B}[S]$ are linearly dependent if and only if there exists a nonzero circulation f such that $\|f\| \subseteq S$. By Lemma 11.4, $\|f\|$ (and therefore S) contains a circuit. Conversely, if S contains a circuit C then $0 \neq f_C \in \mathcal{C}$ and $\|f_C\| = C \subseteq S$, so f_C defines a linear dependence relation (68) among the columns. Hence the columns of $\boldsymbol{B}[S]$ are linearly independent if and only if S is acyclic, proving (i). (Part (i) can also be deduced from Lemma 9.7.)

The proof of (ii) is similar and is left as an exercise. □

11.8 Corollary. *Let $D = (V, E)$ be a digraph with p vertices, q edges, and k connected components (as an undirected graph). Then*

$$\begin{aligned} \dim \mathcal{B} &= p - k \\ \dim \mathcal{C} &= q - p + k. \end{aligned}$$

Proof. For any matrix X, the rank of X is equal to the maximum number of linearly independent columns. Now let $\boldsymbol{B}$ be a basis matrix of $\mathcal{B}$. By Theorem 11.7(i), the rank of $\boldsymbol{B}$ is then the maximum size (number of elements) of an acyclic subset of E. In each connected component D_i of D, the largest acyclic subsets are the spanning trees, whose number of edges is $p(D_i) - 1$, where $p(D_i)$ is the number of vertices of D_i. Hence

$$\begin{aligned} \operatorname{rank} \boldsymbol{B} &= \sum_{i=1}^{k} (p(D_i) - 1) \\ &= p - k. \end{aligned}$$

Since $\dim \mathcal{B} + \dim \mathcal{C} = \dim \mathbb{R}^E = q$ by equation (67) and Theorem 11.3, we have

$$\dim \mathcal{C} = q - (p - k) = q - p + k.$$

(It is also possible to determine $\dim \mathcal{C}$ by a direct argument similar to our determination of $\dim \mathcal{B}$.) □

The number $q - p + k$ (which should be thought of as the number of independent cycles in D) is called the *cyclomatic number* of D (or of its undirected version G, since the direction of the edges have no effect).

Our next goal is to describe explicit bases of $\mathcal{C}$ and $\mathcal{B}$. We begin by defining a *forest* to be an undirected graph without circuits. Thus a forest is a disjoint union of trees. We extend the definition of forest to directed graphs by ignoring the arrows, i.e., a directed graph is a forest if it has no circuits as an undirected graph. Equivalently [why?], $\dim \mathcal{C} = 0$.

Pick a maximal forest T of $D = (V, E)$. Thus T restricted to each component of D is a spanning tree. If e is an edge of D not in T, then it is easy to see that $T \cup e$ contains a unique circuit C_e.

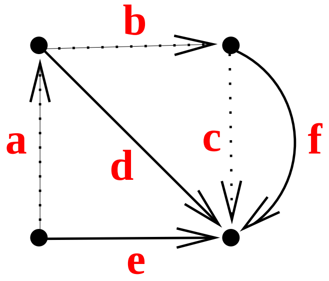
11.9 Theorem. *Let T be as above. Then the set S of circulations f_{C_e}, as e ranges over all edges of D not in T, is a basis for the cycle space $\mathcal{C}$.*

Proof. The circulations f_{C_e} are linearly independent, since for each $e \in E(D) - E(T)$ only f_{C_e} doesn't vanish on e. Moreover,

$$\#S = \#E(D) - \#E(T) = q - p + k = \dim \mathcal{C},$$

so S is a basis. □

11.10 Example. Let D be the digraph shown below, with the edges a, b, c of T shown by dotted lines.



Orient each circuit C_t in the direction of the added edge, i.e., $f_{C_t}(t) = 1$. Then the basis matrix $\boldsymbol{C}$ of $\mathcal{C}$ corresponding to the basis $f_{C_d}, f_{C_e}, f_{C_f}$ is given by

$$\boldsymbol{C} = \left[\begin{array}{rrrrrr} 0 & -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{array} \right]. \qquad (69)$$

We next want to find a basis for the bond space $\mathcal{B}$ analogous to that of Theorem 11.9.

11.11 Lemma. *Let T be a maximal forest of $D = (V, E)$. Let $T^* = D - E(T)$ (the digraph obtained from D by removing the edges of T), called a* cotree *if D is connected. Let e be an edge of T. Then $E(T^*) \cup e$ contains a unique bond.*

Proof. Removing $E(T^*)$ from D leaves a maximal forest T, so removing one further edge e disconnects some component of D. Hence $E(T^*) \cup e$ contains a bond B. It remains to show that B is unique. Removing e from T breaks some component of T into two connected graphs T_1 and T_2 with vertex sets S and $\bar{S}$. It follows [why?] that we must have $B = [S, \bar{S}]$, so B is unique. $\square$

Let T be a maximal forest of the digraph D, and let e be an edge of T. By the previous lemma, $E(T^*) \cup e$ contains a unique bond B_e. Let g_{B_e} be the corresponding element of the bond space $\mathcal{B}$, chosen for definiteness so that $g_{B_e}(e) = 1$.

11.12 Theorem. *The set of functions g_{B_e}, as e ranges over all edges of T, is a basis for the bond space $\mathcal{B}$.*

Proof. The functions g_{B_e} are linearly independent, since only g_{B_e} is nonzero on $e \in E(T)$. Since

$$\#E(T) = p - k = \dim \mathcal{B},$$

it follows that the g_{B_e}'s are a basis for $\mathcal{B}$. $\square$

11.13 Example. Let D and T be as in the previous diagram. Thus a basis for $\mathcal{B}$ is given by the functions $g_{B_a}, g_{B_b}, g_{B_c}$. The corresponding basis matrix is given by

$$\boldsymbol{B} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array} \right].$$

Note that the rows of $\boldsymbol{B}$ are orthogonal to the rows of the matrix $\boldsymbol{C}$ of equation (69), in accordance with Theorem 11.3. Equivalently, $\boldsymbol{B}\boldsymbol{C}^t = \boldsymbol{0}$, the 3×3 zero matrix. (In general, $\boldsymbol{B}\boldsymbol{C}^t$ will have $q - p + k$ rows and $p - k$ columns. Here it is just a coincidence that these two numbers are equal.)

The basis matrices $\boldsymbol{C}_T$ and $\boldsymbol{B}_T$ of $\mathcal{C}$ and $\mathcal{B}$ obtained from a maximal forest T have an important property. A real matrix $m \times n$ matrix A with $m \leq n$ is said to be *unimodular* if every $m \times m$ submatrix has determinant 0, 1, or -1. For instance, the incidence matrix $\boldsymbol{M}(D)$ of a digraph D is unimodular, as proved in Lemma 9.7 (by showing that the expansion of the determinant of a full submatrix has at most one nonzero term).

11.14 Theorem. *Let T be a maximal forest of D. Then the basis matrices $\boldsymbol{C}_T$ of $\mathcal{C}$ and $\boldsymbol{B}_T$ of $\mathcal{B}$ are unimodular.*

Proof. First consider the case $\boldsymbol{C}_T$. Let $\boldsymbol{P}$ be a full submatrix of $\boldsymbol{C}$ (so $\boldsymbol{P}$ has $q-p+k$ rows and columns). Assume $\det \boldsymbol{P} \neq 0$. We need to show $\det \boldsymbol{P} = \pm 1$. Since $\det \boldsymbol{P} \neq 0$, it follows from Theorem 11.7(ii) that $\boldsymbol{P} = \boldsymbol{C}_T[T_1^*]$ for the complement T_1^* of some maximal forest T_1. Note that the rows of the matrix $\boldsymbol{C}_T[T_1^*]$ are indexed by T^* and the columns by T_1^*. Similarly the rows of the basis matrix $\boldsymbol{C}_{T_1}$ are indexed by T_1^* and the columns by E (the set of all edges of D). Hence it makes sense to define the matrix product

$$\boldsymbol{Z} = \boldsymbol{C}_T[T_1^*]\boldsymbol{C}_{T_1},$$

a matrix whose rows are indexed by T^* and columns by E.

Note that the matrix $\boldsymbol{Z}$ is a basis matrix for the cycle space $\mathcal{C}$ since its rows are linear combinations of the rows of the basis matrix $C_{T_1}^*$, and it has full rank since the matrix $\boldsymbol{C}_T[T_1^*]$ is invertible. Now $\boldsymbol{C}_{T_1}[T_1^*] = I_{T_1^*}$ (the identity matrix indexed by T_1^*), so $\boldsymbol{Z}[T_1^*] = \boldsymbol{C}_T[T_1^*]$. Thus $\boldsymbol{Z}$ agrees with the basis matrix $\boldsymbol{C}_T$ in columns T_1^*. Hence the rows of $\boldsymbol{Z} - \boldsymbol{C}_T$ are circulations supported on a subset of T_1. Since T_1 is acyclic, it follows from Lemma 11.4 that the only such circulation is identically 0, so $\boldsymbol{Z} = \boldsymbol{C}_T$.

We have just shown that

$$\boldsymbol{C}_T[T_1^*]\boldsymbol{C}_{T_1} = \boldsymbol{C}_T.$$

Restricting both sides to T^*, we obtain

$$\boldsymbol{C}_T[T_1^*]\boldsymbol{C}_{T_1}[T^*] = \boldsymbol{C}_T[T^*] = I_{T^*}.$$

Taking determinants yields

$$\det(\boldsymbol{C}_T[T_1^*])\det(\boldsymbol{C}_{T_1}[T^*]) = 1.$$

Since all the matrices we have been considering have integer entries, the above determinants are integers. Hence

$$\det \boldsymbol{C}_T[T_1^*] = \pm 1,$$

as was to be proved. (This proof is due to Tutte in 1965.)

A similar proof works for $\boldsymbol{B}_T$. $\square$

11.3 Electrical networks.

We will give a brief indication of the connection between the above discussion and the theory of electrical networks. Let D be a digraph, which for convenience we assume is *connected* and *loopless*. Suppose that at each edge e there is a voltage (potential difference) V_e from $\mathrm{init}(e)$ to $\mathrm{fin}(e)$, and a current I_e in the direction of e (so a negative current I_e indicates a current of $|I_e|$ in the direction opposite to e). Think of V and I as functions on the edges, i.e., as elements of the vector space $\mathbb{R}^E$. There are three fundamental laws relating the quantities V_e and I_e.

Kirchhoff's First Law. $I \in \mathcal{C}_D$. *In other words, the current flowing into a vertex equals the current flowing out. In symbols,*

$$\sum_{\substack{e \\ \mathrm{init}(e)=v}} I_e = \sum_{\substack{e \\ \mathrm{fin}(e)=v}} I_e,$$

for all vertices $v \in V$.

Kirchhoff's Second Law. $V \in \mathcal{C}_D^{\perp} = \mathcal{B}$. *In other words, the sum of the voltages around any circuit (called* loops *by electrical engineers), taking into account orientations, is* 0.

Ohm's Law. *If edge* e *has resistance* $R_e > 0$, *then* $V_e = I_e R_e$.

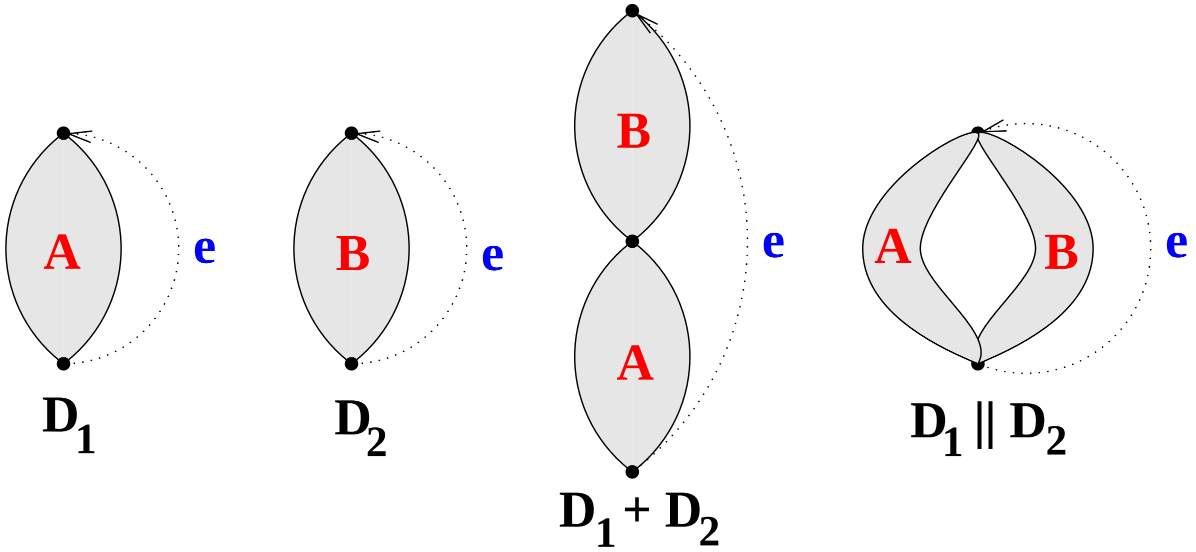
The central problem of electrical network theory dealing with the above three laws[5] is the following: Which of the $3q$ quantities V_e, I_e, R_e need to

[5] Of course the situation becomes much more complicated when one introduces *dynamic* network elements like capacitors, alternating current, etc.

be specified to uniquely determine all the others, and how can we find or stipulate the solution in a fast and elegant way? We will be concerned here only with a special case, perhaps the most important special case in practical applications. Namely, suppose we apply a voltage V_q at edge e_q, with resistances $R_1, \ldots, R_{q-1}$ at the other edges $e_1, \ldots, e_{q-1}$. Let V_i, I_i be the voltage and current at edge e_i. We would like to express each V_i and I_i in terms of V_q and $R_1, \ldots, R_{q-1}$. (By "physical intuition" there should be a unique solution, since we can actually build a network meeting the specifications of the problem.) Note that if we have quantities V_i, I_i, R_i satisfying the three network laws above, then for any scalar α the quantities $\alpha V_i, \alpha I_i, R_i$ are also a solution. This means that we might as well assume that $V_q = 1$, since we can always multiply all voltages and currents afterwards by whatever value we want V_q to be.

The *total resistance* $R(D)$ of the network D just described is the (positive) resistance that would have to be assigned to an edge f parallel to e_q, with all edges $e_1, \ldots, e_{q-1}$ removed, so that the current flowing along f from $\operatorname{init}(e)$ to $fin(e)$ is the current flowing along e_q from $\operatorname{fin}(e)$ to $\operatorname{init}(e)$ in D. Thus by Kirchhoff's laws we have $R(D) = -V_{e_q}/I_{e_q}$.

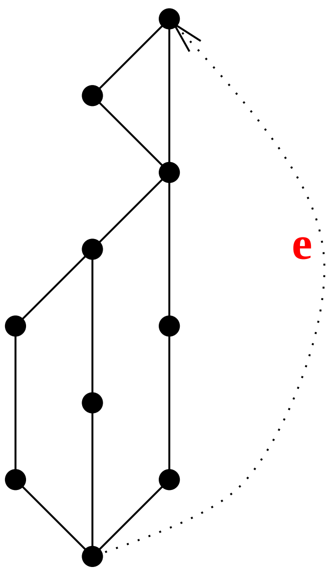
As an illustration of a simple method of computing the total resistance of a network, the following diagram illustrates the notion of a *series connection* $D_1 + D_2$ and a *parallel connection* $D_1 \parallel D_2$ of two networks D_1 and D_2 with a distinguished edge e at which a voltage is applied.



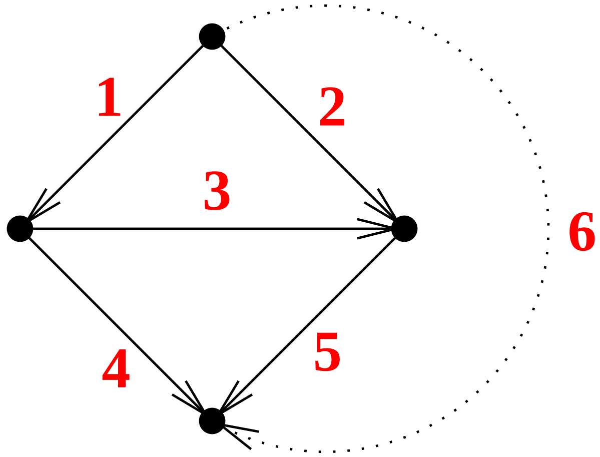
It is well-known and easy to deduce from the three network Laws that

$$\begin{aligned} R(D_1 + D_2) &= R(D_1) + R(D_2) \\ \frac{1}{R(D_1 \parallel D_2)} &= \frac{1}{R(D_1)} + \frac{1}{R(D_2)}. \end{aligned}$$

A network that is built up from a single edge by a sequence of series and parallel connections is called a *series-parallel network*. An example is the following, with the distinguished edge e shown by a broken line from bottom to top.



The simplest network which is not a series-parallel network and has no multiple edges (as an undirected graph) is called the *Wheatstone bridge* and is illustrated below. (The direction of the arrows has been chosen arbitrarily.) We will use this network as our main example in the discussion that follows.



We now return to an arbitrary connected loopless digraph D, with currents I_i, voltages V_i, and resistances R_i at the edges e_i. Recall that we are fixing $V_q = 1$ and $R_1, \ldots, R_{q-1}$. Let T be a spanning tree of D. Since I is a current if and only if it is orthogonal to the bond space $\mathcal{B}$ (Theorem 11.3 and

Kirchhoff's First Law), it follows that any basis for $\mathcal{B}$ defines a complete and minimal set of linear relations satisfied by the I_i's (namely, the relation that I is orthogonal to the basis elements). In particular, the basis matrix $\boldsymbol{C}_T$ defines such a set of relations. For example, if D is the Wheatstone bridge shown above and if $T = \{e_1, e_2, e_5\}$, then we obtain the following relations by adding the edges e_1, e_2, e_5 of T in turn to T^*.

$$\begin{aligned} I_1 - I_3 - I_4 &= 0 \\ I_2 + I_3 + I_4 + I_6 &= 0 \\ I_4 + I_5 + I_6 &= 0 \end{aligned} \tag{70}$$

These three $(= p - 1)$ equations give all the relations satisfied by the I_i's alone, and the equations are linearly independent.

Similary if V is a voltage then it is orthogonal to the cycle space $\mathcal{C}$. Thus any basis for $\mathcal{C}$ defines a complete and minimal set of linear relations satisfied by the V_i's (namely, the relation that V is orthogonal to the basis elements). In particular, the basis matrix C_T defines such a set of relations. Continuing our example, we obtain the following relations by adding the edges e_3, e_4, e_6 of T^* in turn to T.

$$\begin{aligned} V_1 - V_2 + V_3 &= 0 \\ V_1 - V_2 + V_4 - V_5 &= 0 \\ V_2 + V_5 &= 1, \end{aligned} \tag{71}$$

These three $(= q - p + k)$ equations give all the relations satisfied by the V_i's alone, and the equations are linearly independent.

In addition, Ohm's Law gives the $q-1$ equations $V_i = R_i I_i$, $1 \leq i \leq q-1$. We have a total of $(p - k) + (q - p + k) + (q - 1) = 2q - 1$ equations in the $2q-1$ unknowns I_i $(1 \leq i \leq q)$ and V_i $(1 \leq i \leq q-1)$. Moreover, it is easy to see that these $2q - 1$ equations are linearly independent, using the fact that we already know that just the equations involving the I_i's alone are linearly independent, and similarly the V_i's. Hence this system of $2q - 1$ equations in $2q - 1$ unknowns has a unique solution. We have now reduced the problem to straightforward linear algebra. However, it is possible to describe the solution explicitly. We will be content here with giving a formula just for the total resistance $R(D) = -V_q/I_q = -1/I_q$.

Write the $2q-1$ equations in the form of a $(2q-1)\times 2q$ matrix $\boldsymbol{K}$. The columns of the matrix are indexed by $I_1, I_2, \ldots, I_q$, V_1, V_2, $\ldots, V_q$. The last column V_q of the matrix keeps track of the constant terms of the equations. The rows of $\boldsymbol{K}$ are given first by the equations among the I_i's, then the V_i's, and finally Ohm's Law. For our example of the Wheatstone bridge, we obtain the matrix

$$\boldsymbol{K} = \begin{array}{|ccccc|c|ccccc|c|}
I_1 & I_2 & I_3 & I_4 & I_5 & I_6 & V_1 & V_2 & V_3 & V_4 & V_5 & V_6 \\
\hline
1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 1 \\
\hline
R_1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & R_2 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & R_3 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & R_4 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & R_5 & 0 & 0 & 0 & 0 & 0 & -1 & 0
\end{array}$$

We want to solve for I_q by Cramer's rule. Call the submatrix consisting of all but the last column X. Let Y be the result of replacing the I_q column of X by the last column of $\boldsymbol{K}$. Cramer's rule then asserts that

$$I_q = \frac{\det Y}{\det X}.$$

We evaluate $\det X$ by taking a Laplace expansion along the first $p-1$ rows. In other words,

$$\det X = \sum_S \pm \det(X[[p-1], S]) \cdot \det(X[[p-1]^c, \bar{S}]), \tag{72}$$

where (a) S indexes all $(p-1)$-element subsets of the columns, (b) $X[[p-1], S]$ denotes the submatrix of X consisting of entries in the first $p-1$ rows and in the columns S, (c) $X[[p-1]^c, \bar{S}]$ denotes the submatrix of X consisting of entries in the last $2q-p$ rows and in the columns other than S. In order for $\det(X[[p-1],S]) \neq 0$, we must choose $S = \{I_{i_1}, \ldots, I_{i_{p-1}}\}$, where $\{e_{i_1}, \ldots, e_{i_{p-1}}\}$ is a spanning tree T_1 (by Theorem 11.7(i)). In this case, $\det(X[[p-1],S]) = \pm 1$ by Theorem 11.14. If $I_q \notin S$, then the I_q column of $X[[p-1]^c, \bar{S}]$ will be zero. Hence to get a nonzero term in (72), we must have $e_q \in S$. The matrix $X[[p-1]^c, \bar{S}]$ will have one nonzero entry in each of the first $q-p+1$ columns, namely, the resistances R_j where e_j is not an edge of T_1. This accounts for $q-p+1$ entries from the last $q-1$ rows of $X[[p-1]^c, \bar{S}]$. The remaining $p-2$ of the last $q-1$ rows have available only one nonzero entry each, a -1 in the columns indexed by V_j where e_j is an edge of T_1 other than e_q. Hence we need to choose $q-p+1$ remaining entries from rows p through q and columns indexed by V_j for e_j not edge of T_1. By Theorems 11.7(ii) and 11.14, this remaining submatrix has determinant ± 1. It follows that

$$\det(X[[p-1],S]) \cdot \det(X[[p-1]^c, \bar{S}]) = \pm \prod_{e_j \notin E(T_1)} R_j.$$

Hence by (72), we get

$$\det X = \sum_{T_1} \pm \left(\prod_{e_j \notin E(T_1)} R_j \right), \tag{73}$$

where T_1 ranges over all spanning trees of D containing e_q. A careful analysis of the signs[6] shows that all signs in (73) are plus, so we finally arrive at the remarkable formula

$$\det X = \sum_{\substack{\text{spanning trees } T_1 \\ \text{containing } e_q}} \prod_{e_j \notin E(T_1)} R_j.$$

For example, if D is the Wheatstone bridge as above, and if we abbreviate $R_1 = a$, $R_2 = b$, $R_3 = c$, $R_4 = d$, $R_5 = e$, then

$$\det X = abc + abd + abe + ace + ade + bcd + bde + cde.$$

Now suppose we replace column I_q in X by column V_q in the matrix $\boldsymbol{K}$, obtaining the matrix Y. There is a unique nonzero entry in the new column,

[6] To be inserted.

so it must be chosen in any nonzero term in the expansion of $\det Y$. The argument now goes just as it did for $\det X$, except we have to choose S to correspond to a spanning tree T_1 that *doesn't* contain e_q. We therefore obtain

$$\det Y = \sum_{\substack{\text{spanning trees } T_1 \\ \text{not containing } e_q}} \prod_{\substack{e_j \notin E(T_1) \\ e_j \neq e_q}} R_j.$$

For example, for the Wheatstone bridge we get

$$\det Y = ac + ad + ae + bc + bd + be + cd + ce.$$

Recall that $I_q = \det(Y)/\det(X)$ and that the total resistance of the network is $1/I_q$. Putting everything together gives our main result on electrical networks.

11.15 Theorem. *In the situation described above, the total resistance of the network is given by*

$$R(D) = -\frac{1}{I_q} = \frac{\displaystyle\sum_{\substack{\text{spanning trees } T_1 \\ \text{containing } e_q}} \prod_{e_j \notin E(T_1)} R_j}{\displaystyle\sum_{\substack{\text{spanning trees } T_1 \\ \text{not containing } e_q}} \prod_{\substack{e_j \notin E(T_1) \\ e_j \neq e_q}} R_j}.$$

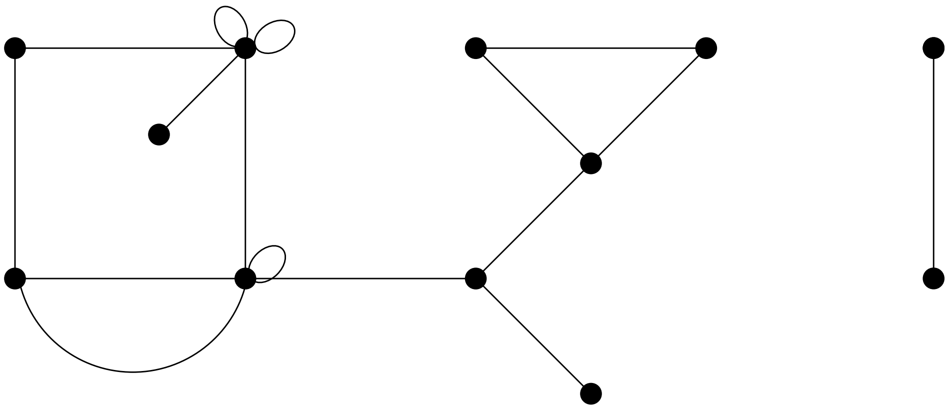
11.16 Corollary. *If the resistances $R_1, \ldots, R_{q-1}$ are all equal to one, then the total resistance of the network is given by*

$$R(D) = -\frac{1}{I_q} = \frac{\text{number of spanning trees containing } e_q}{\text{number of spanning trees not containing } e_q}.$$

In particular, if $R_1 = \cdots = R_{q-1} = 1$, then the total resistance, when reduced to lowest terms a/b, has the curious property that the number $\kappa(D)$ of spanning trees of D is divisible by $a + b$.

11.4 Planar graphs (sketch).

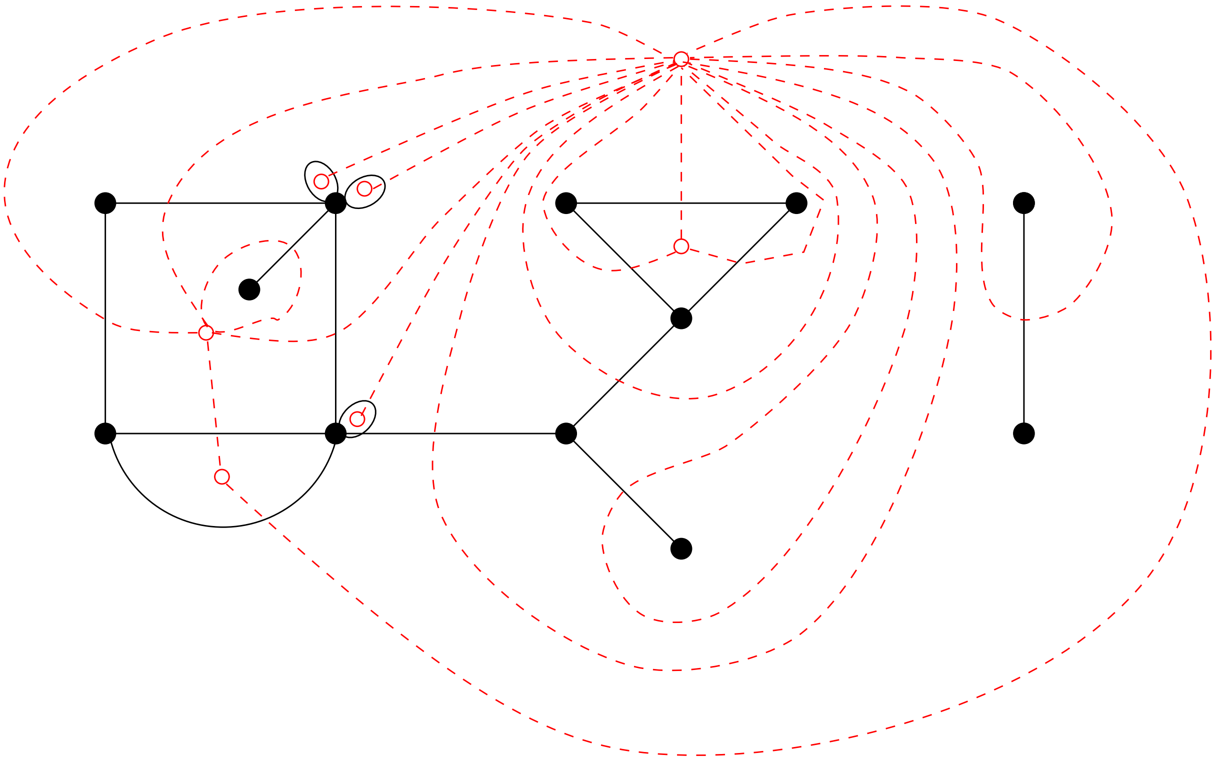
A graph G is *planar* if it can be drawn in the plane $\mathbb{R}^2$ without crossing edges. A drawing of G in this way is called a *planar embedding.*



If the vertices and edges of a planar embedding of G are removed from $\mathbb{R}^2$, then we obtain a disjoint union of open sets, called the *regions* (or *faces*) of G. (More precisely, these open sets are the regions of the planar embedding of G. Often we will not bother to distinguish between a planar graph and a planar embedding if no confusion should result.) Let $R = R(G)$ be the set of regions of G, and as usual $V(G)$ and $E(G)$ denote the set of vertices and edges of G, respectively.

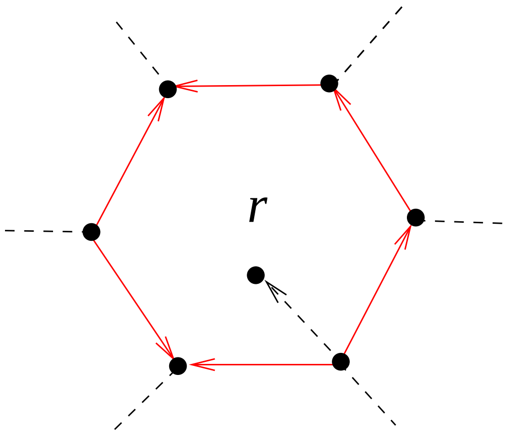
NOTE. If G is simple (no loops or multiple edges) then it can be shown that there exists a planar embedding with edges as straight lines and with regions (regarding as the sequence of vertices and edges obtained by walking around the boundaries of the regions) preserved.

The *dual* G^* of the planar embedded graph G has vertex set $R(G)$ and edge set $E^*(G) = \{e^* : e \in E(G)\}$. If e is an edge of G, then let r and r' be the regions on its two sides. (Possibly $r = r'$; there are five such edges in the example above.) Then define e^* to connect r and r'. We can always draw G^* to be planar, letting e and e^* intersect once. If G is connected then every region of G^* contains exactly one nonisolated vertex of G and $G^{**} \cong G$.

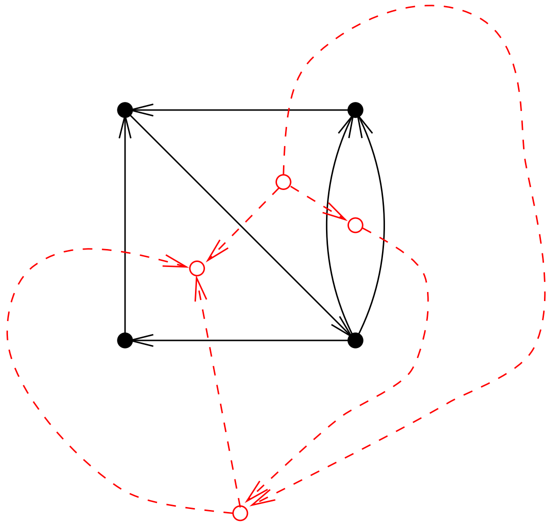


11.17 Example. Let G consist of two disjoint edges. Then G^* has one vertex and two loops, while G^{**} is a three-vertex path. The unbounded region of G^* contains two vertices of G, and $G^{**} \not\cong G$.

Orient the edges of the planar graph G in any way to get a digraph D. Let r be an interior (i.e., bounded) region of D. An *outside edge* of r is an edge e such that r lies on one side of the edge, and a *different* region lies on the other side. The outside edges of any interior region r define a circulation (shown as solid edges in the diagram below), and these circulations (as r ranges over all interior regions of D) form a basis for the cycle space $\mathcal{C}_G$ of G.



Given the orientation D of G, orient the edges of G^* as follows: as we walk along e in the direction of its orientation, e^* points to our *right*.



11.18 Theorem. *Let* $f : E(G) \to \mathbb{R}$. *Define* $f^* : E(G^*) \to \mathbb{R}$ *by* $f^*(e^*) = f(e)$. *Then*

$$\begin{aligned} f \in \mathcal{B}_G &\Leftrightarrow f^* \in \mathcal{C}_{G^*} \\ f \in \mathcal{C}_G &\Leftrightarrow f^* \in \mathcal{B}_{G^*}. \end{aligned}$$

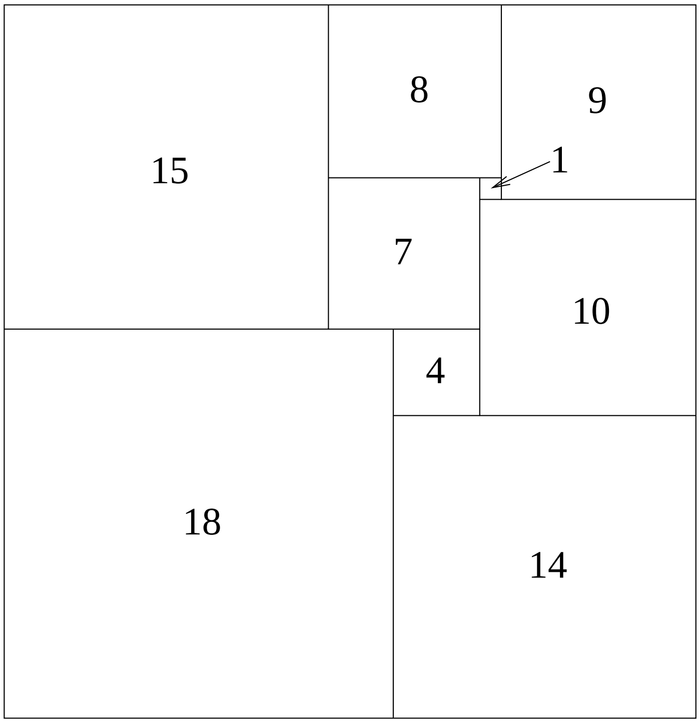
11.19 Proposition. *S is the set of edges of a spanning tree T of G if and only if $S^* = \{e^* : e \in S\}$ is the set of edges of a cotree T^* of G^*.*

11.20 Corollary. $\kappa(G) = \kappa(G^*)$

For nonplanar graphs there is still a notion of a "dual" object, but it is no longer a graph but rather something called a *matroid.* Matroid theory is a flourishing subject which may be regarded as a combinatorial abstraction of linear algebra.

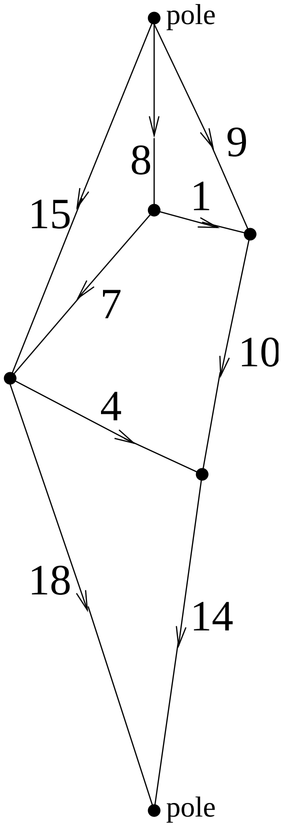
11.5 Squaring the square.

A *squared rectangle* is a rectangle partitioned into finitely many (but more than one) squares. A squared rectangle is *perfect* if all the squares are of different sizes. The earliest squared rectangle was found in 1936; its size is 33×32 and consists of nine squares:



The question then arose: does there exist a perfect squared square? An isolated example with 55 squares was found by Sprague in 1939. Then Brooks, Smith, Stone, and Tutte developed a network theory approach which we now explain.

The *Smith diagram D* of a squared rectangle is a directed graph whose vertices are the horizontal line segments of the squared rectangle and whose squares are the edges, directed from top to bottom. The top vertex (corresponding to the top edge of the rectangle being squared) and the bottom vertex (corresponding to the bottom edge) are called *poles*. Label each edge by the side length of the square to which it corresponds. The figure below shows the Smith diagram of the (perfect) squared rectangle above.



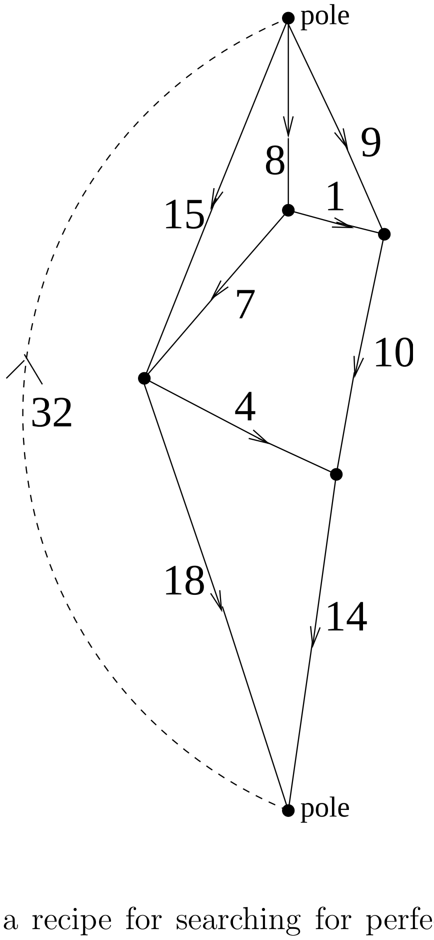
The following result concerning Smith diagrams is straightforward to verify.

11.21 Theorem.

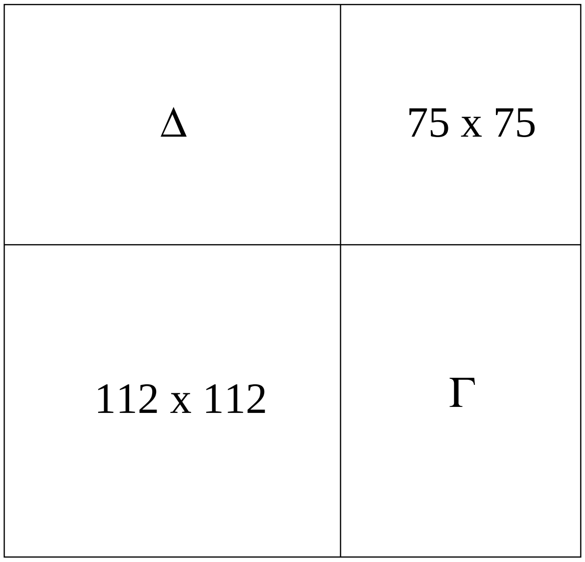
(a) *If we set I_e and V_e equal to the label of edge e, then Kirchhoff's two laws hold (so $R_e = 1$) except at the poles.*

(b) *The Smith diagram is planar and can be drawn without separation of poles. Joining the poles by an edge from the bottom to the top gives a 3-connected graph, i.e., a connected graph that remains connected when one or two vertices are removed.*

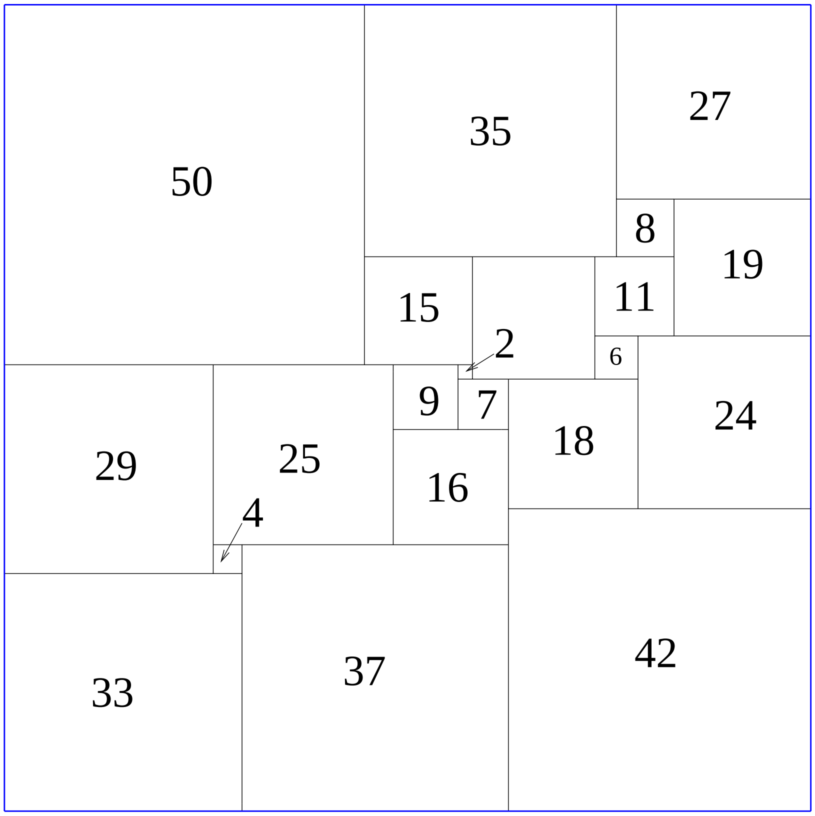
Call the 3-connected graph of Theorem 11.21 the *extended* Smith diagram of the squared rectangle. If we label the new edge e_1 between poles by the horizontal length b of the squared rectangle and set $V_{e_1} = I_{e_1} = b$, then Kirchhoff's two laws hold at *all* vertices.



We therefore have a recipe for searching for perfect squared rectangles and squares: start listing all 3-connected planar graphs. Then choose an edge e_1 to apply a voltage V_1. Put a resistance $R_e = 1$ at the remaining edges e. Solve for I_e $(= V_e)$ to get a squared rectangle, and hope that one of these will be a square. One example Γ found by Brooks et al. was a 112×75 rectangle with 14 squares. It was given to Brooks' mother as a jigsaw puzzle, and she found a different solution Δ! We therefore have found a squared square (though not perfect):



Building on this idea, Brooks et al. finally found two 422×593 perfect rectangles with thirteen squares, all 26 squares being of different sizes. Putting them together as above gives a perfect squared square. This example has two defects: (a) it contains a smaller perfect squared rectangle (and is therefore not *simple*), and (b) it contains a "cross" (four squares meeting a point). They eventually found a perfect squared square with 69 squares without either of these defects. It is now known (thanks to computers) that the smallest order (number of squares) of a perfect squared square is 21. It is unique and happens to be simple and crossfree. See the figure below. It is known that the number (up to symmetry) of simple perfect squared squares of order n for $n \geq 21$ is $1, 8, 12, 26, 160, 441, 1152, \ldots$.


50
35
27
8
19
15
2
11
6
24
29
25
9
7
18
16
4
33
37
42