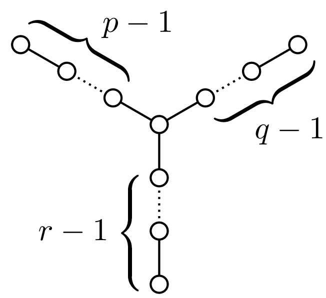
Course: *ADE classifications* **Winter 2025/2026**
Drew Armstrong **RUB Bochum**

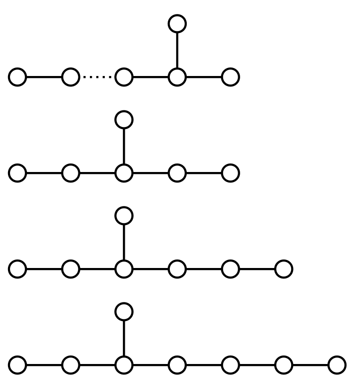
Contents

0 The inequality $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$

Let Y_{pqr} denote the following undirected "Y-shaped" graph with $(p-1)+(q-1)+(r-1)+1 = p+q+r-2$ vertices:



The graphs satisfying $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$ have special names. They are labeled by the letters ADE. The subscript indicates the number of vertices:



A_n — $\{p, q, r\} = \{1, k, n-k+1\}$

D_n — $\{p, q, r\} = \{2, 2, n-2\}$

E_6 — $\{p, q, r\} = \{2, 3, 3\}$

E_7 — $\{p, q, r\} = \{2, 3, 4\}$

E_8 — $\{p, q, r\} = \{2, 3, 5\}$

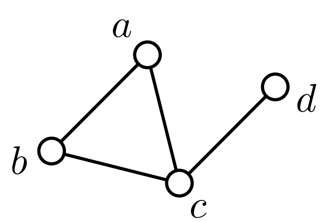
These graphs show up everywhere in mathematics. Their first appearance was in the classification of continuous transformation groups (now called Lie groups) by Lie, Killing, Cartan and Weyl. The graphs themselves were invented independently by Coxeter and Dykin, hence they are somtimes called Coxeter diagrams and sometimes Dynkin diagrams. The ADE labeling scheme goes back to Killing in 1887. You might be wondering what happened to BCFGH, etc. We'll discuss that later.

1 Graphs with small eigenvalues

1.1 Eigenvalues of graphs

Terry Gannon (in *Moonshine Beyond the Monster*) suggests that the simplest characterization of the ADE diagrams has to do with the eigenvalues of their adjacency matrices.

Given a weighted, directed graph G on vertex set $\{v_1, \ldots, v_n\}$, the *adjacency matrix* is the $n \times n$ matrix $A_G = (a_{ij})$ whose ij entry a_{ij} is the weight on the edge from vertex v_i to vertex v_j. We could also let a_{ij} be the **number** of edges from v_i to v_j if we are working with multigraphs. If the graph is undirected then we set $a_{ij} = a_{ji}$ so the matrix is symmetric. If the graph is undirected and simple (no multiple edges and no loops) then A_G will be a symmetric $0, 1$ matrix with zeroes on the diagonal. For example, consider the graph G with vertices $\{a, b, c, d\}$ and edges $\{\{a, b\}, \{a, c\}, \{b, c\}, \{c, d\}\}$:

G = 

$$A_G = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

With this convention, we note that the ij entry of the kth power A_G^k equals the number of paths of length k from v_i to v_j, i.e., the number of paths[1]

$$v_{i_1} - v_{i_2} - \cdots - v_{i_k},$$

with $i_1 = i$ and $i_k = j$, where $v_{i_s} - v_{i_{s+1}}$ is an edge for all s.

Recall that the eigenvalues of a real symmetric matrix are real. Indeed, let A be a real symmetric matrix and let $\mathbf{v} \neq \mathbf{0}$ be a (possibly complex) eigenvector with (possibly complex) eigenvalue λ. Let $*$ denote the conjgate transpose operation, so that $A^* = A$. Then we have

$$\bar{\lambda}\|\mathbf{v}\|^2 = \bar{\lambda}\mathbf{v}^*\mathbf{v} = (\lambda\mathbf{v})^*\mathbf{v} = (A\mathbf{v})^*\mathbf{v} = \mathbf{v}^*A^*\mathbf{v} = \mathbf{v}^*\lambda\mathbf{v} = \lambda\mathbf{v}^*\mathbf{v} = \lambda\|\mathbf{v}\|^2.$$

Since $\mathbf{v} \neq \mathbf{0}$ implies $\|\mathbf{v}\| \neq 0$ we can cancel to obtain $\lambda^* = \lambda$ and hence $\lambda \in \mathbb{R}$. This fact was first proved by Cauchy in 1829, as part of his extension of Euler's principal axes theorem to higher dimensions. Cauchy's original proof was quite complicated.[2] Here is the modern statement and proof of Cauchy's result.

Theorem 1.1 (Principal Axes Theorem)**.** *Let A be a real and symmetric $n \times n$ matrix. Then A has a basis of real orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^n$.*

Proof. The theorem is equivalent to the statement that there exists a real orthogonal matrix $U^TU = I$ such that $U^TAU = \Lambda := \text{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal. Indeed, if $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are the columns of U then $U^TU = I$ is equivalent to the fact that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are orthonormal and the equation $U^TAU = \Lambda$ is equivalent to the statement

$$AU = U\Lambda$$

[1] If the graph is directed the the entries of of powers of A_G count directed paths.

[2] "Dazzled by the brilliance of the new theory of determinants, mathematicians overlooked simple inner product considerations", Hawkins, The Mathematics of Frobenius in Context, page 98.

$$\begin{pmatrix} A\mathbf{u}_1 & \cdots & A\mathbf{u}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{u}_1 & \cdots & \lambda_n\mathbf{u}_n \end{pmatrix}.$$

We will prove this matrix version of the theorem by induction on the size of A.

The characteristic polynomial of A has a complex root, but we know from the previous remark that every eigenvalue is real, hence A has a real eigenvalue, say $\lambda_1 \in \mathbb{R}$. Let $\mathbf{v}_1$ be a (necessarily real) unit vector satisfying $A\mathbf{v}_1 = \lambda_1\mathbf{v}_1$. Since $A^T = A$ we note that A stabilizes the hyperplane $\mathbf{v}_1^\perp$ since for any vector $\mathbf{x}$ with $\mathbf{v}_1^T\mathbf{x} = 0$ we have

$$\mathbf{v}_1^T(A\mathbf{x}) = \mathbf{v}_1^T A^T\mathbf{x} = (A\mathbf{v}_1)^T\mathbf{x} = (\lambda_1\mathbf{v}_1)^T\mathbf{x} = \lambda_1\mathbf{v}_1^T\mathbf{x} = 0.$$

Let $\mathbf{v}_2, \ldots, \mathbf{v}_n$ be any orthonormal basis of the hyperplane $\mathbf{v}_1^\perp$ and let V be the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ so that $V^TV = I$. Note that the ij entry of V^TAV is $\mathbf{v}_i^TA\mathbf{v}_j = (\mathbf{v}_i^TA\mathbf{v}_j)^T = \mathbf{v}_j^TA^T\mathbf{v}_i = \mathbf{v}_j^TA\mathbf{v}_i$. Since $\mathbf{v}_1^TA\mathbf{v}_1 = \lambda_1$ and $\mathbf{v}_1^TA\mathbf{v}_i = 0$ for all $i \geq 2$ we conclude that

$$V^TAV = \left(\begin{array}{c|ccc} \lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right)$$

for some real matrix A', which is symmetric because $(V^TAV)^T = V^TA^TV = V^TAV$ is symmetric. By induction there exists an $(n-1) \times (n-1)$ orthogonal matrix V' such that $(V')^TA'V'$ is diagonal. If we define

$$W := \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & V' & \\ 0 & & & \end{array}\right),$$

then we observe that $W^TW = I$ and $(VW)^T(VW) = W^TV^TVW = W^TW = I$, and

$$(VW)^TA(VW) = W^T\left(\begin{array}{c|ccc} \lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right)W = \left(\begin{array}{c|ccc} \lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & (V')^TA'V' & \\ 0 & & & \end{array}\right)$$

is diagonal. Thus $U := WV$ is the desired orthogonal matrix. □

Let $A^T = A$ and let $U^TU = I$ with $U^TAU = \Lambda := \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Then since $U^T = U^{-1}$ we have for any integer $k \geq 0$ that

$$A^k = (U\Lambda U^{-1})^k = U\Lambda^kU^{-1} = U\Lambda^kU^T.$$

If $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are the (orthonormal) columns of U then this implies

$$A^k = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix} = \sum_{m=1}^n \lambda_m^k \mathbf{u}_m \mathbf{u}_m^T.$$

And if $\mathbf{u}_m^T = \begin{pmatrix} u_{1m} & \cdots & u_{nm} \end{pmatrix}$ then the ij entry of A^k is $\sum_{m=1}^n u_{im} u_{jm} \lambda_m^k$.

Stanley exploited this result in Chapter 1 of his undergraduate textbook *Algebraic Combinatorics* to study walks in graphs. Suppose $A = A_G$ is the adjacency matrix of a simple graph G with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Then for fixed vertices v_i and v_j there exist constants $c_1, \ldots, c_n \in \mathbb{R}$ such that the number walks between v_i and v_j of length k is $c_1\lambda_1^k + \cdots + c_n\lambda_n^k$. Indeed, take $c_i := u_{im}u_{jm}$ from the previous remark. In particular, the growth of walks in G is controlled by the largest eigenvalue of A_G.

It may be interesting to note that the eigenvalues of a **bipartite** graph are symmetric about zero.[3] Proof: Let G be bipartite on two vertex sets X, Y. If we order the vertices of G so that X comes before Y then the adjacency matrix has the form

$$A = \left(\begin{array}{c|c} 0 & B \\ \hline B^T & 0 \end{array}\right).$$

Suppose $A\mathbf{v} = \lambda\mathbf{v}$ and write $\mathbf{v}^T = (\mathbf{v}_1^T | \mathbf{v}_2^T)$ in block form. Then block multiplication gives

$$\left(\begin{array}{c} \lambda\mathbf{v}_1 \\ \hline \lambda\mathbf{v}_2 \end{array}\right) = \lambda\mathbf{v} = A\mathbf{v} = \left(\begin{array}{c|c} 0 & B \\ \hline B^T & 0 \end{array}\right) \left(\begin{array}{c} \mathbf{v}_1 \\ \hline \mathbf{v}_2 \end{array}\right) = \left(\begin{array}{c} B\mathbf{v}_2 \\ \hline B^T\mathbf{v}_1 \end{array}\right),$$

hence $B\mathbf{v}_2 = \lambda\mathbf{v}_1$ and $B^T\mathbf{v}_1 = \lambda\mathbf{v}_2$. On the other hand, if $\tilde{\mathbf{v}}^T = (\mathbf{v}_1^T | -\mathbf{v}_2^T)$ then we have

$$A\tilde{\mathbf{v}} = \left(\begin{array}{c|c} 0 & B \\ \hline B^T & 0 \end{array}\right) \left(\begin{array}{c} \mathbf{v}_1 \\ \hline -\mathbf{v}_2 \end{array}\right) = \left(\begin{array}{c} -B\mathbf{v}_2 \\ \hline B^T\mathbf{v}_1 \end{array}\right) = \left(\begin{array}{c} -\lambda\mathbf{v}_1 \\ \hline \lambda\mathbf{v}_2 \end{array}\right) = -\lambda \left(\begin{array}{c} \mathbf{v}_1 \\ \hline -\mathbf{v}_2 \end{array}\right) = -\lambda\tilde{\mathbf{v}}.$$

We conclude that if λ is an eigenvalue of A then $-\lambda$ is also an eigenvalue. Note that every tree is bipartite, so the eigenvalues of the ADE graphs are real and symmetric about zero.

1.2 Lagrange and sound waves

Let's consider the type A_n chain graphs explicitly. The adjacency matrix is the $n \times n$ matrix

$$A = \begin{pmatrix} 0 & 1 & & \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 0 \end{pmatrix}.$$

[3] The converse is also true. That is, if the eigenvalues of A_G are symmetric about zero then G is bipartite.

According to Goodman-Harpe-Jones (in *Coxeter Graphs and Towers of Algebras*), this matrix was first explicitly diagonalized by Lagrange (in *Researches on the nature and propagation of sound*, 1759). Maybe we'll discuss the physics later. For now I'll just show you the eigenvalues and eigenvectors. I claim that the eigenvectors are

$$\mathbf{v}_k := \begin{pmatrix} \sin(\theta_k) \\ \sin(2\theta_k) \\ \vdots \\ \sin(n\theta_k) \end{pmatrix} \quad \text{for} \quad k = 1, 2, \ldots, n,$$

where $\theta_k := k\pi/(n+1)$. And the eigenvalues are given by the equations $A\mathbf{v}_k = 2\cos(\theta_k)\mathbf{v}_k$. To prove this, we simply add two trigonometric identities:

$$\begin{array}{rcl} \sin(i\theta_k + \theta_k) & = & \sin(i\theta_k)\cos(\theta_k) + \cos(i\theta_k)\sin(\theta_k) \\ \sin(i\theta_k - \theta_k) & = & \sin(i\theta_k)\cos(\theta_k) - \cos(i\theta_k)\sin(\theta_k) \\ \hline \sin((i+1)\theta_k) + \sin((i-1)\theta_k) & = & 2\cos(\theta_k)\sin(i\theta_k). \end{array}$$

Then everything just works out. Remark: Let $\lambda_k = 2\cos(\theta_k)$. Since A is symmetric, we have

$$\lambda_k \mathbf{v}_k^T \mathbf{v}_\ell = (A\mathbf{v}_k)^T \mathbf{v}_\ell = \mathbf{v}_k^T A^T \mathbf{v}_\ell = \mathbf{v}_k^T A \mathbf{v}_\ell = \lambda_\ell \mathbf{v}_k^T \mathbf{v}_\ell.$$

If $k \neq \ell$ then we have $\lambda_k \neq \lambda_\ell$ and hence $\mathbf{v}_k^T \mathbf{v}_\ell = 0$.[4] Thus we obtain some interesting trigonometric identities:

$$\sum_{i=1}^{n} \sin\left(\frac{ik\pi}{n+1}\right) \sin\left(\frac{i\ell\pi}{n+1}\right) = 0 \quad \text{for } k \neq \ell.$$

This is a discrete analogue of the orthogonality of sine waves in Fourier analysis.

1.3 Graphs with small eigenvalues

We showed that the eigenvalues of the type A_n (chain) graph are

$$2\cos\left(\frac{n\pi}{n+1}\right) < 2\cos\left(\frac{(n-1)\pi}{n+1}\right) < \cdots < 2\cos\left(\frac{2\pi}{n+1}\right) < 2\cos\left(\frac{\pi}{n+1}\right).$$

The eigenvalues are symmetric about zero because $\cos(\pi - \theta) = -\cos(\theta)$. As $n \to \infty$ the largest eigenvalue approaches 2 but it never reaches it. In fact, this property **chracterizes** the ADE graphs. Given a graph G we define its *spectral radius* as the maximum of the absolute values of the eigenvalues of the adjacency matrix:

$$\rho(G) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A_G\}.$$

[4] The same proof shows that eigenvectors corresponding to distinct eigenvalues of a real symmetric matrix are orthogonal.

The following result was published by J. H. Smith (*Some properties of the spectrum of a graph*, 1970) but it should really be viewed as a folklore result that is implicit in every ADE type classification.[5]

Theorem 1.2. *Let G be a simple undirected graph. Then*

$$\rho(G) < 2 \quad \Longleftrightarrow \quad G \text{ is a disjoint union of ADE type graphs.}$$

First note that we can restrict our attention to connected graphs. Indeed, if $G = G_1 \cup G_2$ is a disjoint union of graphs then the adjacency matrix is block-diagonal:

$$A_G = \left(\begin{array}{c|c} A_{G_1} & 0 \\ \hline 0 & A_{G_2} \end{array} \right).$$

This implies that the set of eigenvalues of G is the union of the sets of eigenvalues of G_1 and G_2, hence the spectral radius $\rho(G)$ is the maximum of the spectral radii $\rho(G_1)$ and $\rho(G_2)$:

$$\rho(G) = \max\{\rho(G_1), \rho(G_2)\}.$$

The proof of the theorem will follow from a monotonicity property for subgraphs that is part of the Perron-Frobenius family of theorems for positive matrices. I spent far too many hours trying to work out a streamlined proof that I can present in class but I finally decided to state the result without proof. You can find the details in Chapter 6 of *Graphs and Matrices* by Bapat, though even he quotes a few facts without proof.

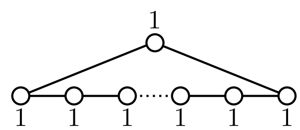
Lemma 1.3 (The Perron-Frobenius Theorem for Graphs)**.** *We will say that a vector is positive when its entries are strictly positive real numbers.*

- *(The Perron Eigenvector) Let G be a connected graph. The adjacency matrix A_G has has a unique (up to scalars) positive eigenvector. The corresponding eigenvalue is the spectral radius $\rho(G)$.*
- *(Strict Monotonicity) Let G be a connected graph and let H be any subgraph. This means that the adjacency matrix A_H is obtained from A_G by taking a principal submatrix (deleting vertices) and shrinking entries (deleting edges). Then we have $\rho(H) \leq \rho(G)$. If $H \neq G$ then $\rho(H) < \rho(G)$.*

An eigenvector $\mathbf{x} = (x_1, \ldots, x_n)$ of the $n \times n$ adjacency matrix A_G can be viewed as a labeling of the vertices. The condition $A_G\mathbf{x} = \lambda\mathbf{x}$ is equivalent to the statement that λx_i equals the

[5] I haven't been able to obtain Smith's paper. Cameron, Goethals and Seidel (*Line Graphs, Root Systems and Elliptic Geometry*, 1975) prove the same result using root systems, which we will discuss later. It is interesting that they do not mention Smith's paper; they merely mention that "graphs whose adjacency matrix has least eigenvalue -2 have gained much attention in these last 15 years". We recommend the book *Coxeter Graphs and Towers of Algebras* (1980) by Goodman, de la Harpe and Jones, for a comprehensive account of the subject.

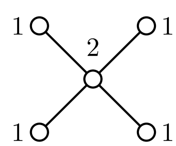
sum of the labels x_j over all j such that v_i, v_j is an edge of G, i.e., the sum of the labels on the neighbors of vertex v_i. As an application, we claim that any cycle graph has spectral radius 2. Indeed, consider the following vertex labeling:



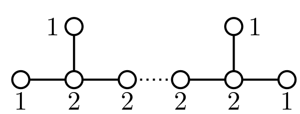
Note that 2 times any vertex label equals the sum of the labels of the neighboring vertices. Since this gives a positive eigenvector of eigenvalue 2 it follows from the first part of the lemma that 2 is the spectral radius.

Now we give the proof of Theorem 1.2.

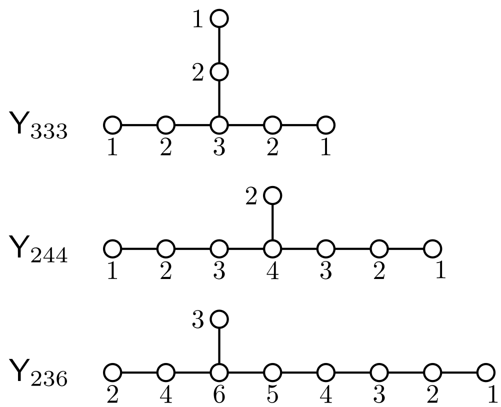
Proof. Let G be a connected graph with spectral radius $\rho(G) < 2$. Then by strict monotonicity of spectral radius G cannot contain a cycle because we just saw that a cycle has spectral radius 2. Next we observe that G cannot contain a vertex of degree ≥ 4, otherwise it must contain the following graph as a subgraph:



But this graph has spectral radius 2 via the displayed vertex labeling. Next we observe that G has at most one vertex of degree 3. If not then G contains the following subgraph, which has spectral radius 2 via the displayed vertex labeling:



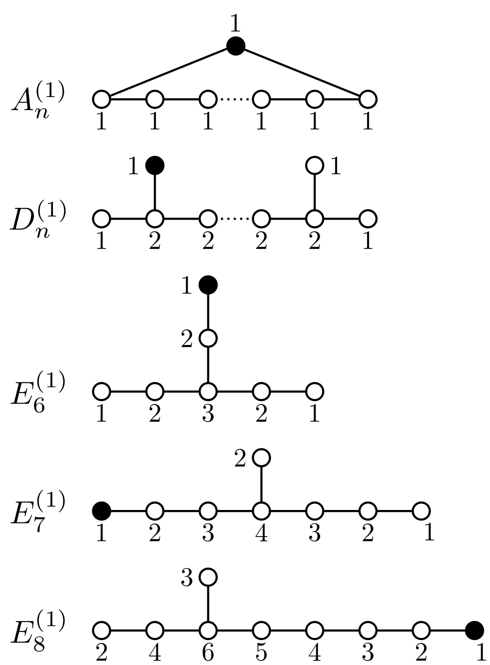
At this point we have shown that G is a Y-shaped graph of the form Y_{pqr}. The rest of the proof follows from the fact that the graphs Y_{333}, Y_{244} and Y_{236} each has spectral radius 2, as shown by the following vertex labelings:



□

This proof is short but mysterious because it involves the existence of some special graphs with special vertex labelings that we seem to have pulled from nowhere. We will investigate this further in the next section. For now, we push the proof a bit further to obtain a classification of graphs with spectral radius equal to 2.

Theorem 1.4. *Let G be a connected simple undirected graph with spectral radius $\rho(G) = 2$. Then G is one of the following* affine ADE graphs*:*



Proof. Note that each of the displayed graphs has spectral radius 2 via the displayed vertex labeling. Conversely, let G be a connected undirected graph with spectral radius $\rho(G) = 2$. If G contains a cycle then it must equal a cycle graph $A_n^{(1)}$. Otherwise it contains some $A_n^{(1)}$ as a strict subgraph, and hence has spectral radius > 2 by strict monotonicity. Next we observe that G has no vertex of degree ≥ 5. Otherwise it contains $D_4^{(1)}$ as a strict subgraph, hence has spectral radius > 2 by strict monotonicity. Similarly, G has at most one vertex of degree 4, otherwise it contains some $D_n^{(1)}$ as a strict subgraph. And if G has a vertex of degree 4 then we must have $G = D_4^{(1)}$ otherwise G contains $D_4^{(1)}$ as a strict subgraph. Furthermore, if $G \neq D_n^{(1)}$ then it has at most one vertex of degree 3, otherwise it contains some $D_n^{(1)}$ as a strict subgraph. We are left with the case of Y-shaped graphs Y_{pqr}, and the proof follows from the fact that the spectral radius is strictly monotone with the arm lengths. □

The relation between a "finite type" ADE graph G and its "affine extension" $G^{(1)}$ comes from the theory of Weyl groups, which we will discuss later.

1.4 Relation to the quantity $\frac{1}{p} + \frac{1}{q} + \frac{1}{r}$

The previous two classification results can be rephrased in the following compact way.

Theorem 1.5. *Let G be a connected, undirected simple graph, and recall that Y_{pqr} denotes the Y-shaped graph with $(p-1)+(q-1)+(r-1)+1$ vertices.*

- *If $\rho(G) < 2$ then $G = Y_{pqr}$ with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$.*
- *If $\rho(G) = 2$ then $G = A_n^{(1)}$, $D_n^{(1)}$ or Y_{pqr} with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$.*

However, the previous proofs give no indication why this should be the case. In this section we explain the appearance of the quantity $\frac{1}{p} + \frac{1}{q} + \frac{1}{r}$.

The *Chebyshev polynomials of the second kind* are defined by the following initial conditions and three-term recurrence:

$$\begin{aligned} U_0(x) &= 1, \\ U_1(x) &= 2x, \\ U_{k+1}(x) &= 2xU_k(x) - U_{k-1}(x). \end{aligned}$$

These polynomials encode the *multiple angle identities* for the sine function,

$$U_k(\cos\theta) = \frac{\sin((k+1)\theta)}{\sin\theta},$$

as one can check by induction. Consider the Y-shaped graph Y_{pqr} for any integers $p, q, r \geq 2$ and let $\lambda = \rho(Y_{pqr}) > 0$ be its spectral radius. From the Perron-Frobenius theorem in the previous section there exists a unique λ-eigenvector $\mathbf{x}$ for the adjacency matrix A with strictly positive entries. We can think of the entries of $\mathbf{x}$ as a vertex labeling. Let $x_0, \ldots, x_{p-1}$, $y_0, \ldots, y_{q-1}$ and $z_0, \ldots, z_{r-1}$ be the labels along the three arms of the graph, with x_0, y_0, z_0

labeling the leaves and $x_{p-1} = y_{q-1} = z_{r-1}$ labeling the center vertex. The eigenvector equation $A\mathbf{x} = \lambda\mathbf{x}$ tells us that

$$\begin{aligned} x_1 &= \lambda x_0, \\ x_{k+1} &= \lambda x_k - x_{k-1} \text{ for } 1 \leq k \leq p-2. \end{aligned}$$

By induction this implies that $x_k = x_0 U_k(\lambda/2)$ for all $0 \leq k \leq p-1$. Similarly, we have $y_k = y_0 U_k(\lambda/2)$ for $0 \leq k \leq q-1$ and $z_k = z_0 U_k(\lambda/2)$ for $0 \leq k \leq r-1$. Let $c := x_{p-1} = y_{q-1} = z_{r-1}$ denote the center label, so that

$$c = x_0 U_{p-1}(\lambda/2) = y_0 U_{q-1}(\lambda/2) = z_0 U_{r-1}(\lambda/2).$$

This determines all vertex labels in terms of the eigenvalue λ and the center label c:

$$\begin{aligned} x_k &= c\, U_k(\lambda/2)/U_{p-1}(\lambda/2), \\ y_k &= c\, U_k(\lambda/2)/U_{q-1}(\lambda/2), \\ z_k &= c\, U_k(\lambda/2)/U_{r-1}(\lambda/2). \end{aligned}$$

We have one more piece of information. The eigenvector equation $A\mathbf{x} = \lambda\mathbf{x}$ at the center vertex tells us that

$$\lambda c = x_{p-2} + y_{q-2} + z_{r-2}.$$

Note that $x_{p-2}/c = x_{p-2}/x_{p-1} = U_{p-2}(\lambda/2)/U_{p-1}(\lambda/2)$. Similarly, since $c = x_{p-1} = y_{q-1} = z_{r-1}$ we have $y_{q-2}/c = U_{q-2}(\lambda/2)/U_{q-1}(\lambda/2)$ and $z_{r-2}/c = U_{r-2}(\lambda/2)/U_{r-1}(\lambda/2)$. Hence dividing both sides of the balance equation by c (which is positive by assumption) gives

$$\lambda = \frac{U_{p-2}(\lambda/2)}{U_{p-1}(\lambda/2)} + \frac{U_{q-2}(\lambda/2)}{U_{q-1}(\lambda/2)} + \frac{U_{r-2}(\lambda/2)}{U_{r-1}(\lambda/2)}.$$

This is an explicit relationship between the Perron-Frobenius eigenvalue λ and the three arm lengths $p, q, r \geq 2$. One can check by induction that $U_k(1) = k+1$ for any $k \geq 0$. If the spectral radius is $\lambda = 2$ then the previous equation implies that

$$\begin{aligned} 2 &= \frac{p-1}{p} + \frac{q-1}{q} + \frac{r-1}{r} \\ 2 &= 1 - \frac{1}{p} + 1 - \frac{1}{q} + 1 - \frac{1}{r} \\ \frac{1}{p} + \frac{1}{q} + \frac{1}{r} &= 1. \end{aligned}$$

The solutions of this Diophantine equation correspond to the affine ADE graphs of type E. In this case, let me also mention that substituting $\lambda = 2$ into the equations for the Perron-Frobenius vertex labels gives $x_k = c(k+1)/p$, $y_k = c(k+1)/q$ and $z_k = c(k+1)/r$. Since the Perron-Frobenius eigenvector is only defined up to scaling, we can choose c so that all Perron-Frobenius labels are positive integers.

Next, suppose that $\lambda < 2$. In this case we can write $\lambda = 2\cos\theta$ for some nonzero angle θ, which is perfectly suited to our Chebyshev formulas because then

$$U_k(\lambda/2) = U_k(\cos\theta) = \frac{\sin((k+1)\theta)}{\sin\theta}.$$

In this case, the balance equation at the central vertex becomes

$$2\cos\theta = \frac{\sin((p-1)\theta)}{\sin(p\theta)} + \frac{\sin((q-1)\theta)}{\sin(q\theta)} + \frac{\sin((r-1)\theta)}{\sin(r\theta)}.$$

The equation $X = x_0 U_{p-1}(\cos\theta) = x_0 \sin(p\theta)/\sin\theta$ and the positivity of the labels c, x_0 implies that $0 < \theta \le \pi/p$. Similarly we must have $\theta \le \pi/q$ and $\theta \le \pi/r$. One can check that $\sin((n+1)t)/\sin(nt)$ is a strictly increasing function of t on $(0, \pi/n]$, hence the following function is strictly decreasing on the interval $(0, \theta]$:

$$F(t) := 2\cos t - \frac{\sin((p-1)t)}{\sin(pt)} - \frac{\sin((q-1)t)}{\sin(qt)} - \frac{\sin((r-1)t)}{\sin(rt)}.$$

It follows that

$$\begin{aligned} \lim_{t\to 0^+} F(t) &> F(\theta) \\ 2 - \frac{p-1}{p} - \frac{q-1}{q} - \frac{r-1}{r} &> 0 \\ -1 + \frac{1}{p} + \frac{1}{q} + \frac{1}{r} &> 0 \\ \frac{1}{p} + \frac{1}{q} + \frac{1}{r} &> 1. \end{aligned}$$

1.5 Marks and Exponents

The subject of ADE classification is full of special integers that satisfy unexpected relationships — a phenomenon called *numerology*. Some of this numerology is already visible in the simple case of graphs. We introduce two concepts now. Given a finite type ADE graph (i.e., with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$) with n vertices, we will prove later that the eigenvalues have the form

$$2\cos(m_1\pi/h), 2\cos(m_2\pi/h), \ldots, 2\cos(m_n\pi/h),$$

for some integers $1 \le m_1 \le \cdots \le m_n$ called the *exponents* and some integer h called the *Coxeter number* of the graph. The spectral radius is $2\cos(\pi/h)$ and the smallest exponent is $m_1 = 1$. Because each ADE graph is bipartite we know that the eigenvalues are symmetric about zero, from which is follows that $m_i + m_{n+1-i} = h$ for all i.

If G is a finite type ADE graph, let $G^{(1)}$ denote the corresponding affine type ADE graph. There is a canonical Perron-Frobenius eigenvector for $G^{(1)}$ with positive, coprime integer entries. (These are the vertex labels in Theorem 1.4.) If G has n vertices then $G^{(1)}$ has $n+1$ labels $c_0, c_1, \ldots, c_n$, which are called the *marks* (or Kac labels) of the graph. It turns out that the label 1 always occurs so we can take $c_0 = 1$. It is a surprising fact that

$$c_0 + c_1 + \cdots + c_n = h,$$

where h is the Coxeter number of the finite type graph G.[6] Another strange identity relates the marks to the arm lengths p, q, r of a finite type ADE graph Y_{pqr}:

$$c_0^2 + c_1^2 + \ldots + c_n^2 = \frac{4}{\frac{1}{p} + \frac{1}{q} + \frac{1}{r} - 1}.$$

This identity comes from the McKay correspondence, which we will discuss in the next chapter. The following table shows the marks and exponents for the finite type ADE graphs.

graph G	marks of $G^{(1)}$	exponents of G	Coxeter number h
A_n	$1,1,\ldots,1$	$1,2,\ldots,n$	$n+1$
D_n	$1,1,1,1,2,\ldots,2$	$n-1,1,3,\ldots,2n-3$	$2(n-1)$
E_6	$1,1,1,2,2,2,3$	$1,4,5,7,8,11$	12
E_7	$1,1,2,2,2,3,4$	$1,5,7,9,11,13,17$	18
E_8	$1,2,2,3,3,4,4,5,6$	$1,7,11,13,17,19,23,29$	30

2 Finite groups of rotations

2.1 Platonic solids

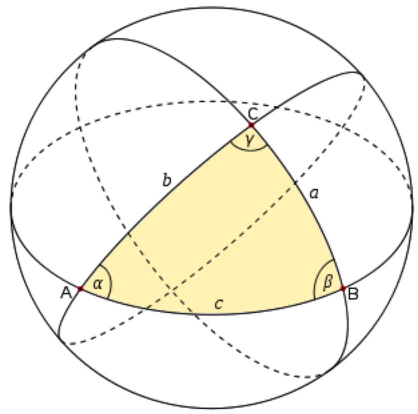
A *regular polytope* is a convex polytope $P \subseteq \mathbb{R}^n$ for which the group of symmetries of P acts transitively on maximal flags of faces. (See the end of this section for details.) A *Platonic solid* is a three dimensional regular polytope. There are only five of these and they were classified in the final book of Euclid's *Elements*. We will see that the Platonic solids are another example of ADE classification. This is described by the following table, which will be explained in this chapter.[7]

type	shape$P$	symmetry group
A_n	one-sided polygon	cyclic
D_n	two-sided polygon	dihedral
E_6	tetrahedron	T
E_7	cube/octahedron	O
E_8	dodecahedron/icosahedron	I

[6]ChatGPT claims to have a proof of this using continuant identities for Chebyshev polynomials but it looks complicated. We will see better reasons later.

[7]The letters T, O, I are called *Schönflies notation* and are used in chemistry and physics. It turns out that $T \cong A_4$, $O \cong S_4$ and $I \cong A_5$, where S_n and A_n are the symmetric and alternating groups.

We first show that regular polyhedra are related to triples $p, q, r \in \mathbb{N}$ satisfying $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$, which is related to the idea of "positive curvature". A *spherical triangle* is a region on the surface of a sphere bounded by three great circles. Unlike in the Euclidean case, Thomas Harriot (1603) showed that the area of a spherical triangle is determined by its interior angles. Consider a triangle on the surface of a sphere with vertices A, B, C and interior angles α, β, γ:



Theorem 2.1 (Harriot's Theorem)**.** *If r is the radius of the sphere and if $\Delta_{\alpha\beta\gamma}$ is the area of the triangle with interior angles α, β, γ (measured within the spherical surface) then*

$$\Delta_{\alpha\beta\gamma} = r^2(\alpha + \beta + \gamma - \pi).$$

The quantity $\alpha + \beta + \gamma - \pi$ is called the **spherical excess** *of the triangle and it measures how far the triangle is from being Euclidean.*

Proof. The three great circles bounding the triangle divide the surface of the sphere into eight triangles, which come in pairs of equal area. Let Δ_α, Δ_β and Δ_γ denote the areas of the triangles that lie across the sides a, b and c from the angles α, β, γ, respectively. Let $S = 4\pi r^2$ denote the surface area of the sphere. Since the eight triangles cover the full sphere, we have

$$2(\Delta_{\alpha\beta\gamma} + \Delta_\alpha + \Delta_\beta + \Delta_\gamma) = 4\pi r^2.$$

On the other hand, the triangles with areas $\Delta_{\alpha\beta\gamma}$ and Δ_α glue together to form a "lune" of angle α. Since this lune covers $\alpha/2\pi$ of the whole sphere, it has area

$$\Delta_{\alpha\beta\gamma} + \Delta_\alpha = \frac{\alpha}{2\pi} S = 2r^2\alpha.$$

Similarly, we have $\Delta_{\alpha\beta\gamma} + \Delta_\beta = 2r^2\beta$ and $\Delta_{\alpha\beta\gamma} + \Delta_\gamma = 2r^2\gamma$. Adding these three equations and multiplying by 2 gives

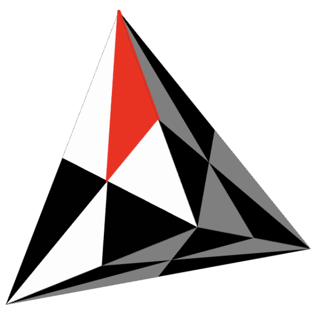
$$6\Delta_{\alpha\beta\gamma} + 2\Delta_\alpha + 2\Delta_\beta + 2\Delta_\gamma = 4r^2(\alpha + \beta + \gamma).$$

Then the first equation tells us that

$$\begin{aligned}4\Delta_{\alpha\beta\gamma} + 4\pi r^2 &= 4\Delta_{\alpha\beta\gamma} + 2(\Delta_{\alpha\beta\gamma} + \Delta_\alpha + \Delta_\beta + \Delta_\gamma)\\ &= 6\Delta_{\alpha\beta\gamma} + 2\Delta_\alpha + 2\Delta_\beta + 2\Delta_\gamma\\ &= 4r^2(\alpha+\beta+\gamma),\end{aligned}$$

and the result follows. □

What does this have to do with Platonic solids? The *barycentric subdivision* of a polytope is defined by placing a vertex at the barycenter of each face of each dimension, and then connecting a set of vertices by a simplex whenever the corresponding faces form a nested chain of subsets. (See below for more detail.) The maximal simplices of the barycentric subdivision correspond to maximal flags of faces. For example, the following figure shows the barycentric subdivision of the surface of a regular tetrahedron, and its projection onto the surface of a sphere.



The red triangle has three vertices — one at a vertex of the tetrahedron, one at the midpoint of an edge, and one at the centroid of a triangle. Around the "vertex vertex" there are 6 triangles, accounting for the fact that this vertex has 3-fold rotational symmetry. Around the "edge vertex" there are 4 triangles, accounting for the fact that the edge has 2-fold rotational symmetry, and around the "face vertex" there are again 6 triangles because the triangular face has 3-fold rotational symmetry. When we blow up the red triangle onto the sphere this tells us that the interior angles of the red spherical triangle are $\pi/3$, $\pi/2$ and $\pi/3$.

In general, each Platonic solid corresponds to a triple of integers p, q, r where each vertex, edge and face has p-, q- and r-fold rotational symmetry, respectively. When we blow up a triangle of the barycentric subdivision onto the surface of the sphere it has interior angles π/p, π/q and π/r. If the radius of the sphere is R then Harriot's formula tells us that

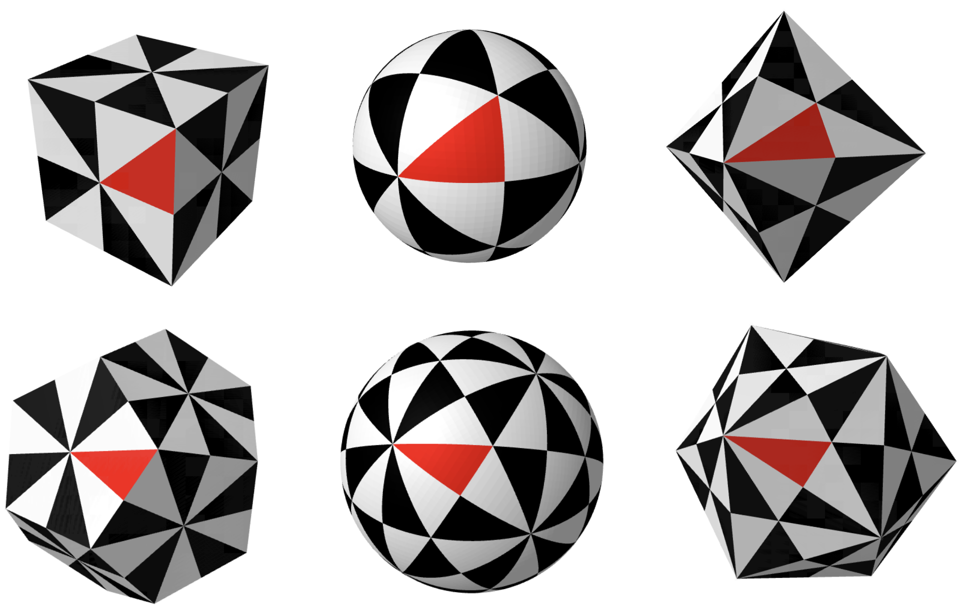
$$\begin{aligned}\text{area of red triangle} &> 0\\ R^2\left(\frac{\pi}{p}+\frac{\pi}{q}+\frac{\pi}{r}-\pi\right) &> 0\\ \frac{1}{p}+\frac{1}{q}+\frac{1}{r}-1 &> 0\end{aligned}$$

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1.$$

The regular tetrahedron corresponds to the triple $(p, q, r) = (3, 2, 3)$. The following table shows the rotational symmetries of the five Platonic solids:

shape	(vertex, edge, face) rotations
tetrahedron	$(3, 2, 3)$
cube	$(3, 2, 4)$
octahedron	$(4, 2, 3)$
dodecahedron	$(3, 2, 5)$
icosahedron	$(5, 2, 3)$

Here are pictures of the remaining four Platonic solids and their projections onto a sphere. Notice that dual pairs share the same projection.



As we mentioned above, the concept of regular polytopes applies in all dimensions, and these have also been classified. According to Coxeter (*Regular Polytopes*, 1948) this classification was independently discovered **at least eight times**, with the earliest discovery due to Schläfli in 1852. Here I will sketch the general classification and we will complete the details in a later chapter. We will approach the classification via groups of symmetries.

A map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is called an *isometry* if it preserves distance:

$$\|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\| \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Given a polytope $P \subseteq \mathbb{R}^n$,[8] we define a *symmetry* of P to be an isometry of $\mathbb{R}^n$ with the property that $\varphi(P) = P$ as sets. The collection of symmetries of P forms a group, which we denote by

$$\text{Sym}(P) = \{\text{isometries } \varphi : \mathbb{R}^n \to \mathbb{R}^n \text{ such that } \varphi(P) = P\}.$$

Basic convex geometry tells us that $\text{Sym}(P)$ permutes the faces of each dimension. We say that a face $F \subseteq P$ is *stabilized* by φ if $\varphi(F) = F$ as sets.[9] The group $\text{Sym}(P)$ also acts on flags of faces of P with fixed dimensions. A *maximal flag* is a nested chain of faces, with one of each dimension:

$$\Phi = \{F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n\},$$

where $\dim(F_d) = d$ and $F_n = P$.[10] Given any $\varphi \in \text{Sym}(G)$ we define the flag $\varphi(\Phi)$ by

$$\varphi(\Phi) = \{\varphi(F_0) \subseteq \varphi(F_1) \subseteq \cdots \subseteq \varphi(F_n)\}.$$

We say that P is *regular* when the group $\text{Sym}(P)$ acts transitively on the set of maximal flags, which means that for any two maximal flags Φ, Φ' there exists at least one group element $\varphi \in \text{Sym}(P)$ such that $\Phi' = \varphi(\Phi)$. I claim that this group element is unique.

To prove this we will use the *barycentric subdivision* of P, defined as follows. To each face $F \subseteq P$ we associate its barycenter (center of mass) $\mathbf{x}_F \in \mathbb{R}^n$. Then to each flag of faces we associate the convex hull of its barycenters:

$$\Phi = \{F_{d_1} \subseteq \cdots \subseteq F_{d_k}\} \quad \rightsquigarrow \quad \Delta_\Phi := \text{conv}\{\mathbf{x}_{F_{i_1}}, \ldots, \mathbf{x}_{F_{i_d}}\}.$$

One can show that the barycenters of a flag are affinely independent, so if Φ is a chain with $k+1$ faces then Δ_Φ is a simplex of dimension k. The collection of simplices corresponding to flags of faces of P is a simplicial complex called the barycentric subdivision:

$$\Delta(P) = \{\text{simplices } \Delta_\Phi \text{ where } \Phi \text{ is a flag of faces of } P\}.$$

The figures above display the barycentric subdivisions of the Platonic solids. If a face $F \subseteq P$ is stabilized by some isometry φ then its barycenter must be **fixed** by φ:

$$\varphi(F) = F \quad \Longrightarrow \quad \varphi(\mathbf{x}_F) = \mathbf{x}_F.$$

Indeed suppose that F has m vertices $\mathbf{v}_1, \ldots, \mathbf{v}_m$. It follows from convex geometry that φ permutes these vertices. Furthermore, we will prove in the next section (see the Isometry

[8] A polytope can defined as an intersection of finitely many closed half spaces, or, equivalently, as the convex hull of a finite set of points. A face of a polytope the intersection of the polytope with some closed half space. In this course will will accept basic facts of convex geometry without proof.

[9] This is different from the idea of being *fixed* by φ which means that $\varphi(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in F$.

[10] Some sources also define $F_{-1} = \emptyset$ because the empty set satisfies the definition of a "face", i.e., the intersection of P with some half space.

Theorem) that every isometry is an affine linear map. Since the barycenter is the average of the vertices and since affine linear maps preserve affine combinations, we have

$$\varphi(\mathbf{x}_F) = \varphi\left(\frac{1}{m}\sum_{i=1}^{m}\mathbf{v}_i\right) = \frac{1}{m}\sum_{i=1}^{m}\varphi(\mathbf{v}_i) = \frac{1}{m}\sum_{i=1}^{m}\mathbf{v}_i = \mathbf{x}_F.$$

If a maximal flag $\Phi = \{F_0 \subseteq \cdots \subseteq F_n\}$ is stabilized by some isometry φ then it follows that φ fixes the barycenters $\mathbf{x}_0, \ldots, \mathbf{x}_n$. But an affine linear map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ that fixes $n+1$ affinely independent points must be the identity map. Finally, suppose we have $\varphi(\Phi) = \gamma(\Phi)$ for some maximal flag Φ and symmetries $\varphi, \gamma \in \mathrm{Sym}(P)$. By the previous remarks this implies that $\varphi = \gamma$ because

$$\varphi(\Phi) = \gamma(\Phi) \quad \Longrightarrow \quad \gamma^{-1}\varphi(\Phi) = \Phi \quad \Longrightarrow \quad \gamma^{-1}\varphi = \mathrm{id} \quad \Longrightarrow \quad \varphi = \gamma.$$

Thus we have proved the following result.

Theorem 2.2. *Let $P \subseteq \mathbb{R}^n$ be a regular polytope. By definition this means that* $\mathrm{Sym}(P)$ *acts transitively on the set of maximal faces of the barycentric subdivision $\Delta(P)$. In fact, this action is* **simplify transitive** *and by choosing a fixed maximal face $\Delta_0 \in \Delta(P)$ we obtain a bijection between group elements and maximal faces:*

$$\begin{array}{rcl} \{\textit{elements of } \mathrm{Sym}(P)\} & \longleftrightarrow & \{\textit{maximal faces of } \Delta(P)\} \\ \varphi & \longleftrightarrow & \varphi(\Delta_0). \end{array}$$

In the pictures of Platonic solids above, the maximal faces are the black and white triangles and the fixed face Δ_0 is the red triangle. Every isometry either preserves orientation or reverses orientation.[11] The collection of orientation preserving isometries is a subgroup of index 2, which denote by $\mathrm{Sym}^+(P) \subseteq \mathrm{Sym}(P)$. The elements of $\mathrm{Sym}^+(P)$ correspond to the black triangles. In this chapter we are only interested in the group $\mathrm{Sym}^+(P)$. In a future chapter we will show that the full symmetry group $\mathrm{Sym}(P)$ is a Coxeter group and we will use the classification of Coxeter groups to derive the classification of regular polytopes. Here is the classification:

Coxeter group	regular polytope
A_n	hypersimplex
B_n/C_n	hypercube/hyperoctahedron
F_4	24-cell
H_3	dodecahedron/icosahedron
H_4	120-cell / 600-cell
$I_2(m)$	regular m-gon

[11] The Isometry Theorem in the next section shows that every isometry has the form $\varphi(\mathbf{x}) = A\mathbf{x} + \mathbf{t}$ for some real orthogonal matrix $A^TA = I$ and vector $\mathbf{t}$. We note that φ is orientation preserving or reversing precisely when $\det(A)$ equals $+1$ or -1, respectively.

I won't explain the details now but I note that the subscript indicates the dimension of the polytope. Thus the Platonic solids correspond to the Coxeter groups of types $A_3, B_3/C_3$ and H_3. This might seem strange because I already said that they correspond to types E_6, E_7, E_8. This is because the Platonic solids participate in **two different ADE type classifications**.[12]

As an amusing consequence of the previous remarks, let $G = \mathrm{Sym}^+(P)$ be the group of orientation preserving symmetries of a Platonic solid corresponding to parameters $p, q, r \geq 1$ with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$. Consider the projection of the barycentric subdivision onto a sphere of radius R. By regularity this divides the surface of the sphere into congruent spherical triangles with interior angles $\pi/p, \pi/q, \pi/r$ and by Harriot's Theorem each triangle has area $R^2(\pi/p+\pi/q+\pi/r-\pi)$. The black triangles correspond to G and there is a bijection between black and white triangles (the two cosets of G have the same size), hence the total number of triangles is $2 \cdot \#G$. Since the whole sphere has surface area $4\pi R^2$ we obtain

$$2 \cdot \#G \cdot R^2(\pi/p + \pi/q + \pi/r - \pi) = 4\pi R^2$$
$$\#G = \frac{2}{\frac{1}{p} + \frac{1}{q} + \frac{1}{r} - 1}.$$

2.2 Isometries are generated by reflections

In this section we give a modern analytic treatment of the concept of "symmetry". By a "symmetry" of an object $X \subseteq \mathbb{R}^n$ we mean an "isometry" of $\mathbb{R}^n$ that preserves X setwise. To be precise, we consider n-dimensional Euclidean space $\mathbb{R}^n$ with respect the standard dot product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$. This allows us to define distances by $\|\mathbf{x} - \mathbf{y}\|^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle$. We say that a map $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is an *isometry* when

$$\|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\| \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Of course the notion of isometry can also be defined in synthetic geometry, without reference to coordinates and inner products. The following fundamental theorem is the connection between synthetic Euclidean geometry and analytic Euclidean geometry.[13]

Theorem 2.3 (The Isometry Theorem)**.** *If $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry then there exists a real orthogonal matrix $A^TA = I$ and a vector $\mathbf{t} \in \mathbb{R}^n$ such that $\varphi(\mathbf{x}) = A\mathbf{x} + \mathbf{t}$ for all $\mathbf{x} \in \mathbb{R}^n$.*

[12] This might be confusing at the moment but I didn't want to hide it from you, since this was a natural moment to mention the classification of regular polytopes.

[13] This is similar to the Fundamental Theorem of Projective Geometry, which says that the group of bijective collineations (maps sending lines to lines) of $\mathbb{RP}^n$ is isomorphic to $PGL_{n+1}(\mathbb{R})$ acting on homogeneous coordinates. The restriction of this result to affine space $\mathbb{R}^n \subseteq \mathbb{RP}^n$ is says that any bijective collineation $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ has the form $\varphi(\mathbf{x}) = A\mathbf{x} + \mathbf{t}$ for some invertible matrix $A \in GL_n(\mathbb{R})$ and vector $\mathbf{t} \in \mathbb{R}^n$. Thus the symmetries of Euclidean geometry form a subgroup of the symmetries of projective geometry. It was Felix Klein who advocated the study of geometry in terms of groups of symmetries.

Proof. Let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be an isometry and define $f(\mathbf{x}) := \varphi(\mathbf{x}) - \varphi(\mathbf{0})$ so that $f(\mathbf{0}) = \mathbf{0}$. We will show that $f(\mathbf{x}) = A\mathbf{x}$ for some real orthogonal matrix $A^T A = I$. If we also define $\mathbf{t} := \varphi(\mathbf{0})$ then it will follow that $\varphi(\mathbf{x}) = f(\mathbf{x}) + \varphi(\mathbf{0}) = A\mathbf{x} + \mathbf{t}$.

We note that f is an isometry since for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ we have

$$\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\varphi(\mathbf{x}) - \varphi(\mathbf{0}) - (\varphi(\mathbf{y}) - \varphi(\mathbf{0}))\| = \|\varphi(\mathbf{x}) - \varphi(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|.$$

This implies that $\|f(\mathbf{x})\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$ because

$$\|f(\mathbf{x})\| = \|f(\mathbf{x}) - \mathbf{0}\| = \|f(\mathbf{x}) - f(\mathbf{0})\| = \|\mathbf{x} - \mathbf{0}\| = \|\mathbf{x}\|.$$

Now from the polarization identity we observe that f preserves inner products:

$$\begin{aligned}\langle f(\mathbf{x}), f(\mathbf{y})\rangle &= (\|f(\mathbf{x})\|^2 + \|f(\mathbf{y})\|^2 - \|f(\mathbf{x}) - f(\mathbf{y})\|^2)/2 \\ &= (\|\mathbf{x}\|^2 + \|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2)/2 \\ &= \langle \mathbf{x}, \mathbf{y}\rangle.\end{aligned}$$

Substituting standard basis vectors for $\mathbf{x}$ and $\mathbf{y}$ shows that $\langle f(\mathbf{e}_i), f(\mathbf{e}_j)\rangle = \langle \mathbf{e}_i, \mathbf{e}_j\rangle = \delta_{ij}$, hence the set $\{f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)\} \subseteq \mathbb{R}^n$ is orthonormal; in particular, it is a basis of $\mathbb{R}^n$.

Next we show that $f : \mathbb{R}^n \to \mathbb{R}^n$ must be a linear map. For any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and standard basis vector $\mathbf{e}_i \in \mathbb{R}^n$ we have

$$\begin{aligned}\langle f(\mathbf{x} + \mathbf{y}), f(\mathbf{e}_i)\rangle &= \langle \mathbf{x} + \mathbf{y}, \mathbf{e}_i\rangle \\ &= \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{e}_i\rangle \\ &= \langle f(\mathbf{x}), f(\mathbf{y})\rangle + \langle f(\mathbf{y}), f(\mathbf{e}_i)\rangle \\ &= \langle f(\mathbf{x}) + f(\mathbf{y}), f(\mathbf{e}_i)\rangle,\end{aligned}$$

and hence $\langle f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - f(\mathbf{y}), f(\mathbf{e}_i)\rangle = 0$. Since the vectors $\{f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)\}$ are a basis for $\mathbb{R}^n$, it follows from non-degeneracy of the inner product that $f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - f(\mathbf{y}) = \mathbf{0}$ and hence $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. By induction this implies that $f(n\mathbf{x}) = nf(\mathbf{x})$ for all integers $n \in \mathbb{Z}$. Then for all rational numbers $m/n \in \mathbb{Q}$ we have

$$nf((m/n)\mathbf{x}) = f(n(m/n)\mathbf{x}) = f(m\mathbf{x}) = mf(\mathbf{x}) \quad \Longrightarrow \quad f((m/n)\mathbf{x}) = (m/n)f(\mathbf{x}).$$

Since the rational numbers are dense in $\mathbb{R}$ and since isometries are continuous (almost by definition), it follows that $f(\alpha\mathbf{x}) = \alpha f(\mathbf{x})$ for all real α. We have shown that f is linear and hence $f(\mathbf{x}) = A\mathbf{x}$ for some real matrix A. Finally, since f preserves inner products we have

$$\mathbf{x}^T A^T A\mathbf{y} = \langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle,$$

which shows that the i, j entry of the matrix $A^T A$ is $\mathbf{e}_i^T A^T A\mathbf{e}_j = \langle \mathbf{e}_i, \mathbf{e}_j\rangle = \delta_{ij}$. $\square$

Thus we can express any isometry of $\mathbb{R}^n$ in terms of orthogonal matrices. Going further, we will show that any isometry is a composition of orthogonal reflections. Given a unit vector $\mathbf{u} \in \mathbb{R}^n$ we recall that the orthogonal reflection across the hyperplane $\mathbf{u}^\perp \subseteq \mathbb{R}^n$ is given by the reflection matrix[14]

$$F_{\mathbf{u}} = I - 2\mathbf{u}\mathbf{u}^T.$$

(We use the letter F for reFlection or Flip. We save the letter R for Rotations in the next section.) To verify this, we note that

$$F_{\mathbf{u}}\mathbf{u} = (I - 2\mathbf{u}\mathbf{u}^T)\mathbf{u} = \mathbf{u} - 2\mathbf{u}(\mathbf{u}^T\mathbf{u}) = \mathbf{u} - 2\mathbf{u}(1) = \mathbf{u} - 2\mathbf{u} = -\mathbf{u},$$

and for any vector $\mathbf{v}$ perpendicular to $\mathbf{u}$ we note that

$$F_{\mathbf{u}}\mathbf{v} = (I - 2\mathbf{u}\mathbf{u}^T)\mathbf{v} = \mathbf{v} - 2\mathbf{u}(\mathbf{u}^T\mathbf{v}) = \mathbf{v} - 2\mathbf{u}(0) = \mathbf{v}.$$

More generally, for any unit vector $\mathbf{u} \in \mathbb{R}^n$ and real number $r \in \mathbb{R}$ we let $F_{\mathbf{u},r}$ denote the reflection across the affine hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{u}^T\mathbf{x} = r\}$, so that $F_{\mathbf{u},0}$ is a linear function with matrix $F_{\mathbf{u}}$. We note that $F_{\mathbf{u},r}$ is an affine linear function:

$$F_{\mathbf{u},r}(\mathbf{x}) = F_{\mathbf{u}}\mathbf{x} + 2r\mathbf{u} = (I - 2\mathbf{u}\mathbf{u}^T)\mathbf{x} + 2r\mathbf{u} = \mathbf{x} - 2(\mathbf{u}^T\mathbf{x} - r)\mathbf{u}.$$

Indeed, for any point $\mathbf{x}$ satisfying $\mathbf{u}^T\mathbf{x} = r$ we verify that

$$F_{\mathbf{u},r}(\mathbf{x}) = \mathbf{x} - 2(r - r)\mathbf{u} = \mathbf{x}.$$

In particular we have $F_{\mathbf{u},r}(r\mathbf{u}) = r\mathbf{u}$. We also note that $F_{\mathbf{u},r}$ swaps the points $\mathbf{0}$ and $2r\mathbf{u}$. The following result is named after Cartan and Dieudonné who proved a much more general version over arbitrary fields. The case of Euclidean isometries was likely known earlier.

Theorem 2.4 (The Cartan-Dieudonné Theorem)**.** *Every isometry of $\mathbb{R}^n$ can be expressed as a composition of k reflections, where $k \leq n+1$. We can assume that all but one of these reflections is linear.*

Proof. Consider any isometry $\varphi : \mathbb{R}^n \to \mathbb{R}^n$. If $\mathbf{u} := \varphi(\mathbf{0})/\|\varphi(\mathbf{0})\|$ and $r := \|\varphi(\mathbf{0})\|/2$ then we note that the affine reflection $F_{\mathbf{u},r}$ swaps the points $\mathbf{0}$ and $\varphi(\mathbf{0})$. Thus the isometry $f := F_{\mathbf{u},r} \circ \varphi$ fixes the origin. By the previous theorem this means that $f(\mathbf{x}) = A\mathbf{x}$ for some real orthogonal matrix $A^TA = I$. Since $\varphi = F_{\mathbf{u},r} \circ f$ we are reduced to proving that any orthogonal matrix A can be expressed as a product of at most n reflection matrices.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of $\mathbb{R}^n$ and consider the unit vector $\mathbf{u}_1 := (A\mathbf{e}_1 - \mathbf{e}_1)/\|A\mathbf{e}_1 - \mathbf{e}_1\|$. We observe that the reflection matrix $F_{\mathbf{u}_1}$ swaps $A\mathbf{e}_1$ and $\mathbf{e}_1$ hence the matrix $A_1 := F_{\mathbf{u}_1}A$ fixes $\mathbf{e}_1$. Since $A_1^TA_1 = I$ this implies that

$$A_1\mathbf{e}_1 = \mathbf{e}_1$$

[14]In computational linear algebra these are called *Householder matrices*.

$$\begin{aligned} A_1^T A_1 \mathbf{e}_1 &= A_1^T \mathbf{e}_1 \\ \mathbf{e}_1 &= A_1^T \mathbf{e}_1. \end{aligned}$$

Then A_1 stabilizes the orthogonal complement $\mathbf{e}_1^\perp$ because if $\mathbf{e}_1^T \mathbf{x} = 0$ then we have

$$\mathbf{e}_1^T (A_1 \mathbf{x}) = (A_1^T \mathbf{e}_1)^T \mathbf{x} = \mathbf{e}_1^T \mathbf{x} = 0.$$

Next we define $\mathbf{u}_2 := (A_1\mathbf{e}_2 - \mathbf{e}_2)/\|A_1\mathbf{e}_2 - \mathbf{e}_2\|$ and $A_2 := F_{\mathbf{u}_2} A_1$, so that A_2 fixes $\mathbf{e}_2$. We observe that A_2 also fixes $\mathbf{e}_1$ because $A_2\mathbf{e}_1 = F_{\mathbf{u}_2} A_1 \mathbf{e}_1 = F_{\mathbf{u}_2}\mathbf{e}_1$ and $\mathbf{u}_2$ is perpendicular to $\mathbf{e}_1$. Thus by a similar argument to the previous, we see that A_2 stabilizes the space spanned by $\mathbf{e}_3, \ldots, \mathbf{e}_n$. If we recursively define $\mathbf{u}_i := (A_{i-1}\mathbf{e}_i - \mathbf{e}_i)/\|A_{i-1}\mathbf{e}_i - \mathbf{e}_i\|$ and $A_i := F_{\mathbf{u}_i} A_{i-1}$ then it follows that A_n fixes the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and hence $A_n = I$. In summary, we have

$$\begin{aligned} F_{\mathbf{u}_n} \cdots F_{\mathbf{u}_1} A &= I \\ A &= F_{\mathbf{u}_1} \cdots F_{\mathbf{u}_n}, \end{aligned}$$

hence A is a product of at most n reflection matrices. (It is possible that some $\mathbf{u}_i$ is the zero vector so $F_{\mathbf{u}_i}$ is the identity matrix.) □

We remark that the same algorithm can be applied to an arbitrary matrix A, resulting in a factorization $A = F_{\mathbf{u}_1} \cdots F_{\mathbf{u}_n} R$ where R is just upper triangular. Taking $Q = F_{\mathbf{u}_1} \cdots F_{\mathbf{u}_n}$ gives the QR-factorization of A. This algorithm was described by Householder in the short paper *Unitary Triangularization of a Nonsymmetric Matrix* (1958), which is why reflection matrices are called "Householder matrices" in numerical linear algebra.

We also remark that the factorization of an orthogonal matrix into reflections is not unique. In fact, there are infinitely many different factorizations for each matrix. Say that a factorization of A is *minimal* when it uses the minimal possible number of reflections. Brady and Watt (*A partial order on the orthogonal group*, 2001) proved that this number is the dimension of the "moved space" $\mathrm{im}\,(A - I)$ and they gave a bijection between minimal factorization of A into reflections and maximal flags of subspaces of $\mathrm{im}\,(A - I)$.

2.3 Euler's rotation theorem

Next we apply the results of the previous section to the study of rotations. Consider the *orthogonal group*:

$$O(n) = \{\text{real } n \times n \text{ matrices } A \text{ satisfying } A^T A = I\}.$$

By the Isometry Theorem this group is isomorphic to the group of isometries of $\mathbb{R}^n$ preserving the origin (or any fixed point). The determinant of an orthogonal matrix is $+1$ or -1 because

$$1 = \det(I) = \det(A^T A) = \det(A^T)\det(A) = \det(A)\det(A) = \det(A)^2.$$

We say that the isometry A is orientation preserving when $\det(A) = 1$ and orientation reversing when $\det(A) = -1$. The orientation preserving isometries form a subgroup, called the *special orthogonal group*:

$$SO(n) = \{A \in O(n) : \det(A) = 1\}.$$

Since the determinant of a reflection matrix is -1, the Cartan-Dieudonné Theorem gives us another point of view on these groups:

$$\begin{aligned} O(n) &= \{\text{products of reflection matrices}\} \\ SO(n) &= \{\text{products of even numbers of reflection matrices}\}. \end{aligned}$$

I claim we can rephrase the second statement as

$$SO(n) = \{\text{products of rotation matrices}\}.$$

The concept of a "rotation" is a bit more subtle than the concept of a "reflection". The data of a rotation is an angle and an oriented 2-plane.[15]

Definition 2.5. Given an ordered pair of orthonormal vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and an angle $\theta \in \mathbb{R}$ we define the matrix $R_{\mathbf{u},\mathbf{v}}(\theta)$ by

$$\begin{cases} R_{\mathbf{u},\mathbf{v}}(\theta)\mathbf{u} &= \cos\theta\mathbf{u} + \sin\theta\mathbf{v}, \\ R_{\mathbf{u},\mathbf{v}}(\theta)\mathbf{v} &= -\sin\theta\mathbf{u} + \cos\theta\mathbf{v}, \end{cases}$$

and $R_{\mathbf{u},\mathbf{v}}(\theta)\mathbf{w} = \mathbf{w}$ for all vectors satisfying $\mathbf{u}^T\mathbf{w} = 0$ and $\mathbf{v}^T\mathbf{w} = 0$. If $\mathbf{w}_1, \ldots, \mathbf{w}_{n-2}$ is any orthonormal basis for the orthogonal complement of the plane $\mathbb{R}\mathbf{u} + \mathbb{R}\mathbf{v}$ and U is the orthogonal matrix with columns $\mathbf{u}, \mathbf{v}, \mathbf{w}_1, \ldots, \mathbf{w}_{n-2}$ then we have

$$R_{\mathbf{u},\mathbf{v}}(\theta) = U \left(\begin{array}{cc|c} \cos\theta & -\sin\theta & \\ \sin\theta & \cos\theta & \\ \hline & & I \end{array}\right) U^T.$$

It follows from this that $R_{\mathbf{u},\mathbf{v}}(\theta)$ is an orthogonal matrix with determinant $+1$ and trace $2\cos\theta + n - 2$. The following beautiful expression called the *Euler-Rodrigues formula*:

$$R_{\mathbf{u},\mathbf{v}}(\theta) = I + \sin\theta(\mathbf{v}\mathbf{u}^T - \mathbf{u}\mathbf{v}^T) + (\cos\theta - 1)(\mathbf{u}\mathbf{u}^T + \mathbf{v}\mathbf{v}^T).$$

It can be proved by observing that the matrix on the right hand side agrees with $R_{\mathbf{u},\mathbf{v}}(\theta)$ on the basis $\mathbf{u}, \mathbf{v}, \mathbf{w}_1, \ldots, \mathbf{w}_{n-2}$. Unfortunately I don't know an intuitive elementary derivation of this formula.[16] This is why I say that rotations are more subtle than reflections.

[15] Some authors call this a "simple rotation" and use the word "rotation" for any element of $SO(n)$.

[16] A higher level derivation views the matrix $K := \mathbf{v}\mathbf{u}^T - \mathbf{u}\mathbf{v}^T$ as an "infinitesimal rotation". Then it uses the formula $-K^2 = \mathbf{u}\mathbf{u}^T + \mathbf{v}\mathbf{v}^T$ to show that $R_{\mathbf{u},\mathbf{v}}(\theta)$ is the matrix exponential $\exp(\theta K)$.

Here is the key result relating rotations and reflections.

Proposition 2.6. *A matrix R is a rotation matrix if and only if it is a product of two reflection matrices. Furthermore, the angle of the rotation is twice the angle between the reflecting hyperplanes.*

Proof. Recall the definition of the reflection matrix $F_{\mathbf{u}} = I - 2\mathbf{u}\mathbf{u}^T$ for a unit vector $\mathbf{u}$. Given an ordered pair of vectors $\mathbf{u}, \mathbf{v}$ satisfying $\mathbf{u}^T\mathbf{u} = \mathbf{v}^T\mathbf{v} = 1$ and $\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = 0$ and an angle θ we note that $\cos(\theta/2)\mathbf{u} + \sin(\theta/2)\mathbf{v}$ is also unit vector. Then one can show using the Euler-Rodrigues formula and brute force[17] that

$$R_{\mathbf{u},\mathbf{v}}(\theta) = F_{\mathbf{w}}F_{\mathbf{u}}, \qquad \text{where } \mathbf{w} = \cos(\theta/2)\mathbf{u} + \sin(\theta/2)\mathbf{v}.$$

Thus any rotation is a product of two reflections. Conversely, given any two unit vectors $\mathbf{u}, \mathbf{w}$ with $\mathbf{u}^T\mathbf{w} = \cos(\theta/2)$, there exists a unique unit vector $\mathbf{v}$ satisfying $\mathbf{w} = \cos(\theta/2)\mathbf{u}+\sin(\theta/2)\mathbf{v}$ and $\mathbf{u}^T\mathbf{v} = 0$, and we conclude that the product $F_{\mathbf{w}}F_{\mathbf{u}}$ equals the rotation $R_{\mathbf{u},\mathbf{v}}(\theta)$. □

Since the group $O(n)$ consists of matrices that can be expressed as a product of k reflection matrices with $k \leq n$, and the subgroup $SO(n)$ consists of matrices that can be expressed as a product of an even number of reflection matrices, we obtain the following corollary.

Corollary 2.7. *Every matrix $A \in SO(n)$ is product of k rotation matrices, with $k \leq \lfloor n/2 \rfloor$.*

The $n = 3$ case is a result of Euler from 1776.

Theorem 2.8 (Euler's Rotation Theorem)**.** *Every orientation preserving isometry of $\mathbb{R}^3$ that fixes the origin is a rotation.*

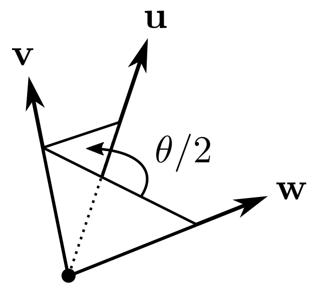
Proof. The Isometry Theorem says that isometries fixing the origin are the same as orthogonal matrices, hence the group of orientation preserving isometries fixing the origin is $SO(3)$. The previous result then tells us that every element of $SO(3)$ can be expressed as a product of 0 or 1 rotation matrices. □

There are certainly easier ways to prove Euler's theorem but we took this opportunity to develop the general version in n-dimensional space because we will use it later. A corollary of Euler's theorem is that the product of two rotations in $\mathbb{R}^3$ is again a rotation, which is certainly not visually obvious. I end this section by giving a nice visual explanation that I learned from John Stillwell's *The Four Pillars of Geometry* (2005, Section 7.5).

We will change the notation slightly to fit the three dimensional geometry. Given any vector $\mathbf{u} \in \mathbb{R}^3$ and angle θ let $R_{\mathbf{v}}(\theta)$ denote the rotation of $\mathbb{R}^3$ counterclockwise by angle $\theta \in \mathbb{R}$ around the oriented line $\mathbb{R}\mathbf{v}$. Given any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ let $F_{\mathbf{uv}}$ denote the orthogonal reflection across the plane $\mathbb{R}\mathbf{u}+\mathbb{R}\mathbf{v}$. Now let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ be any three vectors and let $\theta/2$ be

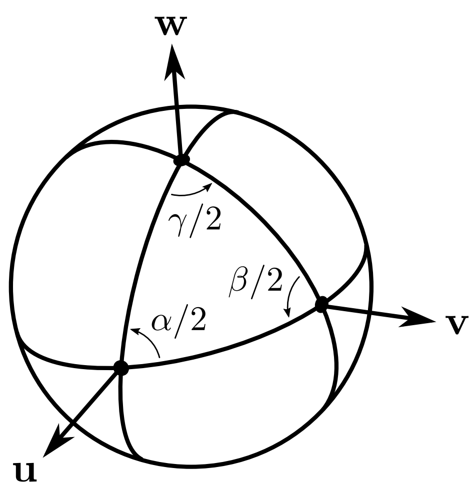
[17] Again, I don't know a really intuitive proof of this formula.

the dihedral angle between the planes $\mathbb{R}\mathbf{u}+\mathbb{R}\mathbf{v}$ and $\mathbb{R}\mathbf{v}+\mathbb{R}\mathbf{w}$, measured at $\mathbf{v}$. Then it follows from Proposition 2.6 the composition of $F_{\mathbf{vw}}$ followed by $F_{\mathbf{vu}}$ equals the rotation $R_{\mathbf{v}}(\theta)$:



$$F_{\mathbf{uv}}F_{\mathbf{vw}} = R_{\mathbf{v}}(\theta)$$

We will use this to compute the product of any two rotations. Consider any vectors $\mathbf{u}, \mathbf{v}$ and angles $\alpha, \beta \in [0, 2\pi)$ so that $\alpha/2, \beta/2 \in [0, \pi)$. Consider the points where the rays $\mathbb{R}_{>0}\mathbf{u}$ and $\mathbb{R}_{>0}\mathbf{v}$ intersect the unit sphere. (The radius of the sphere is not important.) Then there exists a unique third point on the sphere defined by some vector $\mathbf{w}$, and a unique angle $\gamma \in [0, \pi)$ such that the spherical triangle with these three vertices has internal angles $\alpha/2$, $\beta/2$, $\gamma/2$:



Now we consider the reflections $F_{\mathbf{uv}}, F_{\mathbf{vw}}, F_{\mathbf{wu}}$ across the three sides of the triangle. According to the previous remark, we have

$$\begin{aligned} R_{\mathbf{u}}(\alpha) &= F_{\mathbf{wu}}F_{\mathbf{uv}}, \\ R_{\mathbf{v}}(\beta) &= F_{\mathbf{uv}}F_{\mathbf{vw}}, \\ R_{\mathbf{w}}(\gamma) &= F_{\mathbf{vw}}F_{\mathbf{wu}}. \end{aligned}$$

Since every reflection is equal to its own inverse, this gives

$$\begin{aligned} R_{\mathbf{u}}(\alpha)R_{\mathbf{v}}(\beta) &= F_{\mathbf{wu}}F_{\mathbf{uv}}F_{\mathbf{uv}}F_{\mathbf{vw}} \\ &= F_{\mathbf{wu}}F_{\mathbf{vw}} \\ &= (F_{\mathbf{vw}}F_{\mathbf{wu}})^{-1} \\ &= R_{\mathbf{w}}(\gamma)^{-1} \end{aligned}$$

$$= R_{\mathbf{w}}(-\gamma).$$

Hence the composition of the counterclockwise rotations $R_{\mathbf{u}}(\alpha)$ and $R_{\mathbf{v}}(\beta)$ is the **counter-clockwise** rotation around the vector $\mathbf{w}$ by the angle γ.

2.4 Finite groups of rotations

Finally, we show that the finite subgroups of $SO(3)$ — the finite groups of rotations of $\mathbb{R}^3$ — have an ADE type classification. Recall the *orbit-stabilizer theorem.* If a finite group G acts on a set X then for each $x \in X$ we define the orbit $G(x) = \{g(x) : g \in G\} \subseteq X$ and the *stabilizer* $G_x = \{g \in X : g(x) = x\} \subseteq G$, which is a subgroup of G. If $x, y \in X$ are in the same G-orbit then their stabilizers are conjugate subgroups. Indeed, suppose $y = g(x)$. Then we have $G_y = gG_xg^{-1}$ because

$$h \in G_y \Leftrightarrow h(y) = y \Leftrightarrow h(g(x) = g(x) \Leftrightarrow g^{-1}hg(x) = x \Leftrightarrow g^{-1}hg \in G_x \Leftrightarrow h \in gG_xg^{-1}.$$

In particular, this implies that for any elements $x, y \in X$ in the same G-orbit we have $\#G_x = \#G_y$. For each $x \in X$ one can check that the map $g(x) \mapsto gG_x$ is a well-defined bijection between the orbit $G(x)$ and left cosets of the stabilizer G/G_x:

$$\begin{array}{rcl} G(x) & \longleftrightarrow & G/G_x \\ g(x) & \longleftrightarrow & gG_x. \end{array}$$

It follows from Lagrange's theorem that $\#G(x) = \#G/\#G_x$.

Now let $G \subseteq SO(3)$ be any non-trivial finite subgroup of $SO(3)$. We consider the action of G on the points of the unit sphere $S^2 \subseteq \mathbb{R}^3$. Recall from Euler's Rotation Theorem that every non-identity element $g \in G$ is a rotation, hence it fixes exactly two points of S^2, which we call the *poles* of g. Let $P \subseteq S^2$ denote the set of poles of all non-identity elements of G, which is a finite set. For each pole $\mathbf{p}$ we note that the stabilizer $G_{\mathbf{p}}$ is a cyclic group. Suppose that G divides the set of poles into m orbits and choose one pole $\mathbf{p}_1, \ldots, \mathbf{p}_m$ from each orbit, so that

$$P = G(\mathbf{p}_1) \sqcup G(\mathbf{p}_2) \sqcup \cdots \sqcup G(\mathbf{p}_m).$$

In this case I claim that

$$2(\#G - 1) = \sum_{i=1}^{m} \#G(\mathbf{p}_i)(\#G_{\mathbf{p}_i} - 1).$$

To see this we count the following set of pairs in two different ways:

$$\{(g, \mathbf{p}) : g \in G \setminus \{\mathrm{id}\}, \mathbf{p} \in P, g(\mathbf{p}) = \mathbf{p}\}.$$

On the one hand, for each of the $\#G - 1$ non-identity elements $g \in G$ there are exactly two poles, hence the number of pairs is $2(\#G - 1)$. On the other hand, for any pole $\mathbf{p}$ in the orbit $G(\mathbf{p}_i)$, the set of non-identity group elements satisfying $g(\mathbf{p}) = \mathbf{p}$ ist just the stabilizer $G_{\mathbf{p}} \setminus \{\text{id}\}$. But since $\mathbf{p} \in G(\mathbf{p}_i)$ we have $\#G_{\mathbf{p}} = \#G_{\mathbf{p}_i}$. Hence the number of pairs is

$$\sum_{\mathbf{p}\in P}(\#G_{\mathbf{p}} - 1) = \sum_{i=1}^{m} \sum_{\mathbf{p}\in G(\mathbf{p}_i)} (\#G_{\mathbf{p}_i} - 1) = \sum_{i=1}^{m} \#G(\mathbf{p}_i)(\#G_{\mathbf{p}_i} - 1)$$

Now we apply the orbit stabilizer theorem, which says that $\#G(\mathbf{p}_i) \cdot \#G_{\mathbf{p}_i} = \#G$, to get

$$\begin{aligned} 2(\#G - 1) &= \sum_{i=1}^{m} \#G(\mathbf{p}_i)(\#G_{\mathbf{p}_i} - 1) \\ &= \sum_{i=1}^{m} (\#G(\mathbf{p}_i)\#G_{\mathbf{p}_i} - \#G(\mathbf{p}_i)) \\ &= \sum_{i=1}^{m} (\#G - \#G/\#G_{\mathbf{p}_i}) \\ &= \#G \sum_{i=1}^{m} (1 - 1/\#G_{\mathbf{p}_i}) \\ &= \#G(m - \sum_{i=1}^{m} 1/\#G_{\mathbf{p}_i}), \end{aligned}$$

and hence

$$2 - \frac{2}{\#G} = \sum_{i=1}^{m} \left(1 - \frac{1}{\#G_{\mathbf{p}_i}}\right).$$

The case $m = 1$ is impossible since then we would have

$$1 \le 2 - \frac{2}{\#G} = 1 - \frac{1}{\#G_{\mathbf{p}_1}} < 1.$$

I claim that $m \ge 4$ is also impossible. To see this we note that $\#G_{\mathbf{p}_i} \ge 2$ for all i, since by definition we assume that each pole is fixed by at least one non-identity element. This implies that $(1 - 1/\#G_{\mathbf{p}_i}) \ge 1/2$ for all i and hence

$$2 - \frac{2}{\#G} \ge \frac{m}{2}.$$

But if $m \ge 4$ then we obtain the contradiction

$$2 > \left(2 - \frac{2}{\#G}\right) \ge \frac{m}{2} \ge \frac{4}{2} \ge 2.$$

Hence we are left with the cases $m = 2$ and $m = 3$. In the case of two orbits we have

$$2 - \frac{2}{\#G} = \left(1 - \frac{1}{\#G_{\mathbf{p}_1}}\right) + \left(1 - \frac{1}{\#G_{\mathbf{p}_2}}\right)$$

$$\frac{2}{\#G} = \frac{1}{\#G_{\mathbf{p}_1}} + \frac{1}{\#G_{\mathbf{p}_2}}.$$

Since $\#G_{\mathbf{p}_1} \le \#G$ and $\#G_{\mathbf{p}_1} \le \#G$ this implies that $G = G_{\mathbf{p}_1} = G_{\mathbf{p}_2}$. Each of the two orbits contains exactly one pole: $G(\mathbf{p}_1) = \{\mathbf{p}_1\}$ and $G(\mathbf{p}_2) = \{\mathbf{p}_2\}$. (See the picture at the end of this section.) In the case of three orbits we have

$$\begin{aligned} 2 - \frac{2}{\#G} &= \left(1 - \frac{1}{\#G_{\mathbf{p}_1}}\right) + \left(1 - \frac{1}{\#G_{\mathbf{p}_2}}\right) + \left(1 - \frac{1}{\#G_{\mathbf{p}_3}}\right) \\ 1 + \frac{2}{\#G} &= \frac{1}{\#G_{\mathbf{p}_1}} + \frac{1}{\#G_{\mathbf{p}_2}} + \frac{1}{\#G_{\mathbf{p}_3}}. \end{aligned}$$

If we write $p = \#G_{\mathbf{p}_1}$, $q = \#G_{\mathbf{p}_2}$ and $r = \#G_{\mathbf{p}_3}$ then we are back to our favorite inequality:

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1 + \frac{2}{\#G} > 1.$$

Since $p, q, r \ge 2$ the only possibilities are

$$\{p, q, r\} = \{2, 2, *\}, \{2, 3, 3\}, \{2, 3, 4\} \text{ and } \{2, 3, 5\}.$$

A tedious case by case analysis shows that these groups must be the dihedral groups and the three polyhedral groups.[18] Here we regard the dihedral group as the group of **rotational** symmetries of a two-dimensional regular polygon embedded in $\mathbb{R}^3$. The usual two-dimensional reflection symmetries of the polygon are realized by 180° rotations in $\mathbb{R}^3$.

3 McKay correspondence

So far we have seen two examples of ADE type classification:

- Connected simple graphs with spectral radius less than 2.
- Finite groups of rotations of $\mathbb{R}^3$.

In each case we reduced the problem to the classification of integers p, q, r satisfying $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$. This might seem like a mere coincidence but in the late 1970s John McKay found a deeper relationship between subgroups of $SO(3)$ and the Dykin diagrams of types ADE. This "McKay correspondence" is the topic of this chapter. Before we begin I will state the theorem without yet defining all of the terms:

[18] Of course we know that the polyhedral groups exist and have the correct orbit structure. The tedious part is to show that any two groups with the correct orbit structure are isomorphic. This is analogous to a logical gap in Euclid's Elements. Euclid proved, for example, that a regular dodecahedron exists, i.e., a polyhedron having 12 regular pentagonal faces. But he did not prove that any two polyhedra having 12 regular pentagonal faces must be congruent. This gap was later filled by Cauchy's rigidity theorem.

Given a finite "polyhedral group" $G \subseteq SO(3)$ there exists a "binary polyhedral group" $G^* \in SU(2)$ with a two-to-one homomorphism $G^* \to G$, such that the character table of G^* is equal to the matrix of eigenvectors of the corresponding affine Dynkin diagram.

3.1 The spin groups

We can view the group $SO(n)$ as a compact and connected submanifold of the n^2-dimensional Euclidean space of real $n \times n$ matrices. In class I demonstrated the "belt trick", which illustrates that the fundamental group of $SO(3)$ contains an element of order two. In fact, one can prove that $\pi_1(SO(n)) = \mathbb{Z}/2\mathbb{Z}$ for all n. It follows that the unique simply connected covering space (i.e., the universal covering space) has degree 2. Moreover, the group structure of $SO(n)$ has a unique lift making this covering space into a group and the covering map into a group homomorphism. We call this the *spin group* and the *spin homomorphism*:

$$\sigma : \mathrm{Spin}(n) \to SO(n).$$

These groups were first constructed by Cartan in 1913 using the *Clifford algebra* Cl_n which is defined as the (non-commutative) real algebra generated by $n+1$ abstract symbols $\mathbf{1}, \mathbf{e}_1, \ldots, \mathbf{e}_n$ satisfying the relations $\mathbf{e}_i^2 = -\mathbf{1}$, $\mathbf{1}\mathbf{e}_i = \mathbf{e}_i\mathbf{1}$ and $\mathbf{e}_i\mathbf{e}_j = -\mathbf{e}_j\mathbf{e}_i$.[19] Using these relations one can show that Cl_n has dimension 2^n as a real vector space, with basis of *monomials*

$$\mathbf{e}_{i_1}\mathbf{e}_{i_2}\cdots\mathbf{e}_{i_k} \text{ for } 1 \leq i_1 < \cdots < i_k \leq n \text{ and } 0 \leq k \leq n.$$

(We interpret the empty product as the symbol $\mathbf{1}$). Let Cl_n^k be the subspace generated by monomials of degree k, so that $\dim \mathrm{Cl}_n^k = \binom{n}{k}$ and $\mathrm{Cl}_n = \oplus_{k=0}^n \mathrm{Cl}_n^k$. If $x \in \mathrm{Cl}_n^k$ and $y \in \mathrm{Cl}_n^\ell$ are homogeneous elements of degrees k and ℓ then the product xy need not be homogeneous, but we do have $xy \in \oplus_{i=0}^{k+\ell}\mathrm{Cl}_n^i$. It is useful to identify the symbol $\mathbf{1}$ with the real number 1 and the symbols $\mathbf{e}_1, \ldots, \mathbf{e}_n$ with the standard basis of $\mathbb{R}^n$, so that $\mathrm{Cl}_n^0 = \mathbb{R}$ and $\mathrm{Cl}_n^1 = \mathbb{R}^n$.[20]

The Clifford product of any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n = \mathrm{Cl}_n^1$ is

$$\mathbf{u}\mathbf{v} = -\mathbf{u}^T\mathbf{v} + \sum_{i<j}(u_iv_j - u_jv_i)\mathbf{e}_i\mathbf{e}_j,$$

which implies that $\mathbf{v}^2 := \mathbf{v}\mathbf{v} = -\|\mathbf{v}\|^2 \in \mathrm{Cl}_n^0 = \mathbb{R}$ for all $\mathbf{v} \in \mathbb{R}^n$. In particular, we have $\mathbf{u}^2 = -1$ and hence $\mathbf{u}^{-1} = -\mathbf{u}$ for any unit vector $\mathbf{u} \in \mathbb{R}^3$. The most important property of Clifford algebras is that we can use them to represent reflections and rotations of $\mathbb{R}^n$. Given any $\mathbf{u}, \mathbf{x} = \mathrm{Cl}_n^1\mathbb{R}^n$ one can check that the Clifford product $\mathbf{u}\mathbf{x}\mathbf{u}$ is in Cl_n^1 and is the orthogonal

[19] We refer to Baker's *Matrix Groups: An Introduction to Lie Group Theory*, Chapter 5, for the basic theory of Clifford algebras and Spin groups.

[20] More generally we can identify Cl_n^k with the exterior power $\Lambda^k(\mathbb{R}^n)$ and the monomial $\mathbf{e}_{i_1}\mathbf{e}_{i_2}\cdots\mathbf{e}_{i_k}$ with the wedge product $\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}$.

reflection of $\mathbf{x}$ across the hyperplane $\mathbf{u}^\perp$. We define the *pinor group* as the set of Clifford products of unit vectors:[21]

$$\text{Pin}(n) = \{\mathbf{u}_1 \cdots \mathbf{u}_k \in \text{Cl}_n : \mathbf{u}_1, \ldots, \mathbf{u}_k \in \text{Cl}_n^1 \text{ unit vectors and } k \geq 0\}.$$

(The empty product is 1.) This is indeed a group because $-1 = \mathbf{uu}$ is in $\text{Pin}(n)$ and for $s = \mathbf{u}_1 \cdots \mathbf{u}_k$ in $\text{Pin}(n)$ we have $s^{-1} = (-1)^k \mathbf{u}_k \cdots \mathbf{u}_1$. Note that repeated products of unit vectors in Cl_n^1 are not contained in Cl_n^1.

Next we define a group homomorphism fromPin(n) to the orthogonal group $O(n)$ by sending each unit vector $\mathbf{u}$ to the corresponding reflection matrix $F_\mathbf{u} = I - 2\mathbf{uu}^T$:

$$\begin{array}{rccc} \sigma: & \text{Pin}(n) & \to & O(n) \\ & \mathbf{u} & \mapsto & F_\mathbf{u}. \end{array}$$

For any $\mathbf{x} \in \text{Cl}_n^1 = \mathbb{R}^n$ we note that the matrix product $F_\mathbf{u}\mathbf{x}$ is equal to the Clifford product $\mathbf{uxu}$. For a general element $s = \mathbf{u}_1 \cdots \mathbf{u}_k \in \text{Pin}(n)$ we define

$$\sigma(s) := \sigma(\mathbf{u}_1) \cdots \sigma(\mathbf{u}_k) = F_{\mathbf{u}_1} \cdots F_{\mathbf{u}_k} \in O(n).$$

The function σ is well-defined since it can be computed purely in terms of the element $s \in \text{Cl}_n$, and not on the specific factorization into unit vectors:

$$\sigma(s)\mathbf{x} = F_{\mathbf{u}_1} \cdots F_{\mathbf{u}_k}\mathbf{x} = \mathbf{u}_1 \cdots \mathbf{u}_k \mathbf{x} \mathbf{u}_k \cdots \mathbf{u}_1 = (-1)^k s\mathbf{x}s^{-1} \in \text{Cl}_n^1 = \mathbb{R}^n.$$

This formula also implies that σ is a group homomorphism; we call it the *spin homomorphism*. Furthermore, the Cartan-Dieudonné Theorem implies that σ is surjective. Indeed, Cartan-Dieudonné says that any orthogonal matrix $A \in O(n)$ can be expressed as a product of reflection matrices, $A = F_{\mathbf{u}_1} \cdots F_{\mathbf{u}_k}$ for some unit vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ with $k \leq n$, which implies that $A = \sigma(\mathbf{u}_1 \cdots \mathbf{u}_k)$.

Finally, we define the *spin group* as the subgroup of Pin(n) consisting of Clifford products of even numbers of unit vectors:

$$\text{Spin}(n) = \{\mathbf{u}_1 \cdots \mathbf{u}_{2k} \in \text{Cl}_n : \mathbf{u}_1, \ldots, \mathbf{u}_{2k} \in \text{Cl}_n^1 \text{ unit vectors and } k \geq 0\}.$$

Then the spin homomorphism σ restricts to a surjective homomorphism $\text{Spin}(n) \to SO(n)$ onto the special orthogonal group. It is not net clear what this has to do with the topologically constructed universal covering group from the beginning of this section. This is harder to prove, but it turns out that the kernel of σ is just $\{\pm 1\}$ and the topological structure defined on Spin(n) as a subset of $\text{Cl}_n = \mathbb{R}^{2^n}$ realizes the spin homomorphism as the universal covering map for $SO(n)$.

[21]The term "spin" is not due to Cartan. It was added later after these groups found application in quantum mechanics. The joke notation "pin" is based on the analogy with the groups $O(n)$ and $SO(n)$.

3.2 Finite subgroups of $SU(2)$

THIS SECTION NEEDS TO BE WRITTEN

We note that Cl_2 is isomorphic to the quaternions and Spin(3) is isomorphic to $SU(2)$.

Collect the special properties of $SU(2)$ that are necessary for the proof of McKay.

3.3 The character table of a group

Let G be a finite group. A (complex) *representation* of G is a pair (V, ρ) consisting of a finite dimensional vector space V over $\mathbb{C}$ together a group homomorphism $\rho : G \to GL(V)$ from G to the group of invertible linear transformations from V to itself. If V is n-dimensional and if we choose a basis for V then we can think of each endomorphism $\rho(g)$ as an invertible $n \times n$ matrix. It is often convenient to suppress ρ in the notation and simply write $g\mathbf{v}$ instead of $\rho(g)(\mathbf{v})$. Here we imagine that $g \in G$ is a matrix and $v \in V$ is a column vector. Because of the linear properties of matrices we note that

$$g(a\mathbf{u} + b\mathbf{v}) = ag\mathbf{u} + bg\mathbf{v} \quad \text{for all } g \in G, \mathbf{u}, \mathbf{v} \in V, a, b \in \mathbb{C}.$$

When the specific action of G is understood we just say that V is a *G-module*.

Given two G-modules U and V and a $\mathbb{C}$-linear map $\varphi : U \to V$ we say that φ is a *G-linear map* (or a homomorphism of G-modules) if we have

$$\varphi g = g\varphi \quad \text{for all } g \in G.$$

Suppose that $\dim U = m$ and $\dim V = n$. If we choose bases for U and V then we can think of φ as an $n \times m$ matrix. On the left side of the equation above we think of g as an $m \times m$ matrix acting on U and on the right side of the equation we think of g as an $n \times n$ matrix acting on V. We define the "Hom set" as the collection of all G-linear maps from U to V:

$$\mathrm{Hom}_G(U, V) = \{G\text{-linear maps } \varphi : U \to V\}.$$

In fact, the Hom set is a $\mathbb{C}$-vector space. Given G-linear maps $\varphi, \psi : U \to V$ and scalars $a, b \in \mathbb{C}$ we define the G-linear map $a\varphi + b\psi$ by

$$(a\varphi + b\psi)(\mathbf{u}) := a\varphi(\mathbf{u}) + b\psi(\mathbf{u}) \quad \text{for all } \mathbf{u} \in U.$$

The most important idea in representation theory is to think of the dimension the Hom space as a sort of inner product. We will write

$$\boxed{\langle U, V\rangle := \dim \mathrm{Hom}_G(U, V).}$$

Note that $\langle U, V\rangle \leq \dim(U)\dim(V)$ since $\mathrm{Hom}_G(U, V)$ is a subspace of the space of $\mathbb{C}$-linear maps $\mathrm{Hom}_{\mathbb{C}}(U, V)$ which we can identity with the space of $\dim(V) \times \dim(U)$ matrices by choosing bases.

Given a G-module V and a $\mathbb{C}$-subspace $W \subseteq V$ we say that W is a *G-submodule* if for all $g \in G$ and $\mathbf{w} \in W$ we have $g\mathbf{w} \in W$. For any G-linear map $\varphi : U \to V$ I claim that the kernel $\ker \varphi \subseteq U$ and the image $\operatorname{im} \varphi \subseteq V$ are G-submodules. Indeed, for all $\mathbf{u} \in \ker \varphi$ we have $\varphi g\mathbf{u} = g\varphi\mathbf{u} = g\mathbf{0} = \mathbf{0}$, hence $g\mathbf{u} \in \ker \varphi$, and for all $\varphi\mathbf{u} \in \operatorname{im} \varphi$ we have $g\varphi\mathbf{u} = \varphi g\mathbf{u} \in \operatorname{im} \varphi$. A G-module is called *simple* (or *irreducible*) if its only G-submodules are $\{\mathbf{0}\}$ and itself.

Theorem 3.1 (Schur's Lemma). *Let U and V be simple G-modules. Then we have*

$$\langle U, V\rangle = \begin{cases} 1 & U \cong V, \\ 0 & U \not\cong V. \end{cases}$$

Proof. Given a G-linear map $\varphi : U \to V$ we consider the submodules $\ker \varphi \subseteq U$ and $\operatorname{im} \varphi \subseteq V$. Since U is simple we must have $\operatorname{im} \varphi = U$, in which case $\varphi = 0$, or $\ker \varphi = \{\mathbf{0}\}$, in which case φ is injective. Since V is simple we must have $\operatorname{im} \varphi = \{\mathbf{0}\}$, in which case $\varphi = 0$, or $\operatorname{im} \varphi = V$, in which case φ is surjective. We conclude that either $\varphi = 0$ or φ is an isomorphism $U \cong V$.

It remains to show that $\dim \operatorname{Hom}_G(U, U) = 1$ when U is simple. To see this, consider any G-linear map $\varphi : U \to U$. Since $\mathbb{C}$ is algebraically closed there exists an eigenvalue λ. That is, there exists a complex number $\lambda \in \mathbb{C}$ and a nonzero vector $\mathbf{u} \in U$ such that $\varphi\mathbf{u} = \lambda\mathbf{u}$. This implies that the endomorphism $\varphi - \lambda \,\mathrm{id} : U \to U$ has a nonzero kernel. But since U is simple this implies that the kernel is all of U, so that $\varphi\mathbf{u} = \lambda\mathbf{u}$ for all $\mathbf{u}$. We conclude that any G-linear map $U \to U$ is a scalar multiple of the identity map. □

Theorem 3.2 (Maschke's Theorem). *Every finite dimensional G-module is a direct sum of simple G-submodules.*

Proof. Given any G-module V and a nontrivial G-submodule $U \subseteq V$ we will show that there exists a G-submodule $W \subseteq V$ such that $V = U \oplus W$ and then it will follow by induction that V is a direct sum of simple modules. So consider any projection map $P : V \to U$, i.e., any linear map satisfying $P\mathbf{v} \in U$ for all $\mathbf{v} \in V$ and $P\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in U$. Define the G-averaged projection

$$P_G = \frac{1}{\#G} \sum_{g \in G} gPg^{-1}.$$

I claim that $hP_G = P_G h$ for all $h \in G$. Indeed, for any group element $h \in G$ the function $g \mapsto hg$ is a permutation of G, hence

$$hP_G h^{-1} = \frac{1}{\#G} \sum_{g \in G} (hg)P(hg)^{-1} = \frac{1}{\#G} \sum_{g \in G} gPg^{-1} = P_G.$$

It follows that $\ker P_G$ is a G-submodule of V.

On the other hand, I claim that $V = U \oplus \ker P_G$. To see this we observe that P_G satisfies $P_G\mathbf{v} \in U$ for all $\mathbf{v} \in V$ and $P_G\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in U$. Indeed, for any $g \in G$ and $\mathbf{v} \in V$ we have $Pg^{-1}\mathbf{v} \in U$ and hence $gPg^{-1}\mathbf{v} \in U$ since U is a G-submodule. It follows from this that $P_G\mathbf{v} \in U$. Furthermore, for all $g \in G$ and $\mathbf{u} \in U$ we have $g^{-1}\mathbf{u} \in U$ because U is a G-module. Since P fixes U pointwise this implies $Pg^{-1}\mathbf{u} = g^{-1}\mathbf{u}$ and hence $gPg^{-1}\mathbf{u} = gg^{-1}\mathbf{u} = \mathbf{u}$. It follows that

$$P_G\mathbf{u} = \frac{1}{\#G}\sum_{g\in G} gPg^{-1}\mathbf{u} = \frac{1}{\#G}\sum_{g\in G}\mathbf{u} = \mathbf{u}.$$

These facts together imply that $V = U \oplus \ker P_G$. Indeed, for any $\mathbf{v} \in V$ we can write $\mathbf{v} = P_G\mathbf{v} + (\mathbf{v} - P_G\mathbf{v})$ where $P_G\mathbf{v} \in U$ and $(\mathbf{v} - P_G\mathbf{v}) \in \ker P_G$ because

$$P_G(\mathbf{v} - P_G\mathbf{v}) = P_G\mathbf{v} - P_GP_G\mathbf{v} = P_G\mathbf{v} - P_G\mathbf{v} = \mathbf{0}.$$

And the intersection is trivial since if $\mathbf{u} \in U$ and $\mathbf{u} \in \ker P_G$ then $\mathbf{u} = P_G\mathbf{u} = \mathbf{0}$. Thus we have shown that $V = U \oplus W$ where $W := \ker P_G$ is a G-submodule of V. □

At this point it is useful to adopt a categorical point of view. The collection of finite dimensional $\mathbb{C}$-vector spaces and $\mathbb{C}$-linear maps forms a category with three important operations:

$$\oplus, \quad \otimes, \quad \mathrm{Hom}_{\mathbb{C}}(-,-).$$

That is, for any two vector spaces U, V we obtain vector spaces $U \oplus V$, $U \otimes V$ and $\mathrm{Hom}_{\mathbb{C}}(U,V)$. If U and V are G-modules then each of the three vector spaces becomes a G-module in a natural way. The action of G on $\mathrm{Hom}_{\mathbb{C}}(U,V)$ is by “conjugation”. That is, if $\varphi : U \to V$ is a $\mathbb{C}$-linear map then for each $g \in G$ we define $g\varphi : U \to V$ by

$$(g\varphi)(\mathbf{u}) := g\varphi g^{-1}\mathbf{u}.$$

If we let G act trivially on $\mathbb{C}$ then this makes the dual space $U^* := \mathrm{Hom}_{\mathbb{C}}(U,\mathbb{C})$ into a G-module via the *contragedient action* $g\varphi(\mathbf{u}) = \varphi(g^{-1}\mathbf{u})$. The subspace $\mathrm{Hom}_G(U,V) \subseteq \mathrm{Hom}_{\mathbb{C}}(U,V)$ of G-invariant linear maps can be defined as the fixed points of conjugation:

$$\mathrm{Hom}_G(U,V) = \mathrm{Hom}_{\mathbb{C}}(U,V)^G = \{\varphi \in \mathrm{Hom}_{\mathbb{C}}(U,V) : g\varphi g^{-1} = \varphi\}.$$

The three bifunctors $\oplus$, $\otimes$, $\mathrm{Hom}_{\mathbb{C}}(-,-)$ (and their G-module versions) satisfy many basic identities. For example, direct sum and tensor product are commutative up to isomorphism, tensor product distributes over direct sum up to isomorphism, etc. Here are some useful properties that we will use without proof:

- $\mathrm{Hom}_G(U \oplus V, W) \cong \mathrm{Hom}_G(U,W) \oplus \mathrm{Hom}_G(V,W)$
- $\mathrm{Hom}_G(U, V \oplus W) \cong \mathrm{Hom}_G(U,V) \oplus \mathrm{Hom}_G(U,W)$
- $\mathrm{Hom}_{\mathbb{C}}(U,V) \cong U^* \otimes V$

It follows from Schur's lemma and the first two of these properties that the decomposition of a G-module into simple modules is **unique** up to isomorphism and rearrangement of the summands. Indeed, suppose that we have

$$U = \bigoplus_i S_i^{\oplus m_i},$$

where the S_i are non-isomorphic simple G-modules and $S_i^{\oplus m_i}$ is the direct sum of S_i with itself m_i times. Then since the Hom bifunctor preserves direct sums in the second coordinate, and since dimension adds over direct sums, we have

$$\langle S_i, U\rangle = \langle S_i, \bigoplus_j S_j^{\oplus m_j}\rangle = \sum_j m_j \langle S_i, S_j\rangle = m_i.$$

Hence the multiplicity of each simple summand of V is uniquely determined. A similar computation shows that $\langle U, S_i\rangle = m_i$ for all i. More generally, if $U = \oplus S_i^{\oplus m_i}$ and $V = \oplus S_i^{\oplus n_i}$, where S_i and S_j are non-isomorphic simple modules for $i \neq j$, then we have $\langle U, V\rangle = \sum_i m_i n_i = \langle V, U\rangle$.

Now we discuss an amazing simplification. It turns out that we can compute the pairing $\langle U, V\rangle$ purely from the traces of the linear operators $g : U \to U$ and $g : V \to V$. This is the notion of the "character" of a G-module.

Definition 3.3. Given a G-module U we define its *character* $\chi_U : G \to \mathbb{C}$ to be the function that sends each group element $g \in G$ to the trace of g as a linear operator on U:

$$\chi_U(g) := \mathrm{tr}(g|_U).$$

We note that $\chi_U(hgh^{-1}) = \chi_U(g)$ because the trace is invariant under conjugation. We also have $\chi_U(\mathrm{id}) = \dim U$ because the identity element of G corresponds to the identity matrix of size $\dim U$. Next we observe that characters convert the algebraic operations of direct sum, tensor product and duality into addition, multiplication and complex conjugation:

- $\chi_{U\oplus V}(g) = \chi_U(g) + \chi_V(g)$
- $\chi_{U\otimes V}(g) = \chi_U(g)\chi_V(g)$
- $\chi_{U^*}(g) = \chi_U(g^{-1}) = \chi_U(g)^*$

The last of these can be proved by observing that each linear operator $g : U \to U$ has finite order $g^n = \mathrm{id}$ for some n, hence its eigenvalues must be n-th roots of unity. But if λ is a root of unity then $\lambda^{-1} = \lambda^*$. The next result is the key property of characters.

Theorem 3.4 (Inner product of characters)**.** *For G-modules U and V we have*

$$\langle U, V\rangle = \frac{1}{\#G} \sum_{g\in G} \chi_U(g)^* \chi_V(g).$$

Proof. Recall that we have $\mathrm{Hom}_{\mathbb{C}}(U,V) \cong U^* \otimes V$ and $\mathrm{Hom}_G(U,V) = \mathrm{Hom}_{\mathbb{C}}(U,V)^G$, hence

$$\langle U, V\rangle = \dim \mathrm{Hom}_G(U,V) = \dim (U^* \otimes V)^G.$$

The result will follow if we can prove that $\dim W^G = \frac{1}{\#G}\sum_{g\in G}\chi_W(g)$ for any G-module W. To prove this we consider the expression $P_W = \frac{1}{\#G}\sum_{g\in G} g$ as a linear endomorphism of W. Using arguments similar to the proof of Maschke's theorem one can check that P_W is a projection onto the fixed subspace $W^G \subseteq W$. By choosing bases for $\mathrm{im}\, P_W = W^G$ and $\ker P_W$ and using the direct sum $W = \mathrm{im}\, P_W \oplus \ker P_W$ we see that P_W has $\dim W^G$ eigenvalues equal to 1 and the rest equal to 0. It follows that

$$\dim W^G = \mathrm{tr}(P_W) = \frac{1}{\#G}\sum_{g\in G}\mathrm{tr}(g|_W) = \frac{1}{\#G}\sum_{g\in G}\chi_W(g).$$

□

As a first applications of characters we show that there exist finitely many simple modules.

Theorem 3.5 (The regular representation)**.** *The* group algebra *of G is defined as the formal $\mathbb{C}$-linear span of the symbols e_g, one for each group element $g \in G$:*

$$\mathbb{C}G = \left\{ \sum_{g\in G} a_g e_g : a_g \in \mathbb{C} \right\}.$$

This is a vector space of dimension $\#G$. It is also a (noncommutative) ring via the multiplication $e_g e_h := e_{gh}$. The group G acts on $\mathbb{C}G$ via the rule $ge_h := e_{gh}$, which we call the regular representation *of G. The character of the regular representation is*

$$\chi_{\mathrm{reg}}(g) = \begin{cases} \#G & g = \mathrm{id}, \\ 0 & g \neq \mathrm{id}. \end{cases}$$

It follows from this that every simple G-module S is a summand of $\mathbb{C}G$, with multiplicity equal to its dimension.

Proof. Note that each element $g \in G$ acts on $\mathbb{C}G$ by permuting the basis elements. Thus the trace of g is just the number of basis elements fixed by g. The identity element fixes every basis element and a non-identity element fixes no basis elements because $ge_h = e_h$ implies $gh = h$ and hence $g = \mathrm{id}$. Now let S be any simple G-module with character χ_S. Then the multiplicity of S in the direct sum decomposition of $\mathbb{C}G$ is

$$\begin{aligned}\langle \mathbb{C}G, S\rangle &= \frac{1}{\#G}\sum_{g\in G}\chi_{\mathrm{reg}}(g)^*\chi_S(g) \\ &= \frac{1}{\#G}\chi_{\mathrm{reg}}(\mathrm{id})^*\chi_S(\mathrm{id})\end{aligned}$$

$$= \chi_S(\mathrm{id})$$
$$= \dim S.$$

□

In particular, this shows that the number of non-isomorphic simple G-modules is less than or equal to $\#G$. (We will see in a moment that the upper bound is attained exactly when G is abelian.) Furthermore, applying dimension to the simple decomposition of G shows that the sum of the squares of the dimensions of all simple G-modules equals $\#G$:

$$\mathbb{C}G = \bigoplus_S S^{\oplus \dim S}$$
$$\dim \mathbb{C}G = \sum_S \dim S \dim S$$
$$\#G = \sum_S (\dim S)^2.$$

Finally we discuss the character table of G. This is just the matrix that displays all values of all characters. Since characters are constant on conjugacy classes this will be a matrix with a row for each simple character and a column for each conjugacy class in G.

Definition 3.6 (The character table)**.** Let $\chi_1, \ldots, \chi_m$ be the simple characters of G and let $C_1, \ldots, C_n$ be the conjugacy classes of G. The *character table* is the $m \times n$ matrix whose i, j entry is $\chi_i(C_j)$, which we define as $\chi_i(g)$ for any group element $g \in C_j$.

Amazingly, it turns out that the character table is square.

Theorem 3.7. *The number of simple characters equals the number of conjugacy classes.*

Proof. Let $C_1, \ldots, C_n$ be the conjugacy classes of G. We say that $\varphi : G \to \mathbb{C}$ is a *class function* if $\varphi(hgh^{-1}) = \varphi(g)$ for all $g, h \in G$. The set of class functions forms a vector space isomorphic to $\mathbb{C}^n$. Furthermore, we have the following standard Hermitian inner product:[22]

$$\langle \varphi, \psi \rangle := \frac{1}{\#G} \sum_{g \in G} \varphi(g)^* \psi(g).$$

We will be done it we can show that the set of simple characters is a basis for the space of class functions. From Schur's lemma we already know that the simple characters are orthonormal with respect to this Hermitian inner product, hence we only need to show that they span the space of class functions.

[22]The scaling factor $\#G$ does't matter.

Let $\chi_1, \dots, \chi_m$ be the simple characters and suppose for contradiction that there exists some class function not in their span. By projecting this class function onto the orthogonal complement of their span we obtain a nonzero class function $\varphi : G \to \mathbb{C}$ satisfying $\langle \chi_i, \varphi \rangle = 0$ for all i. For any G-module U we consider the linear endomorphism $\varphi_U : U \to U$ defined by

$$\varphi_U := \sum_{g \in G} \varphi(g) g^{-1}.$$

Since φ is a class function we note that φ_U is in fact a G-endomorphism. That is, for all $h \in G$ we have

$$h\varphi_U h^{-1} = \sum_{g \in G} \varphi_U(g) h g^{-1} h^{-1} = \sum_{x \in G} \varphi_U(h^{-1}xh) x^{-1} = \sum_{x \in G} \varphi_i(x) x^{-1} = \varphi_i.$$

Let S_i be the simple G-module with character χ_i and let $\varphi_i := \varphi_{S_i}$. Since φ_i is a G-endomorphism of S_i it follows from Schur's lemma that φ_i acts like a scalar $\lambda_i \in \mathbb{C}$, hence the trace of φ_i is $\lambda_i \dim S_i$. On the other hand, the trace of φ_i is

$$\operatorname{tr}(\varphi_i) = \sum_g \varphi_i(g) \operatorname{tr}(g^{-1}|_{S_i}) = \sum_g \varphi_i(g) \chi_i(g)^* = \#G \cdot \langle \chi_i, \varphi \rangle = 0,$$

hence $\lambda_i = 0$. We have shown that each φ_i is the zero endomorphism on the simple G-module S_i. If follows from Maschke's theorem that φ_U is the zero endomorphism on any G-module U. In particular, this holds for the regular representation $U = \mathbb{C}G$. In this case, the action of φ_U on the basis element $e_{\mathrm{id}} \in \mathbb{C}G$ gives

$$0 = \varphi_U e_{\mathrm{id}} = \sum_{g \in G} \varphi(g) g^{-1} e_{\mathrm{id}} = \sum_{g \in G} \varphi(g) e_{g^{-1}}.$$

Since the elements $e_{g^{-1}} \in \mathbb{C}G$ are linearly independent this implies that each coefficient $\varphi(g)$ is zero. Contradiction. □

Note that this proof does not provide any bijection between the simple characters and the conjugacy classes. Finally, we observe that the orthogonality of simple characters implies another orthogonality relation for the rows of the character table.

Theorem 3.8 (Orthogonality relations for characters)**.** *Let $\chi_1, \dots, \chi_n$ be the simple characters and let $C_1, \dots, C_n$ be the conjugacy classes of G. Let $\chi_i(C_j)$ be the value of χ_i on any element of C_j. Then we have*

$$\sum_{k=1}^{n} \#C_k \cdot \chi_i(C_k) \chi_j(C_k)^* = \begin{cases} \#G & i = j, \\ 0 & i \neq j, \end{cases}$$

and

$$\sum_{k=1}^{n} \chi_k(C_i) \chi_k(C_j)^* = \begin{cases} \#G/\#C_i & i = j, \\ 0 & i \neq j. \end{cases}$$

Proof. From Schur's lemma we have $\langle \chi_i, \chi_j \rangle = \delta_{ij}$ for the simple characters. This implies that

$$\#G \cdot \delta_{ij} = \sum_{g \in G} \chi_i(g)^* \chi_j(g) = \sum_{k=1}^{n} \#C_k \cdot \chi_i(C_k)^* \chi_j(C_k).$$

Now let M be the $n \times n$ matrix whose i, j entry is $\chi_i(C_j)/\sqrt{\#G/\#C_j}$. The previous formula implies that $M^* M = I$. It then follows from the Rank-Nullity theorem that $MM^* = I$, hence

$$\sum_{k=1}^{n} \frac{\chi_k(C_i)}{\sqrt{\#G/\#C_i}} \frac{\chi_k(C_j)^*}{\sqrt{\#G/\#C_j}} = \delta_{ij}$$

$$\sum_{k=1}^{n} \chi_k(C_i)^* \chi_k(C_j) = \frac{\#G}{\sqrt{\#C_i}\sqrt{\#C_j}} \delta_{ij}.$$

□

Remark: It is not necessarily easy to compute the character table of a group, but the orthogonality relations can help.

Examples:

Abelian groups. (Dirichlet characters.)

Murnaghan–Nakayama rule. Conjugacy classes in S_n are parametrized by integer partitions $\lambda = (\lambda_1, \lambda_2, \ldots)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and $\sum_i \lambda_i = n$. Let C_λ the conjugacy class with cycle type λ. Then we can compute $\chi_\lambda(C_\mu)$ using the Murnaghan–Nakayama rule.

Or we can compute it by expressing the Schur polynomial in terms of power sums $s_\lambda = \sum a_{\lambda\mu} p_\mu$. Then we have $\chi_\lambda(C_\mu) = z_\mu a_{\lambda\mu}$, where z_μ is the product $\Pi_{i \geq 1} i^{m_i} m_i!$ where m_i is the number of parts of μ equal to i. Recall that

$$s_\lambda(x_1, \ldots, x_n) = \frac{\det(x_i^{\lambda_j + n - j})}{\det(x_i^{n-j})}.$$

The dimension $\chi_\lambda(\mathrm{id})$ is $z_{1^n} = n!$ times the coefficient of p_1^n in s_λ.

3.4 The McKay correspondence

Given a finite group G with simple modules $S_1, \ldots, S_n$ we define a graph on the set $\{1, \ldots, n\}$ by drawing an edge $i \to j$ weighted by the integer a_{ij} if

$$V \otimes S_i = \bigoplus_j S_j^{\oplus a_{ij}}.$$

We can also express this in terms of characters:

$$\chi_V \chi_i = \sum_j a_{ij} \chi_j.$$

Let $M = (\chi_i(C_j))$ be the character table and let $\mathbf{m}_k = (\chi_i(C_k))$ be its kth column. We recall from the previous section that $\langle \mathbf{m}_k, \mathbf{m}_k \rangle = \#G/\#C_k$ and $\langle \mathbf{m}_k, \mathbf{m}_\ell \rangle = 0$ for $k \neq \ell$. If $A = (a_{ij})$ is the weighted adjacency matrix of the McKay graph then evaluating the previous equation on the kth conjugacy class C_k tells us that

$$A\mathbf{m}_k = \chi_V(C_k)\mathbf{m}_k.$$

In other words, the McKay graph is constructed precisely so that the columns of the character table are an orthogonal basis of eigenvectors for the adjacency matrix. This holds without any hypothesis on the pair (G, V).[23]

Furthermore, if we let $C_1 = \{\mathrm{id}\}$ be the conjugacy class of the identity then we have $\chi_V(C_1) = \dim V$ and $\chi_i(C_1) = \dim S_i$, so taking $k = 1$ gives

$$A \begin{pmatrix} \dim S_1 \\ \vdots \\ \dim S_n \end{pmatrix} = \dim V \begin{pmatrix} \dim S_1 \\ \vdots \\ \dim S_n \end{pmatrix}.$$

Since all of these dimensions are positive, and since A has non-negative entries, the Perron-Frobenius theorem implies that the spectral radius of A equals $\dim V$.

Based on the results of Chapter 1 it is reasonable to search for pairs (G, V) where A is a symmetric $\{0, 1\}$ matrix and where $\dim V = 2$, since in this case we can conclude that the McKay graph of (G, V) is an affine graph of type ADE. We make the following remarks:

- The McKay graph is undirected if and only if $V \cong V^*$.
- The McKay graph is connected if and only if V is *faithful*, i.e., $g|_V = h|_V$ implies $g = h$.
- The McKay graph is loopless if and only if V has no trivial summand.

THE REST OF THIS SECTION IS JUST A SKETCH

There is an easy way to achieve the first two of these: Let G be a finite subgroup of $SU(2)$ with defining representation V. This is faithful by definition, and the function $\varphi : \mathbb{C}^2 \to (\mathbb{C}^2)^*$ that sends $\mathbf{v} = (v_1, v_2)$ to the linear functional $\varphi_{\mathbf{v}}(w_1, w_2) = v_1 w_2 - v_2 w_1$ is a G-isomorphism. We need to check that φ is invertible and that $\varphi_{g\mathbf{v}}(\mathbf{w}) = \varphi_{\mathbf{v}}(g^{-1}\mathbf{w})$ for all $\mathbf{w} \in \mathbb{C}^2$. Let $\varepsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ so that $\varphi_{\mathbf{v}}(\mathbf{w}) = \mathbf{v}^T \varepsilon \mathbf{w}$. Thus need to check that $(g\mathbf{v})^T \varepsilon \mathbf{w} = \mathbf{v}^T \varepsilon g^{-1} \mathbf{w}$. Thus

[23] As long as we work over the complex numbers. More generally, we can use any field of characteristic not dividing $\#G$.

we only need to check that $g^T\varepsilon = \varepsilon g^{-1}$ for all $g \in SU(2)$. This comes from the accidental low-dimensional coincidence that $SU(2) \cong Sp(1)$. In general, the Frobenius-Schur indicator $\nu(\chi) = \frac{1}{\#G}\sum_g \chi(g^2)$ tells us when χ is real ($\nu(\chi) = 1$), symplectic ($\nu(\chi) = -1$), i.e., self-dual with G-invariant alternating form, or none of the above ($\nu(\chi) = 0$).

For the third part suppose there exists a unit vector $\mathbf{u} \in V$ fixed by every $g \in G$. With respect to a Hermitian basis $\mathbf{u}, \mathbf{v}$ each element g has the form $\begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}$. Since g is unitary this forces $b = 1$, hence every g is the identity. Contradiction.

It only remains to show that $a_{ij} \in \{0, 1\}$ for $G \subseteq SU(2)$ and V the defining representation. Here is Steinberg's proof.[24] Note that $a_{ij} = \langle \chi_V \chi_i, \chi_j \rangle$, hence

$$\sum_j a_{ij}^2 = \langle \chi_V \chi_i, \chi_V \chi_i \rangle = \frac{1}{\#G} \sum_g |\chi_V(g)\chi_i(g)|^2.$$

Since $|\chi_V(g)| \leq 2$ (because the eigenvalues are roots of unity λ, λ^{-1}) with equality if and only if $g = \pm I$ this implies

$$\sum_j a_{ij}^2 = \frac{1}{\#G} \sum_g |\chi_V(g)|^2 |\chi_i(g)|^2 < \frac{4}{\#G} \sum_g |\chi_i(g)|^2 = 4\langle \chi_i, \chi_i \rangle = 4.$$

apart from the exceptional case $G = \{\pm I\}$. Since the a_{ij} are nonnegative integers the inequality $\sum_j a_{ij}^2 \leq 3$ implies $a_{ij} \in \{0, 1\}$.

Proof that faithful implies McKay graph connected:[25] Let φ be a faithful character taking exactly r values $a_1, \ldots, a_r \in \mathbb{C}$. Let $A_i = \{g \in G : \varphi(g) = a_i\}$, and let χ be any simple character. Then we have $\langle \varphi^k, \chi \rangle \neq 0$ for some $0 \leq k \leq r-1$. Indeed, suppose for contradiction that $\langle \varphi^k, \chi \rangle = 0$ for all $0 \leq k \leq r-1$. Then we have

$$0 = \#G\langle \varphi^k, \chi \rangle = \sum_g \varphi(g)^k \chi(g)^* = \sum_{i=0}^{r-1} a_i^k b_i,$$

where $b_i = \sum_{g \in A_i} \chi(g)^*$. Since the numbers $a_1, \ldots, a_r$ are distinct, Vandermonde implies that $b_1 = \cdots = b_r = 0$. Let's suppose that $a_1 = \varphi(\text{id}) = d$. Since φ is faithful I claim that $A_1 = \{\text{id}\}$. Indeed, suppose that $\varphi(g) = \varphi(\text{id})$. If $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of g then this implies $\lambda_1 + \cdots + \lambda_d = d$. Since the eigenvalues are roots of unity, this implies $\lambda_1 = \cdots = \lambda_d = 1$, which implies that $g = \text{id}$. Thus we have shown that $0 = b_1 = \chi(\text{id})^*$. Contradiction.

[24] *Finite Subgroups of SU(2), Dynkin Diagrams and Affine Coxeter Elements*, Pacific Journal of Mathematics, Vol. 118, No. 2 (1985), pages 587-624.

[25] A NOTE ON THEOREMS OF BURNSIDE AND BLICHFELDT, RICHARD BRAUER

4 Characters of the symmetric group

The theory of characters of finite groups was developed entirely by Frobenius between the years 1896 and 1900. At the end of this period he completed the most important example, which is the computation of the characters of symmetric groups S_n. In this chapter we present Frobenius' proof of his theorem, as discussed in Curtis (*Pioneers of representation theory*, page 73). Since Frobenius' proof involves key ideas from the theory of symmetric polynomials, we find it a convenient time to introduce this theory.[26]

4.1 Elementary symmetric polynomials

The symmetric group S_n acts on the ring of polynomials $\mathbb{Q}[x_1, \ldots, x_n]$ by permuting variables:

$$\sigma \cdot f(x_1, \ldots, x_n) := f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}).$$

Let $\Lambda_n = \mathbb{Q}[x_1, \ldots, x_n]^{S_n}$ denote the set of polynomials fixed by every element of S_n. Since the action preserves addition, multiplication and scalar multiplication by $\mathbb{Q}$ we observe that Λ_n is a $\mathbb{Q}$-algebra. Define the *elementary symmetric polynomials* $e_k(x_1, \ldots, x_n)$ via the generating function $E(t) = \prod_{i=1}^n (1 + x_i t) = \sum_{k=0}^n e_k(x_1, \ldots, x_n) t^k$. To be explicit, we define

$$e_k(x_1, \ldots, x_n) = \sum_{1 \le i_1 < \cdots < i_k \le n} x_{i_1} \cdots x_{i_k}.$$

Theorem 4.1 (Fundamental Theorem of Symmetric Functions)**.** *Let $t_1, \ldots, t_n$ be another set of variables. Then the $\mathbb{Q}$-algebra homomorphism $\mathbb{Q}[t_1, \ldots, t_n] \to \Lambda_n$ defined by $t_k \mapsto e_k(x_1, \ldots, x_n)$ is a bijection. In other words, the polynomials $e_1, \ldots, e_n$ are an algebraically independent generating set for Λ_n.*

Proof. (Surjective) First we show that this homomorphism is surjective. We will use the lexicographic order on monomials. That is, given $\mathbf{k}, \boldsymbol{\ell} \in \mathbb{N}^n$ we define $\mathbf{k} <_{\text{lex}} \boldsymbol{\ell}$ when $\mathbf{k} \neq \boldsymbol{\ell}$ and if i is minimum such that $k_i \neq \ell_i$ then $k_i < \ell_i$. Given $f(x_1, \ldots, x_n) \in \mathbb{Q}[x_1, \ldots, x_n]$ we will write $\deg(f) = \mathbf{k}$ when

$$f(\mathbf{x}) = c\mathbf{x}^{\mathbf{k}} + \text{lower terms} = c x_1^{k_1} \cdots x_n^{k_n} + \text{lower terms},$$

with $c \neq 0$. One can check that lexicographic order is a well-order on $\mathbb{N}^n$ and that $\deg(fg) = \deg(f) + \deg(g)$ for all $f(\mathbf{x}), g(\mathbf{x}) \in \mathbb{Q}[\mathbf{x}]$. Now let $f(\mathbf{x}) \in \mathbb{Q}[\mathbf{x}]$ be symmetric with $\deg(f) = \mathbf{k} = (k_1, \ldots, k_n) \in \mathbb{N}^n$ and leading term $c\mathbf{x}^{\mathbf{k}}$ with $c \neq 0$. The fact that f is symmetric implies that $k_1 \geq \cdots \geq k_n$. Indeed, suppose for contradiction that $k_i < k_{i+1}$ for some i and let $\mathbf{k}'$ be obtained from $\mathbf{k}$ by switching the entries k_i and k_{i+1}. Then one can check that $\mathbf{k}' >_{\text{lex}} \mathbf{k}$. On

[26] Frobenius' student Schur connected the representations of the finite group S_n to the representations the infinite group GL_n. This theory is also based on symmetric polynomials but in a slightly different way.

the other hand, since f is symmetric and since $\mathbf{k}'$ is obtained from $\mathbf{k}$ by a permutation we see that the coefficients of $\mathbf{x}^{\mathbf{k}}$ and $\mathbf{x}^{\mathbf{k}'}$ are equal and nonzero which contradicts the fact that $\mathbf{k}$ is the highest power occurring in f. Now define $g(\mathbf{x}) = ce_1(\mathbf{x})^{k_1-k_2} \cdots e_{n-1}(\mathbf{x})^{k_{n-1}-k_n} e_n(\mathbf{x})^{k_n}$ and observe that

$$\begin{aligned}
\deg(g) &= (k_1 - k_2) \cdot \deg(e_1) + \cdots + (k_{n-1} - k_n) \cdot \deg(e_n) + k_n \cdot \deg(e_n) \\
&= (k_1 - k_2) \cdot (1, 0, \ldots, 0) + \cdots + (k_{n-1} - k_n) \cdot (1, \ldots, 1, 0) + k_n \cdot (1, \ldots, 1) \\
&= (k_1, \ldots, k_n) \\
&= \mathbf{k}.
\end{aligned}$$

This implies that the polynomial $f(\mathbf{x}) - g(\mathbf{x})$ has degree strictly less than $\mathbf{k}$, hence by induction there exists a polynomial $h(t_1, \ldots, t_n) \in \mathbb{Q}[t_1, \ldots, t_n]$ satisfying $f(\mathbf{x}) - g(\mathbf{x}) = h(e_1, \ldots, e_n)$. Then letting $h'(t_1, \ldots, t_n) = h(t_1, \ldots, t_n) + ct_1^{k_1-k_2} \cdots t_{n-1}^{k_{n-1}-k_n} t_n^{k_n}$ gives $f(\mathbf{x}) = h'(e_1, \ldots, e_n)$ as desired. □

The proof of injectivity uses a lemma called the Jacobian criterion.[27]

Lemma 4.2 (Jacobian criterion)**.** *Given any polynomials $f_1, \ldots, f_n \in \mathbb{Q}[x_1, \ldots, x_n]$ we consider the Jacobian matrix $J_{f,x} = (\partial f_i / \partial x_j)$. If $\det(J_{f,x}) \in \mathbb{Q}[x_1, \ldots, x_n]$ is not the zero polynomial then the ring homomorphism $\varphi_f : \mathbb{Q}[t_1, \ldots, t_n] \to \mathbb{Q}[x_1, \ldots, x_n]$ defined by $\varphi_f(t_k) = f_k$ is injective. In other words, the polynomials $f_1, \ldots, f_n$ are algebraically independent over $\mathbb{Q}$.*

Proof. Suppose that φ_f is not injective and let $F(t_1, \ldots, t_n) \in \mathbb{Q}[t_1, \ldots, t_n]$ be a nonconstant polynomial of minimum total degree such that $F(f_1, \ldots, f_n) \equiv 0$ in $\mathbb{Q}[x_1, \ldots, x_n]$. Differentiating this identity and using to the multivariable chain rule gives $J_{F,f} \cdot J_{f,x} \equiv 0$, where

$$J_{F,f} = \begin{pmatrix} \frac{\partial F}{\partial t_1}(f_1, \ldots, f_n) & \cdots & \frac{\partial F}{\partial t_n}(f_1, \ldots, f_n) \end{pmatrix}$$

is the $1 \times n$ gradient vector of F evaluated at $(f_1, \ldots, f_n)$. We note that this vector is nonzero. Indeed, since F is nonconstant, at least one partial derivative $\partial F / \partial t_i$ is not the zero polynomial. If it happened that $(\partial F / \partial t_i)(f_1, \ldots, f_n) \equiv 0$ in $\mathbb{Q}[x_1, \ldots, x_n]$ then since the total degree of $\partial F / \partial t_i$ is less than the degree of F this would contradict the minimality of the relation $F(f_1, \ldots, f_n) \equiv 0$. Finally, since $\mathbb{Q}[x_1, \ldots, x_n]$ is an integral domain we may consider the matrix equation $J_{F,f} \cdot J_{f,x} \equiv 0$ over the fraction field. Since the matrix $J_{f,x}$ has a nontrivial left kernel we conclude that $\det(J_{f,x}) \equiv 0$. □

Proof. (Injective) To complete the proof of the Fundamental Theorem we must show that the Jacobian matrix of the elementary symmetric polynomials has nonzero determinant. In the section on power sums below we will show that in fact this determinant is equal to the Vandermonde determinant:

$$\det(J_{e,x}) = \det(\partial e_i / \partial x_j) = \prod_{1 \le i < j \le n} (x_i - x_j) \in \mathbb{Q}[x_1, \ldots, x_n],$$

[27] See, for example, Loehr, *Bijective combinatorics*, Theorem 10.83.

which is nonzero. $\square$

As a corollary we get a basis for the ring Λ_n as a vector space over $\mathbb{Q}$. Given an integer partition $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots 0)$ we define the symmetric polynomial

$$e_\lambda(x_1, \ldots, x_n) := \prod_{k \geq 1} e_{\lambda_k}(x_1, \ldots, x_n) \in \Lambda_n.$$

The first part of the Fundamental Theorem (surjectivity) implies that the e_λ are a spanning set for Λ_n. The second part of the Fundamental Theorem (injectivity) implies that the e_λ are linearly independent over $\mathbb{Q}$.

4.2 Complete homogeneous symmetric polynomials

We define the *complete homogeneous symmetric polynomials* $h_k(x_1, \ldots, x_n)$ via the generating function $H(t) = \prod_{i=1}^n (1 - x_i t)^{-1} = \sum_{k \geq 0} h_k(x_1, \ldots, x_n) t^k$. To be explicit, we have

$$h_k(x_1, \ldots, x_n) = \sum_{1 \leq i_1 \leq \cdots \leq i_k \leq n} x_{i_1} \cdots x_{i_k}.$$

Comparing coefficients in the identity $E(t)H(-t) = 1$ gives

$$\begin{aligned} e_1 - h_1 &= 0, \\ e_2 - e_1 h_1 + h_2 &= 0, \\ &\vdots \\ e_k - e_{k-1} h_1 + \cdots + (-1)^{k-1} e_1 h_{k-1} + (-1)^k h_k &= 0, \end{aligned}$$

etc. Solving for $e_1, e_2, \ldots$ gives

$$\begin{aligned} e_1 &= h_1, \\ e_2 &= h_1^2 - h_2, \\ e_3 &= h_1^3 - 2h_1 h_2 + h_3, \end{aligned}$$

and by induction we conclude that $e_1, e_2, \ldots, e_n \in \mathbb{Q}[h_1, \ldots, h_n]$. Similarly we have $h_1, \ldots, h_n \in \mathbb{Q}[e_1, \ldots, e_n]$ so that the rings $\mathbb{Q}[e_1, \ldots, e_n]$ and $\mathbb{Q}[h_1, \ldots, h_n]$ are equal, and both are equal the ring of symmetric polynomials Λ_n. I claim that the polynomials $h_1, \ldots, h_n$ are also algebraically independent over $\mathbb{Q}$. To see this we will use an interesting lemma.

Lemma 4.3 (Noetherian implies Hopfian)**.** *Let R be a Noetherian ring. Then any surjective ring homomorphism $\varphi : R \to R$ must be injective.*

Proof. We note that $\ker \varphi \subseteq \ker \varphi^2 \subseteq \ker \varphi^3 \subseteq \cdots$. Since R is Noetherian we have $\ker \varphi^m = \ker \varphi^{m+1}$ for some m. We wish so show that $\ker \varphi = \{0\}$, so consider any element $r \in \ker \varphi$. Since φ^m is surjective (because φ is) we have $r = \varphi^m(s)$ for some $s \in R$, hence $\varphi^{m+1}(s) = \varphi(r) = 0$. Then since $\ker \varphi^{m+1} \subseteq \ker \varphi^m$ we have $\varphi^m(s) = 0$, i.e., $r = 0$. $\square$

Now consider the evaluation homomorphisms φ_e and φ_h from $\mathbb{Q}[t_1,\ldots,t_n]$ to Λ_n sending $t_k \mapsto e_k(x_1,\ldots,x_n)$ and $t_k \mapsto h_k(x_1,\ldots,x_n)$, respectively. The Fundamental Theorem in the previous section says that φ_e is bijective. Thus we can define the ring homomorphism $\varphi := \varphi_e^{-1} \circ \varphi_h$ from $\mathbb{Q}[t_1,\ldots,t_n]$ to itself. We saw above that φ_h is surjective, hence φ is surjective. By the Hilbert basis theorem and the previous lemma this implies that φ is injective, and hence φ_h is also injective. Thus we have shown that $h_1,\ldots,h_n$ are algebraically independent over $\mathbb{Q}$.

But we can give a more elementary proof based on the "reciprocity" between $E(t)$ and $H(t)$. Since the $e_1,\ldots,e_n$ are algebraically independent we can define a ring homomorphism $\omega : \Lambda_n \to \Lambda_n$ by $\omega(e_k) := h_k$ for all k. I claim that $\omega(h_k) = e_k$ and hence ω is an involutive automorphism. To see this we apply ω to the reciprocity relation $E(t)H(-t) = 1$ to get

$$h_k - h_{k-1}\omega(h_1) + \cdots + (-1)^{k-1}h_1\omega(h_{k-1}) + (-1)^k\omega(h_k) = 0.$$

Then multiplying both sides by $(-1)^k$ gives

$$\omega(h_k) - \omega(h_{k-1})h_{k-1} + \cdots + (-1)^{k-1}\omega(h_1)h_{k-1} + (-1)^k h_k = 0.$$

Since the polynomials $\omega(h_1),\omega(h_2),\ldots$ satisfy the same recurrence relation as the polynomials $e_1,e_2,\ldots$ we conclude by induction that $\omega(h_k) = e_k$. Thus ω is a $\mathbb{Q}$-algebra automorphism exchanging the generators $e_1,e_2,\ldots,e_n$ and $h_1,\ldots,h_n$. Hence the algebraic independence of the h_k is equivalent to the algebraic independence of the e_k. As with the polynomials $e_\lambda(x_1,\ldots,x_n)$, it follows from this that the polynomials $h_\lambda(x_1,\ldots,x_n) = \prod_{k\geq 1} h_{\lambda_k}(x_1,\ldots,x_n)$ are a vector space basis for Λ_n as λ ranges over integer partitions.

Remark: We note that

$$e_k(1,1,\ldots,1) = \binom{n}{k} \qquad \text{and} \qquad h_k(1,1,\ldots,1) = \binom{n+k-1}{k},$$

hence the involution ω is some kind of algebraic version of the combinatorial "reciprocity"

$$\binom{-n}{k} = (-1)^k\binom{n+k-1}{k}$$

relating the counting of sets and multisets.

4.3 Power sum symmetric polynomials

Another basis of Λ_n is given by the *power sum symmetric polynomials*

$$p_k(x_1,\ldots,x_n) = x_1^k + x_2^k + \cdots + x_n^k.$$

The *Newton identities* express the p_k in terms of the e_k. To prove them, we note that

$$\log E(t) = \log\prod_{i=1}^n(1+x_it) = \sum_{i=1}^n \log(1+x_it),$$

and hence

$$\begin{aligned}
E'(t)/E(t) &= \frac{d}{dt}\log E(t) \\
&= \sum_{i=1}^{n} \frac{d}{dt}\log(1+x_i t) \\
&= \sum_{i=1}^{n} \frac{x_i}{1+x_i t} \\
&= \sum_{i=1}^{n} x_i \sum_{k\geq 0}(-x_i t)^k \\
&= \sum_{k\geq 0}(-t)^k \sum_{i=1}^{n} x_i^{k+1} \\
&= P(-t),
\end{aligned}$$

where we define the generating function

$$P(t) = \sum_{k\geq 0} p_{k+1}(x_1,\ldots,x_n)t^k.$$

Then comparing coefficients in the identity $E'(t) = E(t)P(-t)$ gives

$$\begin{aligned}
e_1 &= p_1, \\
2e_2 &= e_1p_1 - p_2, \\
3e_3 &= e_2p_1 - e_1p_2 + p_3, \\
&\vdots \\
ke_k &= e_{k-1}p_1 - e_{k-2}p_2 + \cdots + (-1)^{k-1}e_1p_{k+1} + (-1)^k p_k,
\end{aligned}$$

etc. This shows that that $\mathbb{Q}[p_1,\ldots,p_n] = \mathbb{Q}[e_1,\ldots,e_n] = \Lambda_n$.[28] We now use the Jacobian criterion to show that the power sums are algebraically independent. After that we will complete our proof that the elementary symmetric polynomials are algebraically independent.

Note that the Jacobian matrix of the power sums $p_1,\ldots,p_n$ with respect to the variables $x_1,\ldots,x_n$ is the Vandermonde matrix with scaled rows:

$$J_{p,x} = (\partial p_i/\partial x_j) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2x_1 & 2x_2 & \cdots & 2x_n \\ 3x_1^2 & 3x_2^2 & \cdots & 3x_n^2 \\ \vdots & \vdots & & \vdots \\ nx_1^{n-1} & nx_2^{n-1} & \cdots & nx_n^{n-1}. \end{pmatrix}$$

[28] However, we remark that $\mathbb{Z}[p_1,\ldots,p_n]$ is a strict subring of $\mathbb{Z}[e_1,\ldots,e_n]$.

Hence the determinant is the Vandermonde determinant scaled by $1 \cdot 2 \cdots n = n!$:

$$\det(J_{p,x}) = \pm n! \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

Since this is not the zero polynomial we conclude from the Jacobian criterion that $p_1, \ldots, p_n$ are algebraically independent over $\mathbb{Q}$. In other words, the polynomials

$$p_\lambda(x_1, \ldots, x_n) := \prod_{k \geq 1} p_{\lambda_k}(x_1, \ldots, x_n)$$

are $\mathbb{Q}$-linearly independent as λ ranges over all integer partitions. This gives another vector space basis for Λ_n.

Next we compute the Jacobian determinant of $p_1, \ldots, p_n$ with respect to $e_1, \ldots, e_n$. By the Newton identities above we have

$$p_k = (-1)^{k-1} k e_k + \text{some polynomial in } e_1, \ldots, e_{k-1}.$$

It follows that the matrix $J_{p,e} = (\partial p_i / \partial e_j)$ is upper triangular with diagonal entries $\partial p_k / \partial e_k = (-1)^{k-1} k$, hence $\det(J_{p,e}) = \pm n!$. It follows from the chain rule that

$$\begin{aligned} J_{p,e} \cdot J_{e,x} &= J_{p,x} \\ \det(J_{p,e}) \det(J_{e,x}) &= \det(J_{p,x}) \\ \pm n! \det(J_{e,x}) &= n! \prod_{1 \leq i < j \leq n} (x_i - x_j) \\ \det(J_{e,x}) &= \pm \prod_{1 \leq i < j \leq n} (x_i - x_j). \end{aligned}$$

Since this is not the zero polynomial, we conclude from the Jacobian criterion that $e_1, \ldots, e_n$ are algebraically independent. This completes our proof of the Fundamental Theorem.

The power sums have a similar relationship to the complete homogeneous symmetric polynomials $h_k(x_1, \ldots, x_n)$. Differentiating the identity $H(t)E(-t) = 1$ gives $H'(t)E(-t) - H(t)E'(-t) = 0$ and hence $H'(t)/H(t) = E'(-t)/E(-t) = P(t)$. Then comparing coefficients in the identity $H'(t) = H(t)P(t)$ gives

$$\begin{aligned} h_1 &= p_1, \\ 2h_2 &= h_1 p_1 + p_2, \\ 3h_3 &= h_2 p_1 + h_1 p_2 + p_3, \\ &\vdots \\ kh_k &= h_{k-1} p_1 + h_{k-2} p_2 + \cdots + h_1 p_{k+1} + p_k, \end{aligned}$$

etc. This is another form of the "Newton identities". Applying the involution $\omega(e_k) = h_k$ and comparing both forms of the Newton identities shows that $\omega(p_k) = (-1)^{k-1} p_k$ for all $k \geq 1$.

4.4 Alternating polynomials and Schur polynomials

We say that a polynomial $f(x_1,\ldots,x_n) \in \mathbb{Q}[x_1,\ldots,x_n]$ is *alternating* if it becomes negative after switching any two inputs; equivalently, if $\sigma \cdot f = \text{sgn}(\sigma)f$ for all $\sigma \in S_n$. We note that this condition is preserved by addition, multiplication and scalar multiplication, hence the alternating polynomials form a $\mathbb{Q}$-subalgebra of the symmetric polynomials Λ_n. We will see below that the relationship between alternating and symmetric polynomials is quite simple.

Determinants are a natural way to produce alternating polynomials. For example, we have the *Vandermonde determinant*:

$$\Delta(x_1,\ldots,x_n) := \det \begin{pmatrix} x_1^{n-1} & \cdots & x_1 & 1 \\ x_2^{n-1} & \cdots & x_2 & 1 \\ \vdots & \vdots & & \vdots \\ x_n^{n-1} & \cdots & x_n & 1 \end{pmatrix} \in \mathbb{Q}[x_1,\ldots,x_n].$$

Note that Δ is alternating because the determinant of a matrix is an alternating function of its rows. I claim that

$$\Delta(x_1,\ldots,x_n) = \prod_{1\le i<j\le n}(x_i - x_j).$$

To prove this we just need to show that Δ is divisible by $x_i - x_j$ for each pair $1 \le i < j \le n$. Then the result will follow from three facts:

- $\mathbb{Q}[x_1,\ldots,x_n]$ is a unique factorization domain (Gauss' Lemma).
- $\Delta(x_1,\ldots,x_n)$ and $\prod_{1\le i<j\le n}(x_i - x_j)$ are both homogeneous of degree $\binom{n}{2}$.
- The monomial $x_1^{n-1}x_2^{n-2}\cdots x_{n-1}^1$ has coefficient $+1$ in both polynomials.

So fix $1 \le i \le n$ and consider Δ as an element of the ring $R[x_i]$ where R is the ring of polynomials in the variables $\{x_1,\ldots,x_n\} \setminus \{x_i\}$ over $\mathbb{Q}$. We can also view Δ as an element of $K[x_i]$ where K is the fraction field of the domain R. For any $j \neq i$ we may consider the evaluation map $K[x_i] \to K$ defined by sending $x_i \mapsto x_j$. Since Δ is alternating this map sends Δ to zero. Since K is a field this implies that the kernel is the principal ideal generated by the polynomial $x_i - x_j \in K[x_i]$. We have shown that $\Delta = (x_i - x_j)g(x_1,\ldots,x_n)$ for some rational function $g(x_1,\ldots,x_n) = p(x_1,\ldots,x_n)/q(x_1,\ldots,x_n)$ with p,q coprime elements of $\mathbb{Q}[x_1,\ldots,x_n]$. Since Δ is a polynomial this implies that q divides $(x_i - x_j)p$ in $\mathbb{Q}[x_1,\ldots,x_n]$. Finally, since q is coprime to both $x_i - x_j$ and p this forces q to be constant. Thus we have shown that $x_i - x_j$ divides Δ in $\mathbb{Q}[x_1,\ldots,x_n]$.

The following more general result has essentially the same proof.

Theorem 4.4 (Alternating polynomials)**.** *Every alternating polynomial has a unique expression of the form*

$$f(x_1,\ldots,x_n) = \Delta(x_1,\ldots,x_n)g(x_1,\ldots,x_n),$$

where $g(x_1,\ldots,x_n)$ is a symmetric polynomial and $\Delta(x_1,\ldots,x_n)=\prod_{1\leq i<j\leq n}(x_i-x_j)$ is the Vandermonde determinant.

Proof. If $f(x_1,\ldots,x_n)$ is alternating then the same argument as above shows that x_i-x_j divides f in $\mathbb{Q}[x_1,\ldots,x_n]$ for all $1\leq i<j\leq n$. By unique factorization this implies that $f=\Delta g$ for some polynomial g. But this g is symmetric because

$$\begin{aligned}
\sigma\cdot f &= \sigma\cdot(\Delta g)\\
\sigma\cdot f &= (\sigma\cdot\Delta)(\sigma\cdot g)\\
\operatorname{sgn}(\sigma)f &= \operatorname{sgn}(\sigma)\Delta(\sigma\cdot g)\\
f &= \Delta(\sigma\cdot g)\\
\Delta g &= \Delta(\sigma\cdot g)\\
g &= \sigma\cdot g.
\end{aligned}$$

□

This leads to a natural basis for the $\mathbb{Q}$-algebra of alternating polynomials, which then lifts to a basis of Λ_n called the "Schur polynomials". Let $f(x_1,\ldots,x_n)$ be alternating and write

$$f(x_1,\ldots,x_n)=\sum_{\alpha\in\mathbb{N}^n}c_\alpha\mathbf{x}^\alpha,$$

with $c_\alpha\in\mathbb{Q}$ and $\mathbf{x}^\alpha:=x_1^{\alpha_1}\cdots x_n^{\alpha_n}$. Since f is alternating then we have $c_{\sigma\cdot\alpha}=\operatorname{sgn}(\sigma)c_\alpha$ for all $\sigma\in S_n$ and $\alpha\in\mathbb{N}^n$. If $\alpha=(\alpha_1,\ldots,\alpha_n)$ with $\alpha_i=\alpha_j$ for some $i\neq j$ then I claim that $c_\alpha=0$. Indeed, let σ be the tranposition (ij). Then we have $\sigma\cdot\alpha=\alpha$ and $\operatorname{sgn}(\sigma)=-1$, hence $c_\alpha=-c_\alpha$. Thus every monomial in f must have distinct exponents. Consider the set of strictly decreasing sequences of non-negative integers:

$$\mathbb{N}^n_>:=\{(\alpha_1,\ldots,\alpha_n)\in\mathbb{N}^n:\alpha_1>\cdots>\alpha_n\}.$$

Then we can write

$$f(x_1,\ldots,x_n)=\sum_{\alpha\in\mathbb{N}^n_>}c_\alpha\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)x_{\sigma(1)}^{\alpha_1}\cdots x_{\sigma(n)}^{\alpha_n}=\sum_{\alpha\in\mathbb{N}^n_>}c_\alpha\Delta_\alpha(x_1,\ldots,x_n),$$

so that the *generalized Vandermonde determinants*

$$\Delta_\alpha(x_1,\ldots,x_n):=\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)x_{\sigma(1)}^{\alpha_1}\cdots x_{\sigma(n)}^{\alpha_n}=\det\begin{pmatrix}x_1^{\alpha_1}&\cdots&x_1^{\alpha_n}\\ \vdots&&\vdots\\ x_n^{\alpha_1}&\cdots&x_n^{\alpha_n}\end{pmatrix}$$

are a vector space basis for the $\mathbb{Q}$-algebra of alternating polynomials, as α ranges over $\mathbb{N}^n_>$.[29]

[29] To prove linear independence we observe that $x_1^{\alpha_1}\cdots x_n^{\alpha_n}$ is the lexicographic leading term of Δ_α. If $\alpha\in\mathbb{N}^n_>$ is the lex-largest exponent appearing in a relation $f=\sum_{c\in\mathbb{N}^n_>}c_\beta\Delta_\beta=0$ then we have $f=c_\alpha\mathbf{x}^\alpha+$ lower terms $=0$, which implies that $c_\alpha=0$. Now use induction.

From this we obtain a new basis of symmetric polynomials. Note that we have a bijection $\mathbb{N}^n_> \to \mathbb{N}^n_\geq$ from **strictly** decreasing sequences to **weakly** decreasing sequences defined by

$$\begin{aligned}(\alpha_1, \ldots, \alpha_n) &\mapsto (\alpha_1 - (n-1), \alpha_2 - (n-2), \ldots, \alpha_{n-1} - 1, \alpha_n),\\ \alpha &\mapsto \alpha - \delta,\end{aligned}$$

where

$$\delta = (n-1, \ldots, 2, 1, 0).$$

We can think of $\mathbb{N}^n_\geq$ as the set of integer partitions with at most n parts. Note that $\Delta_\delta = \Delta$ is the classical Vandermonde determinant. For any $\alpha \in \mathbb{N}^n_>$ the theorem on alternating polynomials above tells us that $\Delta_\alpha = \Delta\, s_\alpha(x_1, \ldots, x_n)$ for some symmetric polynomial s_α, which we call a *Schur polynomial*. It is more traditional to index this polynomial by the integer partition $\lambda = \alpha - \delta \in \mathbb{N}^n_\geq$. Thus we define

$$s_\lambda(x_1, \ldots, x_n) = \Delta_{\delta+\lambda}(x_1, \ldots, x_n)/\Delta_\delta(x_1, \ldots, x_n).$$

It follows from the fact that the generalized Vandermonde determinants Δ_α ($\alpha \in \mathbb{N}^n_>$) are a $\mathbb{Q}$-basis for alternating polynomials that the Schur polynomials s_λ ($\lambda \in \mathbb{N}^n_\geq$) are a $\mathbb{Q}$-basis for symmetric polynomials.

Note that the Schur polynomials s_λ are defined in a quite different manner from the symmetric polynomials $e_\lambda, h_\lambda, p_\lambda$ in that there is no analogue of the polynomials e_k, h_k, p_k. The key property of the Schur polynomials that we need in this chapter is a "Vandermonde type" identity in two sets of variables, which is attributed to Cauchy.[30]

Theorem 4.5 (Cauchy identity)**.** *Let $\mathbb{N}^n_\geq$ be the set of weakly decreasing sequences of n non-negative integers, i.e., the set of integer partitions with at most n parts, and consider two independent sets of variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Then we have*

$$\prod_{i=1}^n \prod_{j=1}^n \frac{1}{1 - x_i y_j} = \sum_{\lambda \in \mathbb{N}^n_\geq} s_\lambda(x_1, \ldots, x_n) s_\lambda(y_1, \ldots, y_n).$$

Proof. Consider the expression

$$F(x_1, \ldots, x_n, y_1, \ldots, y_n) := \det\left(\frac{1}{1 - x_i y_j}\right) \prod_{i=1}^n \prod_{j=1}^n (1 - x_i y_j)$$

Note that each of the $n!$ terms of the $n \times n$ determinant has denominator that divides the product $\prod_{i=1}^n \prod_{j=1}^n (1 - x_i y_j)$, hence F is a polynomial in $\mathbb{Q}[x_1, \ldots, x_n, y_1, \ldots, y_n]$. Furthermore, we see that F is an alternating polynomial $x_1, \ldots, x_n$ since it switches two rows of the determinant and leaves the product invariant. Similarly, F is alternating in $y_1, \ldots, y_n$. It

[30]See historical notes in Stanley's EC2.

follows from a slight modification of the theorem on alternating polynomials that $F(\mathbf{x},\mathbf{y}) = \Delta(\mathbf{x})\Delta(\mathbf{y})C(\mathbf{x},\mathbf{y})$, where $\Delta(\mathbf{x}) = \prod_{1\le i<j\le n}(x_i - x_j)$, $\Delta(\mathbf{y}) = \prod_{1\le i<j\le n}(y_i - y_j)$ and $C(\mathbf{x},\mathbf{y}) \in \mathbb{Q}[\mathbf{x},\mathbf{y}]$. By comparing degrees one can check that $C(\mathbf{x},\mathbf{y})$ is constant and a more intricate analysis shows that in fact $C(\mathbf{x},\mathbf{y}) = 1$. At this point we have

$$\prod_{i=1}^{n}\prod_{j=1}^{n}\frac{1}{1-x_iy_j} = \frac{1}{\Delta(\mathbf{x})\Delta(\mathbf{y})}\det\left(\frac{1}{1-x_iy_j}\right).$$

The result will now follow if we can show that

$$\det\left(\frac{1}{1-x_iy_j}\right) = \sum_{\alpha\in\mathbb{N}^n_>}\Delta_\alpha(\mathbf{x})\Delta_\alpha(\mathbf{y}).$$

To prove this we expand each entry of the matrix using the geometric series

$$\frac{1}{1-x_ix_j} = \sum_{m\ge 0}(x_iy_j)^m.$$

Then the jth column of the matrix is the infinite linear combination

$$\sum_{m\ge 0} y_j^m \begin{pmatrix} x_1^m \\ \vdots \\ x_n^m \end{pmatrix}.$$

Since the determinant is multilinear in its columns, we have[31]

$$\det\left(\frac{1}{1-x_iy_j}\right) = \sum_{m_1,\ldots,m_n\ge 0} y_1^{m_1}\cdots y_n^{m_n}\det\begin{pmatrix} x_1^{m_1} & \cdots & x_1^{m_n} \\ \vdots & & \vdots \\ x_n^{m_1} & \cdots & x_n^{m_n} \end{pmatrix}.$$

Note that the $\mathbf{x}$-determinant vanishes unless the m_i are distinct. By summing over distinct $m_1,\ldots,m_n$ and permuting them into decreasing order we obtain

$$\begin{aligned}
\det\left(\frac{1}{1-x_iy_j}\right) &= \sum_{\sigma\in S_n}\sum_{\alpha\in\mathbb{N}^n_>}\mathbf{y}^{\sigma\cdot\alpha}\Delta_{\sigma\cdot\alpha}(\mathbf{x}) \\
&= \sum_{\sigma\in S_n}\sum_{\alpha\in\mathbb{N}^n_>}\mathbf{y}^{\sigma\cdot\alpha}\operatorname{sgn}(\sigma)\Delta_\alpha(\mathbf{x}) \\
&= \sum_{\alpha\in\mathbb{N}^n_>}\Delta_\alpha(\mathbf{x})\left(\sum_{\sigma\in S_n}\operatorname{sgn}(\sigma)\mathbf{y}^{\sigma\cdot\alpha}\right) \\
&= \sum_{\alpha\in\mathbb{N}^n_>}\Delta_\alpha(\mathbf{x})\Delta_\alpha(\mathbf{y}).
\end{aligned}$$

□

[31]Here I ignore any subtleties involved with formal power series-valued determinants.

4.5 The Frobenius character formula

In this section we present Frobenius' theorem on the characters of the symmetric group together with Frobenius' original proof.[32] Recall that the conjugacy classes of S_n are parametrized by "cycle types". Indeed, each permutation is a product of commuting cycles, and for any permutation σ and cycle $(i_1, \ldots, i_k)$ we have

$$\sigma \circ (i_1, \ldots, i_k) \circ \sigma^{-1} = (\sigma(i_1), \ldots, \sigma(i_k)).$$

We record the cycle type of a permutation $\pi \in S_n$ as an integer partition $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq 0) = (1^{m_1} 2^{m_2} \cdots)$, where m_i is the number of cycles of length i and where λ_i is the length of the ith longest cycle that occurs. We write $\lambda \vdash n$ to denote the fact that $\sum_i \lambda_i = n$. One can check that the number of permutations that commute with π is equal to

$$z_\lambda := 1^{m_1} m_1! 2^{m_2} m_2! \cdots,$$

hence the number of permutations of cycle type $\lambda \vdash n$ is $n!/z_\lambda$. Recall from the last chapter that the number of simple characters of a finite group equals the number of conjugacy classes. The conjugacy classes of S_n are in bijection with partitions of n (cycle types):

$$\mathrm{Par}(n) = \{\lambda \vdash n\}.$$

This suggests that one might be able to use the symmetric polynomials $e_\lambda, h_\lambda, p_\lambda, s_\lambda$ for $\lambda \vdash n$ to construct the characters of S_n. The following theorem does exactly this.

Theorem 4.6 (Frobenius character theorem, 1900). *Given an integer partition $\mu \vdash n$ we can expand the power sum symmetric polynomial $p_\mu(x_1, \ldots, x_n)$ in the Schur basis:*[33]

$$p_\mu(x_1, \ldots, x_n) = \sum_{\lambda \vdash n} \chi^\lambda(\mu)\, s_\lambda(x_1, \ldots, x_n).$$

For each $\lambda \vdash n$ we define the function $\chi^\lambda : S_n \to \mathbb{Q}$ by $\chi^\lambda(g) := \chi^\lambda(\mu)$ when g has cycle type $\mu \vdash n$. Then the functions χ^λ for $\lambda \vdash n$ are the simple characters of S_n.[34]

Here is a sketch of the proof we will give:

(1) Show that the functions χ^λ are an orthonormal basis for the space of class functions on S_n, with respect to the usual inner product of class functions.

(2) Show that each χ^λ is a $\mathbb{Z}$-linear combination of simple characters.

[32] See Curtis, *Pioneers of representation theory*, page 73.

[33] We remark that each of the polynomials $e_\lambda, h_\lambda, p_\lambda, s_\lambda$ is homogeneous of total degree $|\lambda| = \sum_i \lambda_i$. Thus the s_λ expansion of p_μ involves only partitions with $|\lambda| = |\mu|$.

[34] In fact, the Schur polynomials are a $\mathbb{Z}$-basis of Λ_n so the characters of S_n are integer valued, and one might hope for a combinatorial interpretation of these numbers. The Murnaghan-Nakayama rule gives such an interpretation, but we don't discuss it here.

It follows that the change of basis matrix between the functions χ^λ and the simple characters of S_n is an orthogonal matrix with integer entries, hence it is a signed permutation matrix, which implies that each χ^λ equals ± 1 times a simple character. Finally, one can check that $\chi^\lambda(\mathrm{id}) = \chi^\lambda(1, 1, \ldots, 1)$ is positive, hence χ^λ equals $+1$ times a simple character.

Proof. The key to the proof of part (1) is the "Cauchy kernel"

$$\Omega(\mathbf{x}, \mathbf{y}) := \prod_{i=1}^{n} \prod_{j=1}^{n} \frac{1}{1 - x_i y_j}.$$

We saw in the previous chapter that $\Omega(\mathbf{x}, \mathbf{y}) = \sum_\lambda s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y})$. On the other hand, taking the logarithm of Ω gives

$$\begin{aligned} \log \Omega(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \geq 1} \frac{(x_i y_j)^k}{k} \\ &= \sum_{k \geq 1} \frac{p_k(x_1, \ldots, x_n) p_k(y_1, \ldots, y_n)}{k}. \end{aligned}$$

Then taking the exponential gives

$$\begin{aligned} \Omega(\mathbf{x}, \mathbf{y}) &= \exp\left(\sum_{k \geq 1} \frac{p_k(x_1, \ldots, x_n) p_k(y_1, \ldots, y_n)}{h} \right) \\ &= \sum_{m \geq 0} \frac{1}{m!} \left(\sum_{k \geq 1} \frac{p_k(x_1, \ldots, x_n) p_k(y_1, \ldots, y_n)}{k} \right)^m \\ &= \sum_{m \geq 0} \frac{1}{m!} \sum_{\sum m_i = m} \binom{m}{m_1, m_2, \ldots} \left(\frac{p_1(\mathbf{x}) p_1(\mathbf{y})}{1} \right)^{m_1} \left(\frac{p_2(\mathbf{x}) p_2(\mathbf{x})}{2} \right)^{m_2} \cdots \\ &= \sum_{m \geq 0} \sum_{\sum m_i = m} \frac{(p_1(\mathbf{x}) p_1(\mathbf{y}))^{m_1} (p_2(\mathbf{x}) p_2(\mathbf{x}))^{m_2} \cdots}{1^{m_1} m_1! 2^{m_2} m_2! \cdots} \\ &= \sum_{\mu} \frac{1}{z_\mu} p_\mu(\mathbf{x}) p_\mu(\mathbf{y}), \end{aligned}$$

where $\lambda = (1^{m_1} 2^{m_2} \cdots)$ runs over all integer partitions. Hence we have

$$\sum_{\mu} \frac{1}{z_\mu} p_\mu(\mathbf{x}) p_\mu(\mathbf{y}) = \sum_{\lambda} s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y}).$$

Each side is an infinite sum. By looking at the homogeneous parts of bi-degree (n, n) we obtain a finite sum over partitions of n:

$$\sum_{\mu \vdash n} \frac{1}{z_\mu} p_\mu(\mathbf{x}) p_\mu(\mathbf{y}) = \sum_{\lambda \vdash n} s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y}).$$

Substituting $p_\mu(\mathbf{x}) = \sum_{\lambda_1 \vdash n} \chi^{\lambda_1}(\mu) s_{\lambda_1}(\mathbf{x})$ and $p_\mu(\mathbf{y}) = \sum_{\lambda_2 \vdash n} \chi^{\lambda_2}(\mu) s_{\lambda_1}(\mathbf{y})$ gives

$$\sum_{\lambda_1, \lambda_2 \vdash n} \left(\sum_{\mu \vdash n} \frac{1}{z_\mu} \chi^{\lambda_1}(\mu) \chi^{\lambda_2}(\mu) \right) s_{\lambda_1}(\mathbf{x}) s_{\lambda_2}(\mathbf{y}) = \sum_{\lambda} s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y}).$$

Since the sets $\{s_\lambda(\mathbf{x}) : \lambda \vdash n\}$ and $\{s_\lambda(\mathbf{y}) : \lambda \vdash n\}$ are linearly independent over $\mathbb{Q}$, one can check that the set $\{s_{\lambda_1}(\mathbf{x}) s_{\lambda_2}(\mathbf{y}) : \lambda_1, \lambda_2 \vdash n\}$ is linearly independent over $\mathbb{Q}$, hence we have

$$\sum_{\mu \vdash n} \frac{1}{z_\mu} \chi^{\lambda_1}(\mu) \chi^{\lambda_2}(\mu) = \begin{cases} 1 & \lambda_1 = \lambda_2, \\ 0 & \lambda_1 \neq \lambda_2. \end{cases}$$

Finally, recall the inner product of class functions. Given two functions $\varphi, \psi : G \to \mathbb{C}$ on a finite group G we defined the Hermitian inner product

$$\langle \varphi, \psi \rangle = \frac{1}{\#G} \sum_{g \in G} \varphi(g) \psi(g)^*,$$

and we showed that the simple characters of G are orthonormal with respect to this inner product. Since the number of permutations of cycle type μ is $n!/z_\mu$, we also have

$$\begin{aligned} \langle \chi^{\lambda_1}, \chi^{\lambda_2} \rangle &= \frac{1}{\#S_n} \sum_{g \in S_n} \chi^{\lambda_1}(g) \chi^{\lambda_2}(g) \\ &= \frac{1}{n!} \sum_{\mu \vdash n} \frac{n!}{z_\mu} \chi^{\lambda_1}(\mu) \chi^{\lambda_2}(\mu) \\ &= \sum_{\mu \vdash n} \frac{1}{z_\mu} \chi^{\lambda_1}(\mu) \chi^{\lambda_2}(\mu). \end{aligned}$$

Thus we have shown that the functions χ^λ for $\lambda \vdash n$ are an orthonormal basis for the space of class functions.

To prove part (2) we need to show that each function χ^λ is a $\mathbb{Z}$-linear combination of simple characters of S_n. Since any character is a $\mathbb{Z}$-linear combination of simple characters we only need to show that χ^λ is a $\mathbb{Z}$-linear combination of characters. In fact, we will show that

$$\chi^\lambda(\mu) = \sum_{\tau \in S_n} \operatorname{sgn}(\tau) \psi^\lambda_\tau(\mu),$$

where the functions $\psi^\lambda_\tau : S_n \to \mathbb{Z}$ are certain characters to be defined shortly (some of which may be the zero function).

Given any non-negative integer vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ we let X_α denote the set of *ordered partial set partitions* of $\{1, \ldots, n\}$ of type α; that is the set of n-tuples $(B_1, \ldots, B_n)$, where some B_i may be empty, where $\{1, \ldots, n\}$ is the disjoint union of the non-empty parts, and where $\#B_i = \alpha_i$ for all i. The group S_n acts on X_α by permuting entries:[35]

$$\sigma \cdot (B_1, \ldots, B_n) := (\sigma(B_1), \ldots, \sigma(B_n)).$$

Then we define $\phi^\alpha(\sigma)$ as the number of fixed points:[36]

$$\phi^\alpha(\sigma) := \#\{B \in X_\alpha : \sigma \cdot B = B\}.$$

This is a character of S_n. Indeed, for any action of S_n on any finite set X we define a representation of S_n on the vector space $\mathbb{C}[X]$ of formal $\mathbb{C}$-linear combinations of elements of X. This turns elements $\sigma \in S_n$ into permutation matrices of size $\#X \times \#X$ and the trace of the matrix of σ is just the number of basis vectors (i.e., elements of X) fixed by σ.

If $\mu \vdash n$ is the cycle type of a permutation $\sigma \in S_n$ then I claim that $\phi^\alpha(\sigma)$ equals the coefficient of $\mathbf{x}^\alpha$ in the power sum symmetric polynomial $p_\mu(x_1, \ldots, x_n)$:

$$[\mathbf{x}^\alpha] p_\mu(x_1, \ldots, x_n) = \phi^\alpha(\sigma).$$

To see this, we first observe that a partition $(B_1, \ldots, B_n)$ is fixed by $\sigma \in S_n$ if and only if each B_i is a union of cycles of σ. Thus for a given σ we see that $\phi^\alpha(\sigma)$ is the number of ways to distribute the cycles of σ into n ordered boxes so that the total length of the cycles in the ith box equals α_i. Suppose σ has m_i cycles of length i, so the cycle type is $\mu = (1^{m_1} 2^{m_2} \cdots)$ with $m_1 + 2m_2 + \cdots = n$. Then we have

$$p_\mu(x_1, \ldots, x_n) = \prod_{k \geq 1} (x_1^k + \cdots + x_n^k)^{m_k}.$$

For any exponent vector $\alpha \in \mathbb{N}^n$ we note that the coefficient of $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ in the expansion of p_μ solves exactly the same counting problem. That is, we can associate each factor of $p_k = x_1^k + \cdots + x_n^k$ in the product p_μ with a specific k-cycle of σ. To obtain a monomial in the expansion of p_μ we choose a summand from factor p_k. Choosing the summand x_i^k from a given factor p_k corresponds to putting the associated k-cycle into the ith box.

Finally, we relate these numbers to the functions χ^λ. Recall from the definitions that

$$p_\mu(x_1, \ldots, x_n) = \sum_{\lambda \vdash n} \chi^\lambda(\mu) s_\lambda(x_1, \ldots, x_n),$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. And recall that we defined the Schur polynomials as as ratios of Vandermonde determinants:

$$s_\lambda(x_1, \ldots, x_n) = \Delta_{\lambda+\delta}(x_1, \ldots, x_n) / \Delta_\delta(x_1, \ldots, x_n),$$

[35] Here S_n acts on subsets of $\{1, \ldots, n\}$ in the obvious way.

[36] Frobenius also viewed ϕ^α as the character of S_n "induced" from the trivial character on the subgroup $S_{B_1} \times \cdots \times S_{B_n}$. In these notes we will have no need for induced characters.

where $\Delta_\alpha(x_1,\dots,x_n) = \det(x_i^{\alpha_j}) = \sum_{\tau\in S_n} \operatorname{sgn}(\tau)\mathbf{x}^{\tau\cdot\alpha}$ and $\delta = (n-1,\dots,1,0)$. Multiplying both sides by the standard Vandermonde determinant $\Delta = \Delta_\delta$ gives

$$\Delta(\mathbf{x})p_\mu(\mathbf{x}) = \sum_{\lambda\vdash n} \chi^\lambda(\mu)\Delta_{\lambda+\delta}(\mathbf{x}) = \sum_{\lambda\vdash n} \chi^\lambda(\mu) \sum_{\tau\in S_n} \operatorname{sgn}(\tau)\mathbf{x}^{\tau\cdot(\lambda+\delta)}.$$

From this we note that $\chi^\lambda(\mu)$ is the coefficient of $\mathbf{x}^{\lambda+\delta}$ in $\Delta(\mathbf{x})p_\mu(\mathbf{x})$. On the other hand, we showed above that $\phi^\alpha(\mu)$ is the coefficient of $\mathbf{x}^\alpha$ in $p_\mu(\mathbf{x})$. Hence we have

$$\begin{aligned}
\chi^\lambda(\mu) &= [\mathbf{x}^{\lambda+\delta}]\Delta(\mathbf{x})p_\mu(\mathbf{x}) \\
&= [\mathbf{x}^{\lambda+\delta}] \sum_{\tau\in S_n} \operatorname{sgn}(\tau)\mathbf{x}^{\tau\cdot\delta}p_\mu(\mathbf{x}) \\
&= \sum_{\tau\in S_n} \operatorname{sgn}(\tau)[\mathbf{x}^{\lambda+\delta}]\mathbf{x}^{\tau\cdot\delta}p_\mu(\mathbf{x}) \\
&= \sum_{\tau\in S_n} \operatorname{sgn}(\tau)[\mathbf{x}^{(\lambda+\delta)-\tau\cdot\delta}]p_\mu(\mathbf{x}) \\
&= \sum_{\tau\in S_n} \operatorname{sgn}(\tau)\phi^{(\lambda+\delta)-\tau\cdot\delta}(\mu).
\end{aligned}$$

where we define $\phi^{(\lambda+\delta)-\tau\cdot\delta}(\mu) = 0$ if the vector $(\lambda+\delta)-\tau\cdot\delta$ has any negative entries. Finally, we define for each partition $\lambda \vdash n$ and and permutation $\tau \in S_n$ the function $\psi_\tau^\lambda(\mu) := \phi^{(\lambda+\delta)-\tau\cdot\delta}(\mu)$ which is either a character or zero. □

4.6 The Frobenius characteristic map

The ideas in the proof of his character theorem led Frobenius to consider a certain linear map from the set of class functions $CF(S_n) = \{f : S_n \to \mathbb{Q}\}$ to symmetric polynomials $\Lambda_n = \mathbb{Q}[x_1,\dots,x_n]^{S_n}$. More specifically, let $\Lambda_n^{(k)}$ denote the subspace of Λ_n consisting of symmetric polynomials of total degree k. Then we define the *Frobenius characteristic*[37]

$$\operatorname{ch} : CF(S_n) \to \Lambda_n^{(n)}$$

by sending each class function φ to the symmetric polynomial

$$\operatorname{ch}(\varphi) = \sum_{\lambda\vdash n} \frac{\varphi(\lambda)}{z_\lambda}\, p_\lambda(x_1,\dots,x_n),$$

where we let $\varphi(\lambda)$ denote the value of φ on any permutation of cycle type λ. This map is an isomorphism of $\mathbb{Q}$-linear spaces because the polynomials $p_\lambda(x_1,\dots,x_n)$ with $\lambda \vdash n$ are a basis

[37]We should mention that the number of variables in the symmetric polynomials is not important in this context. Only the degree of the polynomials matters. It is common to let the number of variables go to infinity.

for $\Lambda_n^{(n)}$. Note that we have $\mathrm{ch}(\iota_\lambda) = p_\lambda/z_\lambda$ where $\iota_\lambda : S_n \to \mathbb{Q}$ is the indicator function that equals 1 on permutations of type λ and 0 otherwise.

The definition of the Frobenius characteristic map seems arbitrary at first but we will try to demonstrate in this section why it is natural. First, we note that the simple characters get sent to Schur polynomials:

$$\mathrm{ch}(\chi^\lambda) = s_\lambda(x_1, \ldots, x_n).$$

The quickest way to see this is to use the orthogonality relations. From the definition of χ^λ we have $p_\mu = \sum_{\lambda \vdash n} \chi^\lambda(\mu)\, s_\lambda$. We can rewrite this as the matrix equation $(p_\mu) = C \cdot (s_\lambda)$ where C is the character table of S_n, i.e., the square matrix with μ, λ entry $\chi^\lambda(\mu)$. The orthogonality relations in the above proof say that CC^T is the diagonal matrix with diagonal entries z_μ. It follows that C^{-1} is the matrix with μ, λ entry $\chi^\mu(\lambda)/z_\lambda$. Then the matrix equation $(s_\mu) = C^{-1} \cdot (p_\lambda)$ says that

$$s_\mu(x_1, \ldots, x_n) = \sum_{\lambda \vdash n} \frac{\chi^\mu(\lambda)}{z_\lambda} p_\lambda(x_1, \ldots, x_n) = \mathrm{ch}(\chi^\mu),$$

as desired. Recall that the space of class functions $CF(S_n)$ contains the characters as a discrete cone, i.e., the set of $\mathbb{N}$-linear combinations of the simple characters. The image of this cone is the set of $\mathbb{N}$-linear combinations of Schur polynomials. This is the reason for the concept of "Schur positivity" in symmetric function theory — the symmetric polynomials with non-negative integer Schur coefficients are exactly those that correspond to characters.

Next we consider the permutation characters ϕ^α from the proof of the character theorem. I claim that these correspond to the complete homogeneous symmetric polynomials:

$$\mathrm{ch}(\phi^\alpha) = h_\alpha(x_1, \ldots, x_n) := \prod_i h_{\alpha_i}(x_1, \ldots, x_n).$$

To prove this, recall that we defined $\phi^\alpha(\lambda)$ as the coefficient of $\mathbf{y}^\alpha$ in $p_\lambda(y_1, \ldots, y_n)$. It follows that $\mathrm{ch}(\phi^\alpha)(\mathbf{x})$ is the coefficient of $\mathbf{y}^\alpha$ in the Cauchy kernel:

$$\begin{aligned}
[\mathbf{y}^\alpha]\Omega(\mathbf{x}, \mathbf{y}) &= [\mathbf{y}^\alpha] \sum_\mu \frac{1}{z_\lambda} p_\lambda(\mathbf{x}) p_\lambda(\mathbf{y}) \\
&= \sum_\mu \frac{[\mathbf{y}^\alpha] p_\lambda(\mathbf{y})}{z_\lambda} p_\lambda(\mathbf{x}) \\
&= \sum_\mu \frac{\phi^\alpha(\lambda)}{z_\lambda} p_\lambda(\mathbf{x}) \\
&= \mathrm{ch}(\phi^\alpha)(\mathbf{x}).
\end{aligned}$$

On the other hand, we can expand the Cauchy kernel in terms of h_α. Recall that

$$H(t) = \prod_{i=1}^n \frac{1}{1 - x_i t} = \sum_{k \geq 0} h_k(x_1, \ldots, x_n) t^k.$$

Thus we have

$$\begin{aligned}\Omega(\mathbf{x},\mathbf{y}) &= \prod_{i=1}^{n}\prod_{j=1}^{n}\frac{1}{1-x_iy_j}\\ &= \prod_{j=1}^{n}\sum_{k\geq 0} h_k(x_1,\ldots,x_n)y_j^k\\ &= \sum_{\alpha\in\mathbb{N}^n_>} h_\alpha(x_1,\ldots,x_n)\mathbf{y}^\alpha,\end{aligned}$$

and hence $[\mathbf{y}^\alpha]\Omega(\mathbf{x},\mathbf{y}) = h_\alpha(x_1,\ldots,x_n)$. As an interesting consequence, if we apply the Frobenius characteristic to both sider of the Frobenius character formula

$$\chi^\lambda = \sum_{\tau\in S_n}\mathrm{sgn}(\tau)\phi^{(\lambda+\delta)-\tau\cdot\delta}$$

then we obtain the expansion of the Schur polynomials in terms of the h-basis:

$$s_\lambda = \sum_{\tau\in S_n}\mathrm{sgn}(\tau)h_{(\lambda+\delta)-\tau\cdot\delta} = \det(h_{\lambda_i+j-i}),$$

where we adopt the convention that $h_k(x_1,\ldots,x_n) = 0$ when $k < 0$. This is called the *Jacobi-Trudi* identity.

Next we consider the elementary polynomials e_λ. Recall that we have a $\mathbb{Q}$-algebra involution $\omega : \Lambda_n \to \Lambda_n$ defined by $\omega(h_k) = e_k$, which implies that $\omega(h_\lambda) = e_\lambda$. We would like to understand the involution $\mathrm{ch}^{-1}\circ\omega\circ\mathrm{ch}$ on class functions. I claim that

$$\mathrm{ch}^{-1}(\omega(\mathrm{ch}(\varphi)))(\sigma) = \varphi(\sigma)\mathrm{sgn}(\sigma) \quad \text{for all } \sigma\in S_n \text{ and } \varphi\in CF(S_n).$$

More compactly: $\omega(\mathrm{ch}(\varphi)) = \mathrm{ch}(\varphi\cdot\mathrm{sgn})$. In other words, the involution ω on symmetric polynomials corresponds to "tensoring with the sign representation". To prove this, we recall the identity $\omega(p_k) = (-1)^{k-1}p_k$ from the section on power sums. We also note that the sign of a k-cycle in S_n is $(-1)^{k-1}$. If $\mathrm{sgn}(\lambda)$ denotes the sign of a permutation with cycle type $\lambda\vdash n$ then this implies that $\omega(p_\lambda) = \mathrm{sgn}(\lambda)p_\lambda$, hence for any class function φ we have

$$\begin{aligned}\mathrm{ch}(\varphi\cdot\mathrm{sgn}) &= \sum_{\lambda\vdash n}\frac{\varphi(\lambda)\mathrm{sgn}(\lambda)}{z_\lambda}\,p_\lambda\\ &= \sum_{\lambda\vdash n}\frac{\varphi(\lambda)}{z_\lambda}\,\omega(p_\lambda)\\ &= \omega\left(\sum_{\lambda\vdash n}\frac{\varphi(\lambda)}{z_\lambda}\,p_\lambda\right)\\ &= \omega(\mathrm{ch}(\varphi)).\end{aligned}$$

Since $\omega(h_\lambda) = e_\lambda$, this result tells us that $\text{ch}^{-1}(e_\lambda)$ is the character $\phi^\lambda \cdot \text{sgn}$.

Finally, we define the *Hall inner product* on symmetric polynomials via the standard inner product on class functions. That is, for any symmetric polynomials f, g we define

$$\langle f, g\rangle := \langle \text{ch}^{-1}(f), \text{ch}^{-1}(g)\rangle = \sum_{\lambda \vdash n} \frac{1}{z_\lambda} \text{ch}^{-1}(f)(\lambda)\text{ch}^{-1}(g)(\lambda).$$

Since the Schur polynomials correspond to the simple characters of S_n, this is also the unique inner product on Λ_n making the Schur polynomials into an orthonormal basis:

$$\langle s_\lambda, s_\mu\rangle = \begin{cases} 1 & \lambda = \mu, \\ 0 & \lambda \neq \mu. \end{cases}$$

The Hall inner product explains the mysterious significance of the Cauchy kernel $\Omega(\mathbf{x}, \mathbf{y}) = \prod_{i,j}(1 - x_i y_j)^{-1}$. And we will motivate this by an analogy with elementary linear algebra. Let U and V be square matrices over $\mathbb{Q}$ with columns $\mathbf{u}_i$ and $\mathbf{v}_i$ and let I be the identity matrix. Then we have

$$I = UV^T \quad \Longleftrightarrow \quad I = U^T V.$$

In terms of the column vectors this becomes

$$I = \sum_i \mathbf{u}_i \mathbf{v}_i^T \quad \Longleftrightarrow \quad \mathbf{u}_i^T \mathbf{v}_j = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

In particular, if the $\mathbf{u}_i$ are an orthonormal basis then $I = \sum_i \mathbf{u}_i \mathbf{u}_i^T$. In the context of general vector spaces we should replace the dot product $\mathbf{u}_i^T \mathbf{v}_j$ with an inner product $\langle \mathbf{u}_i, \mathbf{v}_j\rangle$ and the rank one matrix $\mathbf{u}_i \mathbf{v}_i^T$ with the tensor product $\mathbf{u}_i \otimes \mathbf{v}_i$.

Now consider the rings of symmetric polynomials $\Lambda_n(\mathbf{x})$ and $\Lambda_n(\mathbf{y})$ in two different sets of variables $\mathbf{x}$ and $\mathbf{y}$. We will think of elements of $\Lambda_n(\mathbf{x})$ as "column vectors" and elements of $\Lambda_n(\mathbf{y})$ as "row vectors" and we will view the bijection $f(\mathbf{x}) \leftrightarrow f(\mathbf{y})$ as "transposition". One can use the Cauchy identity $\Omega(\mathbf{x}, \mathbf{y}) = \sum_\lambda s_\lambda(\mathbf{x}) s_\lambda(\mathbf{y})$ to prove the following result relating the Cauchy kernel and the Hall inner product: Given any two bases $u_\lambda(\mathbf{x}), v_\lambda(\mathbf{x})$ for the vector space $\Lambda_n(\mathbf{x})$ we have

$$\Omega(\mathbf{x}, \mathbf{y}) = \sum_\lambda u_\lambda(\mathbf{x}) v_\lambda(\mathbf{y}) \quad \Longleftrightarrow \quad \langle u_\lambda(\mathbf{y}), v_\mu(\mathbf{y})\rangle = \begin{cases} 1 & \lambda = \mu, \\ 0 & \lambda \neq \mu. \end{cases}$$

In this sense, the Cauchy kernel is analogous to the "identity matrix". The key formula that makes the proofs work is

$$\langle f(\mathbf{x}), \Omega(\mathbf{x}, \mathbf{y})\rangle_{\mathbf{x}} = f(\mathbf{y}),$$

where $\langle -, -\rangle_{\mathbf{x}}$ is the Hall inner product with respect to the $\mathbf{x}$ variables. In the language of functional analysis, this says that $\Omega(\mathbf{x}, \mathbf{y})$ is a "reproducing kernel" for the Hall inner product. This explains the use of the word "kernel".

All of this is quite intricate, but it is in some sense implicit in the notion of "representations of S_n". Given enough time, anyone interested in the subject would have to rediscover these ideas in some form. The modern context for these ideas is the theory of Hopf algebras.

5 Groups generated by reflections

6 Rings of invariants

7 Diagonal invariants