NST IB Mathematical Methods III

Easter Term 2014
Dr Adrian Kent

1. Small Oscillations

2. Groups

3. Representation Theory

Recommended books:

H.F. Jones, *Groups, Representations and Physics*, 2nd edn. (Inst. of Physics, 1998) [Nicely written. Good for the group theory and representation theory covered in this course, and beyond it to topics like SU(2) and SO(3). Approximately £35 new (as of April 2014).]

K.F. Riley, M.P. Hobson & S.J. Bence *Mathematical Methods for Physics and Engineering*, 3rd edn. (Cambridge University Press, 2006)

Freely available online to University of Cambridge users via Cambridge Books Online. Can be accessed by searching for ebook versions in the University Library online catalogue and following the links displayed. [Huge book, 1333 pp., *not* too expensive if you wish to buy a copy – approximately £33-40 new (as of April 2014). Good reference for most of the material in this course and many other topics as well.]

J. Mathews & R.L. Walker, *Mathematical Methods*, 2nd edn. (Benjamin/Cummings, 1970) [Not strongly recommended. Very expensive, approximately £100 new (as of April 2014).]

[†]21/4/2012 edn. Notes originated by N S Manton and S T C Siklos. Revised by H Osborn, R M Williams and M E McIntyre, with further small changes by M B Green, N S Manton and A Kent. For online copies, go to
http:www.damtp.cam.ac.uk/user/examples and find N23L.

Chapter 1

1. Small oscillations

This chapter contains the first set of syllabus items: small oscillations and equilibrium; normal modes and normal coordinates, examples, e.g. vibrations of linear molecules such as CO_2.

Administrative note: All footnotes are non-examinable: you don't have to read them. Other sections in small print are either (a) administrative, like this section, or (b) non-examinable if flagged as such, like the Preamble below, or (c) worked examples that will not be covered in the lectures, or (d) reminders of relevant parts of earlier courses. For this course it's recommended that you revise *vector spaces and matrices*, including *orthogonal and unitary transformations*. Also needed will be *Lagrange's equations* from dynamics.

As far as I am aware these printed notes are accurate. However, if anyone detects an error or obscurity, no matter how minor, please email me as soon as possible (`A.P.A.Kent@damtp.cam.ac.uk`). Or, if you prefer anonymity, email `nst@maths.cam.ac.uk`. Any necessary corrections will go into the online copy of the notes (N23L at http://www.damtp.cam.ac.uk/user/examples). See the date at bottom left of page 1.

Preamble (non-examinable)

Science, like ordinary perception, works by fitting models to data. (Models are partial, approximate representations of reality, with some models fitting better than others. An important class of models in the natural sciences comes under the heading *classical theory of small oscillations*, small-oscillations theory for short. It is the first topic in these lectures.

The classical theory has countless applications, ranging from probing the Sun's interior to environmental noise control, the design of earthquake-resistant structures, jet-engine and mobile-phone design — and on and on. And it is an essential prelude to the quantum theory of molecular vibrations and energy levels. That in turn tells you how lasers work. It also explains the physics of the qubits (the quantum version of a bit) and quantum logic gates from which some types of quantum computer are or (it is hoped) will be built. It explains too why, for instance, oxygen and nitrogen but not CO_2 are transparent to infrared radiation — this involves small oscillations on femtosecond timescales — and why life as

we know it depends on temperatures staying within a certain range. Small-oscillations theory underlies all of this.

1.1 Examples

Example 1.1 *The simple pendulum*

For small oscillations in a fixed vertical plane, the pendulum is a system with one degree of freedom exhibiting *harmonic motion*, in the sense of simple sinusoidal motion.

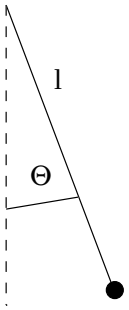


Fig. 1: A simple pendulum of length l with a mass m.

We regard the pendulum as a massless rod of constant length l, pivoted to a fixed point at one end and with a point mass m attached at the other. For angular displacement $\theta(t)$ the equation of motion is[1]

$$ml\ddot{\theta} = -mg\sin\theta$$

where the dots mean differentiation with respect to time t. For small θ, the equation linearizes to

$$\ddot{\theta} = -\frac{g}{l}\theta$$

whose general solution is

$$\theta = A\sin\omega(t - t_0), \qquad \text{where } \omega^2 = g/l\ .$$

The solution has two arbitrary constants of integration, the amplitude A and the phase or *phase angle*, ωt_0. ⌟

[1]The tangential acceleration is simply $l\ddot{\theta}$ because of the condition l = constant, a *constraint* on the motion; without that constraint there'd be another contribution $2\dot{l}\dot{\theta}$.

Example 1.2 *Coupled pendula: equal masses*

First we take identical pendula with masses m on massless rods of length l, joined by a massless spring with restoring force $k \times$ extension. The length of the spring is such that the system is in static equilibrium when the pendula are vertical. Again we assume that motion is confined to a vertical plane.

The extension of the spring is $l(\theta_2 - \theta_1)$, to sufficient accuracy in the linearization for small θ_1 and θ_2.

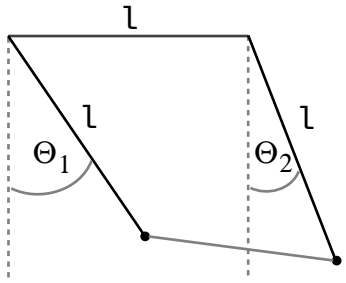


Fig. 2: Coupled pendula of equal masses and lengths l.

For small oscillations, therefore, the equations of motion are

$$ml\ddot{\theta}_1 = -mg\theta_1 + kl(\theta_2 - \theta_1) \quad (1.1)$$
$$ml\ddot{\theta}_2 = -mg\theta_2 + kl(\theta_1 - \theta_2). \quad (1.2)$$

A typical solution for θ_1 would look quite complicated:

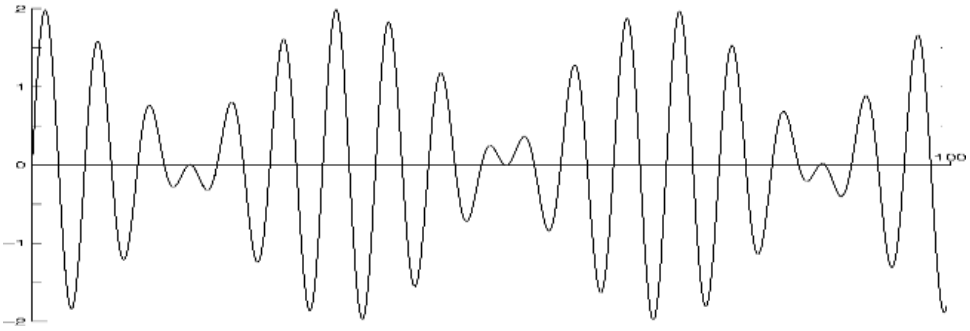


Fig. 3: The generic behaviour for $\theta_1(t)$ is not harmonic motion. (The expression (1.3) below is plotted for $\omega = 1$, $\Omega = \sqrt{1.4}$, $A = B = 1$.)

Energy is transferred back and forth between the two pendula. This is *not* harmonic (i.e. simple sinusoidal) motion. However, there *are* harmonic solutions of the coupled equations, which are obvious both from inspection of the equations and from physical intuition:

(i) In-phase solution, $\theta_1 = \theta_2$. The joining spring exerts no force, since its length doesn't change. Each of the equations reduces to the same form, e.g.

$$ml\ddot{\theta}_1 = -mg\theta_1 \ ,$$

as if the two pendula were uncoupled. So $\theta_1 = \theta_2 = A\sin\omega(t - t_0)$, where again A and t_0 are arbitrary constants and $\omega^2 = g/l$.

(ii) 180° out-of-phase solution, $\theta_1 = -\theta_2$. Each equation again reduces to the same form, e.g.

$$ml\ddot{\theta}_1 = -(mg + 2kl)\theta_1 \ ,$$

with solution

$$\theta_1 = B\sin\Omega(t - t_1),$$

where $\Omega^2 = g/l + 2k/m$. Notice that $\Omega > \omega$, though not by much if $2kl/mg$ is small.

Each of these two special solutions has a single, pure frequency, namely

$$\omega = \sqrt{g/l} \quad \text{in case (i)}, \qquad \Omega = \sqrt{g/l + 2k/m} \quad \text{in case (ii)}.$$

The two simple modes of oscillation represented by these special solutions are called **normal modes** of oscillation. Their frequencies are called **normal frequencies**, short for 'normal-mode frequencies'.

The general solution for θ_1 can be written as linear combinations of the normal-mode solutions, involving the four arbitrary constants A, B, t_0 and t_1, i.e.

$$\theta_1 = A\sin\omega(t - t_0) + B\sin\Omega(t - t_1) \ . \tag{1.3}$$

It is the general solution because (i) it certainly satisfies the equations, being just the sum of two solutions — the **principle of superposition** applying because the equations are linear — and (ii) it contains the correct number, four, of arbitrary constants. The corresponding general solution for θ_2 is evidently

$$\theta_2 = A\sin\omega(t - t_0) - B\sin\Omega(t - t_1) \ . \tag{1.4}$$

This follows from the relations $\theta_1 = \theta_2$ and $\theta_1 = -\theta_2$ already noted, respectively, for the in-phase ($A \neq 0,\ B = 0$) and out-of-phase ($A = 0,\ B \neq 0$) solutions.

If Ω/ω is not equal to any rational number (ratio of integers), then the general solution is not periodic. The case of figure 3 above illustrates this point, since $\Omega/\omega = \sqrt{1.4} = \sqrt{(7/5)}$.

The special linear combinations $\theta_1 + \theta_2$ and $\theta_1 - \theta_2$, which do oscillate simple harmonically and are, therefore, periodic — specifically, $\theta_1 + \theta_2 = 2A \sin \omega(t - t_0)$, and $\theta_1 - \theta_2 = 2B \sin \Omega(t - t_1)$ — are called the **normal coordinates** for this problem. ⌟

Example 1.3 *Coupled pendula with unequal masses; Lagrange's equations*

The setup is as in Example 1.2 except that the masses are taken to be different, m_1 and m_2 say. This time we use a more sophisticated approach, which will point toward the general theory.

Although we could use the natural coordinates $\mathbf{x}_1$ and $\mathbf{x}_2$ (the position vectors of the masses) together with the two constraints (the two rods have constant lengths), it is easier to use the generalized coordinates θ_1 and θ_2 as before. We can then use a standard result of dynamics, *Lagrange's equations*, a recipe for writing down the equations of motion for a very general class of systems.

(Recall that **generalized coordinates** in dynamics means *any* set of numbers that specifies the **configuration** of the system at a given instant. The configuration means the spatial locations and orientations of all its components, hence the distribution of mass in the system, taking account of any constraints — in this case the fact that l = constant. The generalized coordinates can be angles, lengths, areas, volumes or whatever is relevant and convenient. Here, of course, it's angles that are relevant and convenient.)

In our case there are two generalized coordinates and therefore two Lagrange's equations, one of which is

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}_1}\right) - \frac{\partial \mathcal{L}}{\partial \theta_1} = 0$$

and the other the same with suffix 2 instead of 1. Here the *Lagrangian function* $\mathcal{L}$, 'Lagrangian' for short, is defined to be the difference

$$\mathcal{L} = T - V$$

where T and V are the kinetic and potential energies of the whole system *expressed in terms of the generalized coordinates and their time derivatives*, in our case θ_1, θ_2, $\dot\theta_1$, and $\dot\theta_2$.

(The partial derivatives in Lagrange's equations then make sense. For instance $\partial\mathcal{L}/\partial\dot\theta_1$ means, by definition, differentiation that varies $\dot\theta_1$ while artificially holding θ_1, θ_2, and $\dot\theta_2$ constant. That is, these partial derivatives refer only to the prescribed functional dependence of $\mathcal{L}$ upon its four arguments θ_1, θ_2, $\dot\theta_1$, and $\dot\theta_2$. At this stage the symbols θ_1, θ_2, $\dot\theta_1$, and $\dot\theta_2$ mean just the four arguments of $\mathcal{L}$, not the functions of time required to describe the dynamical evolution.)

For our system the total kinetic energy T is exactly $\frac{1}{2}m_1l^2\dot\theta_1^2 + \frac{1}{2}m_2l^2\dot\theta_2^2$, a function of $\dot\theta_1, \dot\theta_2$ alone, and the total potential energy is evidently a function $V(\theta_1, \theta_2)$ of θ_1 and θ_2 alone — i.e. V, being the gravitational plus the elastic energy, is a function of the system configuration alone and not of its rate of change. Therefore $\mathcal{L}$ has the form

$$\mathcal{L} = \tfrac{1}{2}m_1l^2\dot\theta_1^2 + \tfrac{1}{2}m_2l^2\dot\theta_2^2 - V(\theta_1, \theta_2)\ .$$

(Thus, for instance, $\partial\mathcal{L}/\partial\dot\theta_1 = m_1l^2\dot\theta_1$. And remember that the d/dt in Lagrange's equations is a different animal altogether; it does, of course, refer to the time dependence of the dynamical evolution; thus $(d/dt)\,(\partial\mathcal{L}/\partial\dot\theta_1) = m_1l^2\ddot\theta_1$.)

The total potential energy $V(\theta_1,\, \theta_2)$ is exactly

$$\begin{aligned} V \;=\; & m_1gl(1-\cos\theta_1) + m_2gl(1-\cos\theta_2) \\ & +\tfrac{1}{2}k\left\{\left[(b - l\sin\theta_1 + l\sin\theta_2)^2 + (l\cos\theta_2 - l\cos\theta_1)^2\right]^{1/2} - b\right\}^2, \end{aligned}$$

where b is the natural length of the spring, which is also the separation of the two pivots. Since the spring has restoring force $k \times$ extension, its potential energy is $\frac{1}{2}k \times \text{extension}^2$, hence the rather complicated second line, involving the sum of the squares of the horizontal and vertical displacements. Then, in the same way as before, the extension of the spring is $l\theta_2 - l\theta_1$, to sufficient approximation (being the difference between the horizontal displacements). Thus for small amplitude, correct to leading order, which for the energies is *quadratic* order. the total potential energy is approximately

$$V = \tfrac{1}{2}m_1gl\theta_1^2 + \tfrac{1}{2}m_2gl\theta_2^2 + \tfrac{1}{2}kl^2(\theta_2 - \theta_1)^2.$$

In summary, therefore, the Lagrangian function is

$$\mathcal{L}(\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2) = \tfrac{1}{2}m_1 l^2 \dot{\theta}_1^2 + \tfrac{1}{2}m_2 l^2 \dot{\theta}_2^2 - \tfrac{1}{2}m_1 g l \theta_1^2 - \tfrac{1}{2}m_2 g l \theta_2^2 - \tfrac{1}{2}kl^2(\theta_2 - \theta_1)^2 \ ,$$

and the equations of motion are

$$m_1 l\ddot{\theta}_1 = -m_1 g\theta_1 - kl(\theta_1 - \theta_2) \tag{1.5}$$
$$m_2 l\ddot{\theta}_2 = -m_2 g\theta_2 + kl(\theta_1 - \theta_2) \tag{1.6}$$

or, in matrix form,

$$\begin{pmatrix} m_1 l & 0 \\ 0 & m_2 l \end{pmatrix} \begin{pmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{pmatrix} = \begin{pmatrix} -m_1 g - kl & kl \\ kl & -m_2 g - kl \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \ . \tag{1.7}$$

This matrix equation can be rewritten as

$$\mathbf{T}\ddot{\mathbf{x}} = -\mathbf{V}\mathbf{x}$$

where the column vector $\mathbf{x} = (\theta_1,\, \theta_2)^{\mathsf{T}}$ and where the matrices

$$\mathbf{T} = \begin{pmatrix} m_1 l^2 & 0 \\ 0 & m_2 l^2 \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} -m_1 g l - kl^2 & kl^2 \\ kl^2 & -m_2 g l - kl^2 \end{pmatrix}, \tag{1.8}$$

have been defined in a standard way, such that the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}T_{ij}\dot{\theta}_i\dot{\theta}_j - \tfrac{1}{2}V_{ij}\theta_i\theta_j \tag{1.9}$$

to quadratic order (summation convention, i, j each running from 1 to 2).

There are various ways of solving these equations. The most elementary is to use (1.5) to eliminate θ_2 from (1.6). This results in a fourth order equation with constant coefficients for θ_1, the general solution of which will be a sum of four complex exponential terms.

(As usual it will be convenient to use complex exponentials in place of sines and cosines. After solving the equations we can take the real part, which also satisfies the equations because they are linear equations with real coefficients.)

Since we know from the above reasoning that the solution will contain only exponentials, we may as well try immediately a solution of the form

$$\theta_1 = a_1 e^{i\omega t}, \qquad \theta_2 = a_2 e^{i\omega t} \ ,$$

where a_1 and a_2 are constants. By definition this will be a normal mode, because the coordinates will oscillate sinusoidally with the same frequency. For such a solution we can replace each time derivative in equation (1.7) with $i\omega$ and cancel the exponentials, giving

$$-\omega^2 \begin{pmatrix} m_1 l & 0 \\ 0 & m_2 l \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} -m_1 g - kl & kl \\ kl & -m_2 g - kl \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{1.10}$$

i.e.

$$\begin{pmatrix} \omega^2 m_1 l - m_1 g - kl & kl \\ kl & \omega^2 m_2 l - m_2 g - kl \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = 0 \ . \tag{1.11}$$

This equation has nontrivial solutions only if ω^2 takes values satisfying

$$\begin{vmatrix} \omega^2 m_1 l - m_1 g - kl & kl \\ kl & \omega^2 m_2 l - m_2 g - kl \end{vmatrix} \ = \ 0 \ . \tag{1.12}$$

Calculating the determinant gives a quadratic equation in ω^2,

$$(\omega^2 m_1 l - m_1 g - kl)(\omega^2 m_2 l - m_2 g - kl) - (kl)^2 = 0 \ ,$$

i.e.

$$m_1 m_2 (\omega^2 l - g)^2 - (m_1 + m_2)(\omega^2 l - g) kl = 0 \ ,$$

which has solutions

$$\omega_1^2 \ = \ \frac{g}{l} \qquad \text{and} \qquad \omega_2^2 \ = \ \frac{g}{l} \ + \ \frac{k(m_1 + m_2)}{m_1 m_2}$$

generalizing those on page 6. Corresponding to each of these solutions, there is a pair of amplitudes

$$(a_1^{(1)}, a_2^{(1)}) \quad \text{and} \quad (a_1^{(2)}, a_2^{(2)})$$

that can be determined up to an overall scale factor from equation (1.11). Thus for ω_1 we have

$$\begin{pmatrix} -kl & kl \\ kl & -kl \end{pmatrix} \begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \end{pmatrix} = 0 \ , \tag{1.13}$$

so that $a_1^{(1)} = a_2^{(1)}$. This corresponds to the two masses swinging in phase. For ω_2, we have

$$\begin{pmatrix} (m_1/m_2)kl & kl \\ kl & (m_2/m_1)kl \end{pmatrix} \begin{pmatrix} a_1^{(2)} \\ a_2^{(2)} \end{pmatrix} = 0 \ , \tag{1.14}$$

so that $m_1 a_1^{(2)} = -m_2 a_2^{(2)}$. This corresponds to the two masses swinging 180° out of phase, with scaled amplitudes. By picturing the situation and remembering Newton's third law — the spring exerts equal and opposite *forces* — one can see that the *accelerations* and therefore the amplitudes must scale in just this way. (It's easiest to consider first the limit $g/l \to 0$.)

As in the case of equal masses, we can find the normal coordinates — the special linear combinations of θ_1 and θ_2 that oscillate simple harmonically, with a single, pure frequency, in this case either ω_1 or ω_2. This can be done by taking linear combinations of equations (1.5) with (1.6). Adding them gives

$$l(m_1\ddot{\theta}_1 + m_2\ddot{\theta}_2) = -g(m_1\theta_1 + m_2\theta_2)$$

whereas dividing the first by m_1, the second by m_2, and subtracting gives

$$l(\ddot{\theta}_1 - \ddot{\theta}_2) = -\big(g + kl(m_1^{-1} + m_2^{-1})\big)(\theta_1 - \theta_2) \ .$$

So $(m_1\theta_1 + m_2\theta_2)$, $(\theta_1 - \theta_2)$ are the normal coordinates for this problem. ⌟

Remark: In more complicated small-oscillations problems, it is worth paying attention to symmetry properties. There are many important problems with symmetries; that is why *group theory* is so important in physics and chemistry. The usefulness of symmetry considerations will be illustrated in a worked example below (equilateral triangle problem, Example 1.5) and at the end of the course.

The pendulum problem just analysed has an obvious reflectional symmetry in the case $m_1 = m_2$. In that respect it is like the CO_2 molecule (Example 1.4 below). The equations of motion express the symmetry in that they are unaltered by the transformation $\theta_1 \leftrightarrow -\theta_2$, when $m_1 = m_2$. (This is shorthand for making the substitutions $\theta_1 = -\theta_2^{\rm new}$, $\theta_2 = -\theta_1^{\rm new}$ and then dropping the superscripts.) Notice that it is the second of the normal modes found above that's invariant under this particular transformation. That is, any motion described by the second normal mode alone looks exactly the

same when reflected about a vertical plane halfway between the two pivots. Typically, for each symmetry of the system, there will be at least one invariant mode.

1.2 General analysis of small oscillations

1.2.1 Lagrange's equations

We consider a system with N degrees of freedom, i.e. one that can be described by means of N generalized coordinates

$$q_1,\ q_2,\ \ldots,\ q_N.$$

We may refer to these coordinates as the vector $\mathbf{q}$, meaning column vector as before.

Let $V(\mathbf{q})$ be the potential energy of the system and assume that the coordinates have been chosen so that $\mathbf{q} = \mathbf{0}$ is a position of stable equilibrium. Expanding $V(\mathbf{q})$ by Taylor series about the origin for small $\mathbf{q}$ gives (using the summation convention and writing $\partial^2 V/\partial q_i \partial q_j|_{\mathbf{q}=\mathbf{0}} = V_{ij}$)

$$V(\mathbf{q}) = V(\mathbf{0}) + \tfrac{1}{2} V_{ij} q_i q_j + \cdots$$

where V_{ij} are the components of a constant, symmetric, semi-positive matrix $\mathbf{V}$. Once again there are no linear terms, since $\partial V/\partial q_i = 0$ at $\mathbf{q} = \mathbf{0}$, this being the condition for equilibrium. We will take $V(\mathbf{0}) = 0$, without loss of generality, since this constant plays no part in determining the motion.

Note: the fact that $\mathbf{V}$ is semi-positive or non-negative, i.e. $V_{ij} q_i q_j \geqslant 0$ for all $\mathbf{q}$, implies that all the forces derived from the potential are either restoring forces or zero, i.e. that the equilibrium is semi-stable.

Let T be the kinetic energy of the system, which will generally depend on $\mathbf{q}$ and $\dot{\mathbf{q}}$. To second order, however, we assume that it can be written in the form

$$T = \tfrac{1}{2} T_{ij} \dot{q}_i \dot{q}_j,$$

where T_{ij} are the components of a *constant*, symmetric positive-definite matrix $\mathbf{T}$.

(Unlike $\mathbf{V}$, $\mathbf{T}$ is taken to be positive definite. That is, $T_{ij} \dot{q}_i \dot{q}_j > 0$ for all nonzero $\dot{\mathbf{q}}$. Thus, all modes of oscillation contribute to the kinetic energy

at lowest order. If $\mathbf{q}$ has dimensions of length, then $\mathbf{T}$ has dimensions of mass; if $\mathbf{q}$ is dimensionless (e.g. angles), then $\mathbf{T}$ has dimensions of moment of inertia (mass $\times$ length^2). If the q_i have a variety of dimensions — e.g. lengths, angles, areas — then the matrix elements T_{ij} have a corresponding variety of dimensions and similarly for the V_{ij}.)

The Lagrangian for the system is, to first nontrivial order (i.e. quadratic),

$$\mathcal{L} = T - V = \tfrac{1}{2} T_{ij} \dot{q}_i \dot{q}_j - \tfrac{1}{2} V_{ij} q_i q_j \ .$$

Lagrange's equations are

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i}\right) - \frac{\partial \mathcal{L}}{\partial q_i} = 0 \qquad (i = 1, ...N) \ .$$

Therefore the equations of motion (equivalent to Newton's laws) are the following system of N coupled second order linear equations:

$$T_{ij} \ddot{q}_j + V_{ij} q_j = 0 \, . \tag{1.15}$$

1.2.2 Normal modes

The idea of normal modes is the same as before. The normal modes are those special solutions of (1.15) that oscillate with a single, pure frequency. To find them try complex exponentials or, equally well,

$$q_i(t) = Q_i \sin \omega (t - t_0) \ ,$$

where $\mathbf{Q}$ is independent of t. Substitution into the equations of motion gives

$$-\omega^2 T_{ij} Q_j + V_{ij} Q_j = 0 \tag{1.16}$$

or in matrix notation, with $\mathbf{Q} = \{Q_1,\ Q_2,\ \ldots\ ,\ Q_N\}^{\mathsf{T}}$,

$$(-\omega^2 \mathbf{T} + \mathbf{V})\mathbf{Q} = \mathbf{0} \ . \tag{1.17}$$

Since we are looking for a nonzero solution $\mathbf{Q}$, linear algebra tells us that

$$\det(-\omega^2 \mathbf{T} + \mathbf{V}) = 0 \ . \tag{1.18}$$

This is a polynomial of degree N in ω^2. The solutions are the normal frequencies squared. *The normal frequencies must be real* (their squares non-negative), since we can multiply (1.16) by Q_i and sum to give

$$\omega^2 = \frac{V_{ij}Q_iQ_j}{T_{ij}Q_iQ_j} \ ,$$

This expression[2] is non-negative under our assumptions about $\mathbf{T}$ and $\mathbf{V}$.

In the special case $\mathbf{T} = \mathbf{I}$, the solutions for ω^2 are simply the eigenvalues of the symmetric matrix $\mathbf{V}$. So in the general case where $\mathbf{T} \neq \mathbf{I}$ we'll call them the **generalized eigenvalues** of the problem (1.17), and the corresponding vectors $\mathbf{Q}$ the **generalized eigenvectors**.[3]

Let the normal frequencies, defined as the positive square roots of the generalized eigenvalues, be

$$\omega_1, \ \omega_2, \ \omega_3, \ \ldots, \ \omega_N \ ,$$

and let the corresponding generalized eigenvectors be

$$\mathbf{Q}^{(1)}, \ \mathbf{Q}^{(2)}, \ \ldots, \ \mathbf{Q}^{(N)}$$

each with i^{th} component

$$Q_i^{(1)}, \ Q_i^{(2)}, \ \ldots, \ Q_i^{(N)}$$

$(i = 1, \ldots, N)$ so that the generalized eigenvalue for the m^{th} normal mode satisfies

[2]The expression is known as Rayleigh's formula, or Rayleigh's quotient: for $\omega \neq 0$ it has the useful 'stationarity property' of being $O(\epsilon^2)$ accurate when $\mathbf{Q}$ is $O(\epsilon)$ accurate. You don't need to remember any of that for the examination. Just for fun, though, consider the effect of replacing the true $\mathbf{Q}$ for a normal mode by slightly inaccurate values $\mathbf{Q} + \delta\mathbf{Q}$, with small δQ_i all $O(\epsilon)$ as $\epsilon \to 0$. Rewrite Rayleigh's formula as $\omega^2 T_{ij}Q_jQ_i - V_{ij}Q_jQ_i = 0$, using the symmetry of T_{ij} and V_{ij}. Small changes satisfy $2\omega\delta\omega T_{ij}Q_jQ_i + 2\left(\omega^2 T_{ij}Q_j - V_{ij}Q_j\right)\delta Q_i = O(\epsilon^2)$, again using symmetry. But (1.16) shows that the expression in brackets is zero and hence that $\delta\omega = O(\epsilon^2)$. This is related to a much more general principle of classical and quantum mechanics, the *principle of least action* or stationarity of the so-called 'action integral' $\int \mathcal{L}\, dt$, and of quantum-mechanical 'path integrals'. See Chapter 19 of Feynman's great *Lectures in Physics*, vol. II.

[3]We can always reduce the problem to the case $\mathbf{T} = \mathbf{I}$, by making a rotation of coordinates that diagonalizes the symmetric matrix $\mathbf{T}$, and then rescaling the new coordinates so that the diagonal entries in the new $\mathbf{T}$ are all 1. If a further rotation (which does not affect $\mathbf{T}$) is now made to diagonalize $\mathbf{V}$, then the resulting new coordinates are normal coordinates, as can easily be seen from the equations of motion.

$$\omega_m^2 \;=\; \frac{V_{ij}Q_i^{(m)}Q_j^{(m)}}{T_{ij}Q_i^{(m)}Q_j^{(m)}} \tag{1.19}$$

as already noted. Then we can write the general solution in the form

$$\mathbf{q}(t) = \sum_{m=1}^{N} A^{(m)}\mathbf{Q}^{(m)} \sin\omega_m(t - t_0^{(m)}) \;, \tag{1.20}$$

where the constant $A^{(m)}$ is the amplitude of the m^{th} normal mode, and the constant $\omega_m t_0^{(m)}$ its phase. This assumes that none of the generalized eigenvalues ω_m^2 are zero. Once we have obtained the frequencies ω_m and generalized eigenvectors $\mathbf{Q}^{(m)}$, our expression for the general solution thus depends on the $2N$ unknowns $A^{(m)}$ and $t_0^{(m)}$,

For so-called **zero modes**, for which $\omega_m^2 = 0$ by definition, the sine function in the term $A^{(m)}\sin\omega_m(t-t_0^{(m)})$ must be replaced by a linear function $B^{(m)}(t - t_0^{(m)})$. This can be shown by taking the limit $\omega_m \to 0$, with $B^{(m)}$ defined as $A^{(m)}\omega_m$ and held constant. We still have a total of $2N$ unknown constants, since for each zero mode we now have $B^{(m)}$ and $t_0^{(m)}$.

The solution (1.20) is not, of course, generally periodic, just as the function in figure 3 on page 5 is not periodic. Only if all the frequency ratios are rational do we have periodicity.

Remark on terminology: The polynomial equation (1.18) for the generalized eigenvalues ω_m^2 is sometimes called the 'characteristic equation'.

1.2.3 Orthogonality

Suppose that two of the generalized eigenvectors, say $\mathbf{Q}^{(1)}$, $\mathbf{Q}^{(2)}$, have distinct normal frequencies $\omega_1 \neq \omega_2$. Then $\mathbf{Q}^{(1)}$ is orthogonal to $\mathbf{Q}^{(2)}$ in a certain sense. From (1.17) we have

$$(-\omega_1^2\mathbf{T} + \mathbf{V})\mathbf{Q}^{(1)} = \mathbf{0}$$

and

$$(-\omega_2^2\mathbf{T} + \mathbf{V})\mathbf{Q}^{(2)} = \mathbf{0} \;.$$

If we premultiply these equations by the row vectors $-(\mathbf{Q}^{(2)})^{\mathsf{T}}$ and $(\mathbf{Q}^{(1)})^{\mathsf{T}}$ respectively and add the results, then the $\mathbf{V}$ terms cancel because of the symmetry of $\mathbf{V}$ (which implies that $Q_i^{(2)}V_{ij}Q_j^{(1)} = Q_i^{(1)}V_{ij}Q_j^{(2)}$). Similarly using the symmetry of $\mathbf{T}$ we get

$$(\omega_1^2 - \omega_2^2)\;(\mathbf{Q}^{(1)})^{\mathsf{T}}\mathbf{T}\mathbf{Q}^{(2)} = 0 \;.$$

Since $(\omega_1^2 - \omega_2^2) \neq 0$, we have $(\mathbf{Q}^{(1)})^{\mathsf{T}}\mathbf{T}\mathbf{Q}^{(2)} = 0$. We say that $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ are 'orthogonal with respect to $\mathbf{T}$'.

If $\omega_1^2 = \omega_2^2$ then the proof fails. But it can be shown from standard linear algebra that there exist linearly independent $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$. From these, an orthogonal pair can always be constructed ('Gram–Schmidt orthogonalization', see below). So by rescaling the $\mathbf{Q}^{(m)}$ such that $(\mathbf{Q}^{(m)})^{\mathsf{T}}\mathbf{T}\mathbf{Q}^{(m)} = 1$ for each m, we finally have a set of generalized eigenvectors $\mathbf{Q}^{(m)}$ that satisfy

$$(\mathbf{Q}^{(m)})^{\mathsf{T}}\mathbf{T}\mathbf{Q}^{(n)} = \delta_{mn} \tag{1.21}$$

where δ_{mn} is the Kronecker delta. We then say that the $\mathbf{Q}^{(m)}$ make up an **orthonormal** set of generalized eigenvectors.

These results are slight generalizations of the standard linear-algebra results about the eigenvalues and eigenvectors of symmetric matrices $\mathbf{V} = \mathbf{V}^{\mathsf{T}}$. Indeed, in the important special case $\mathbf{T} = \mathbf{I}$ they are exactly the standard results, applied to the matrix $\mathbf{V}$.

Just a reminder: first, the eigenvalues of a real symmetric matrix are real, and non-negative if the matrix is non-negative in the sense that its quadratic form is always $\geqslant 0$. Pairs of eigenvectors $\mathbf{e}^{(m)}$, $\mathbf{e}^{(n)}$ corresponding to distinct eigenvalues are orthogonal and can be rescaled to be orthonormal, $(\mathbf{e}^{(m)})^{\mathsf{T}}\mathbf{I}\,\mathbf{e}^{(n)} = (\mathbf{e}^{(m)})^{\mathsf{T}}\mathbf{e}^{(n)} = \delta_{mn}$.

Second, if there is a doubly repeated eigenvalue, ω_1^2 say, then there is a whole plane of eigenvectors spanned by a pair of linearly independent eigenvectors corresponding to ω_1^2, say $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$. This means that three numbers β_1, β_2, γ can be found such that $\beta_1\mathbf{e}^{(1)} + \beta_2\mathbf{e}^{(2)}$ and $\gamma\mathbf{e}^{(2)}$ are normalised and orthogonal to each other as well as to the eigenvectors for different frequencies. If there is a triply repeated eigenvalue then the same applies except that 'plane' becomes 'volume', and so on. The general procedure can be found in textbooks under the heading 'Gram–Schmidt orthogonalization'.

Third and finally, when $\mathbf{T} \neq \mathbf{I}$ then everything goes through with just one minor modification. Whenever **inner products** of the form $(\mathbf{e}^{(m)})^{\mathsf{T}}\mathbf{e}^{(n)}$ occur — also called **scalar products** and written as $\mathbf{e}^{(m)}\cdot\mathbf{e}^{(n)}$ or $(\mathbf{e}^{(m)}, \mathbf{e}^{(n)})$ — they are replaced by **generalized inner products** of the form $[\mathbf{e}^{(m)}, \mathbf{e}^{(n)}] = (\mathbf{e}^{(m)})^{\mathsf{T}}\mathbf{T}\,\mathbf{e}^{(n)}$ just as in (1.21). It all works perfectly because of the positive-definiteness of $\mathbf{T}$.

The idea of 'generalized inner products' will crop up again in the group-theoretic sections. There, $\mathbf{T}$ will often be Hermitian, $\mathbf{T} = \mathbf{T}^{\dagger}$, where † means the transposed complex conjugate, $\mathbf{T}^{\dagger} = \mathbf{T}^{*\mathsf{T}}$.

1.2.4 Normal coordinates

Normal coordinates $\alpha^{(m)}(t)$ are, as before, just those linear combinations of the original generalized coordinates $q_j(t)$ that oscillate with the single, pure frequency ω_m, and satisfy the same simple-harmonic equation as the simple

pendulum. As soon as we are in possession of an orthonormal set of generalized eigenvectors $\mathbf{Q}^{(m)}$ we can immediately find the normal coordinates. For if we multiply the j^{th} component of (1.20) by $Q_i^{(n)}T_{ij}$, sum over j, and use (1.21), we immediately obtain

$$Q_i^{(n)}T_{ij}q_j(t) \;=\; A^{(n)}\sin\omega_n(t-t_0^{(n)}) \;, \tag{1.22}$$

showing that the left-hand side is a normal coordinate for any choice of n. Let us call this n^{th} normal coordinate $\alpha^{(n)}(t)$; i.e.,

$$\alpha^{(n)}(t) = Q_i^{(n)}T_{ij}q_j(t) \;. \tag{1.23}$$

(We have assumed here there are no zero modes; if there are then the right hand side of (1.22) is replaced by a linear function for the relevant n.)

Alternatively, we can go through the whole derivation from first principles, as follows, underlining the fact that $\alpha^{(n)}(t)$ satisfies the simple-pendulum equation. Let

$$q_i(t) = \sum_{m=1}^{N}\alpha^{(m)}(t)Q_i^{(m)} \;,$$

which is possible because there are N linearly independent vectors $\mathbf{Q}_i^{(m)}$. Then substituting into the equation of motion (1.15) gives

$$\sum_{m=1}^{N}\left[\ddot{\alpha}^{(m)}(t)T_{ij}Q_j^{(m)} + \alpha^{(m)}(t)V_{ij}Q_j^{(m)}\right] = 0 \;,$$

which on using (1.16) becomes

$$\sum_{m=1}^{N}\left[\ddot{\alpha}^{(m)}(t) + \omega_m^2\alpha^{(m)}(t)\right]T_{ij}Q_j^{(m)} = 0.$$

Multiplying by $Q_i^{(n)}$ and using the orthonormality property (1.21) gives

$$\ddot{\alpha}^{(m)}(t) + \omega_m^2\alpha^{(m)}(t) = 0$$

with solution

$$\alpha^{(m)}(t) = A^{(m)}\sin\omega_m(t-t_0^{(m)}) \;,$$

echoing (1.22). That is, $\alpha^{(m)}(t)$ are the normal coordinates. They can be written in terms of $q_i(t)$ using orthonormality:

$$q_i(t) = \sum_{m=1}^{N} \alpha^{(m)}(t) Q_i^{(m)}$$

$$\begin{aligned} \Rightarrow \quad q_i(t) T_{ij} Q_j^{(n)} &= \sum_{m=1}^{N} \alpha^{(m)}(t) Q_i^{(m)} T_{ij} Q_j^{(n)} \\ &= \sum_{m=1}^{N} \alpha^{(m)}(t) \delta_{mn} \\ &= \alpha^{(n)}(t) \ , \end{aligned} \tag{1.24}$$

reproducing (1.23) apart from renaming the indices.

Administrative note: *The next two examples will be only sketched in the lectures.* I recommend you try to solve them unseen, then use the following notes as worked examples.

The first example is simple and straightforward.

The second example is trickier. The key is to use symmetry, intuitively, to guess most of the normal modes. Question 7 of the 2003 exam is about another example quite like it (copies from `www.maths.cam.ac.uk/undergrad/nst/pastpapers`).

Not lectured on: There follows a **non-examinable** section in small print on the 'adiabatic invariant' for small oscillations under conditions (e.g. constraints) that are changing slowly. This just for fun — a beautiful application of the stationarity property of the Rayleigh quotient (1.19).

Example 1.4 *Vibrations of the* CO_2 *molecule*

The problem is to find the normal frequencies and normal modes of the following system of masses and springs, a simple model of the carbon dioxide molecule. (The mass M of a real carbon atom is roughly three-quarters of the mass m of a real oxygen atom.) Attention is restricted to the in-line or 1-dimensional vibrations.

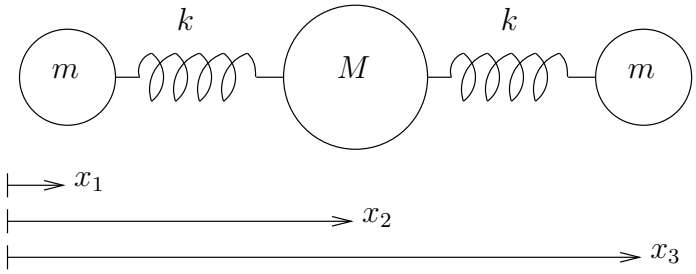


Fig. 4: A simple model of a CO_2 molecule.

Consider vibrations along the line of the molecule and let x_1, x_2 and x_3 be the displacements from equilibrium of the atoms. These are the generalized coordinates q_1, q_2 and q_3 that are most convenient here. Then

$$T = \tfrac{1}{2}m\dot{x}_1^2 + \tfrac{1}{2}M\dot{x}_2^2 + \tfrac{1}{2}m\dot{x}_3^2 \quad \Rightarrow \quad \mathbf{T} = \begin{pmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{pmatrix}$$

and

$$V = \tfrac{1}{2}k(x_2 - x_1)^2 + \tfrac{1}{2}k(x_2 - x_3)^2 \quad \Rightarrow \quad \mathbf{V} = k\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} .$$

To find the normal frequencies, we solve the characteristic equation (1.18), which in this case becomes

$$\det\left[-\omega^2\begin{pmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{pmatrix} + k\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}\right] = 0$$

i.e.

$$\begin{vmatrix} k - \omega^2 m & -k & 0 \\ -k & 2k - \omega^2 M & -k \\ 0 & -k & k - \omega^2 m \end{vmatrix} = 0$$

which, after a little algebra, reduces to

$$\omega^2(k - \omega^2 m)(2km + kM - \omega^2 mM) = 0 .$$

Thus the normal frequencies in ascending order are given by

$$\omega_1^2 = 0, \qquad \omega_2^2 = k/m, \qquad \omega_3^2 = k(2M^{-1} + m^{-1}) .$$

To find the normal modes for each normal frequency, we solve equation (1.17).
For ω_1, we have

$$\begin{pmatrix} k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & k \end{pmatrix}\begin{pmatrix} Q_1^{(1)} \\ Q_2^{(1)} \\ Q_3^{(1)} \end{pmatrix} = 0$$

so, up to an overall multiple,

$$\mathbf{Q}^{(1)} \propto \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

The overall multiple can be fixed by requiring

$$(\mathbf{Q}^{(1)})^{\mathbf{T}}\mathbf{T}\mathbf{Q}^{(1)} = 1 .$$

Thus the normalized normal mode has generalized eigenvector

$$\mathbf{Q}^{(1)} = (2m+M)^{-\frac{1}{2}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

This normal mode, a **zero mode**, is just a rigid translation of the system.

For ω_2, we have

$$\begin{pmatrix} 0 & -k & 0 \\ -k & 2k - kM/m & -k \\ 0 & -k & 0 \end{pmatrix} \begin{pmatrix} Q_1^{(2)} \\ Q_2^{(2)} \\ Q_3^{(2)} \end{pmatrix} = 0$$

so that the normalized normal mode has generalized eigenvector

$$\mathbf{Q}^{(2)} = (2m)^{-\frac{1}{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} .$$

This mode is invariant under the reflectional symmetry of the problem.

For ω_3, we have

$$\begin{pmatrix} -2km/M & -k & 0 \\ -k & -kM/m & -k \\ 0 & -k & -2km/M \end{pmatrix} \begin{pmatrix} Q_1^{(3)} \\ Q_2^{(3)} \\ Q_3^{(3)} \end{pmatrix} = 0$$

so the normalized normal mode has generalized eigenvector

$$\mathbf{Q}^{(3)} = (2m + 4m^2/M)^{-\frac{1}{2}} \begin{pmatrix} 1 \\ -2m/M \\ 1 \end{pmatrix} .$$

You should check that the three modes are pairwise orthogonal with respect to $\mathbf{T}$, as they should be because ω_1, ω_2 and ω_3 are all distinct.

To find the normal coordinates, we may use (1.23). Note that the first coordinate corresponds to a zero mode and thus gives us a linear function, corresponding to the fact that

the centre of mass moves with constant velocity. For $\alpha^{(1)}(t)$ we have

$$\begin{aligned}\alpha^{(1)}(t) &= \mathbf{x}^{\mathbf{T}}\mathbf{T}\mathbf{Q}^{(1)} \\ &= \begin{pmatrix} x_1, & x_2, & x_3 \end{pmatrix} \begin{pmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{pmatrix} (2m+M)^{-1/2} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \frac{mx_1 + mx_3 + Mx_2}{\sqrt{(2m+M)}} = B_1(t - t_0^{(1)}) \ .\end{aligned}$$

Similarly,

$$\alpha^{(2)}(t) = \sqrt{(m/2)}\ (x_1 - x_3) = A_2 \sin\omega_2(t - t_0^{(2)})$$

and

$$\alpha^{(3)}(t) = \frac{x_1 - 2x_2 + x_3}{\sqrt{(2/m + 4/M)}} = A_3 \sin\omega_3(t - t_0^{(3)})$$

⌟

Example 1.5 *Normal modes of a 3-particle-spring system in* 2 *dimensions*

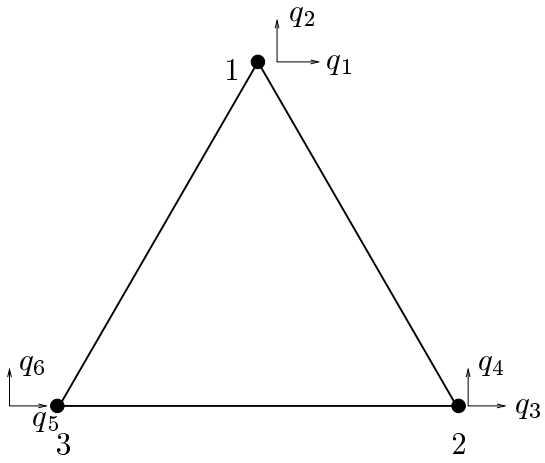


Fig. 5: Coordinates for the 3 particle-spring system.

Three particles each of mass m, resting on a frictionless level plane surface, are interconnected by identical springs of length l and spring constant k. The equilibrium position is an equilateral triangle. Let $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ be the cartesian coordinates of the three particles in the general motion.

(We could try to produce a neater system of equations by using more symmetrical coordinates; polar coordinates for each particle, for example. Generally, unless there is a really obvious set of coordinates, it is easier to bash through using cartesians.)

For the general motion, let

$$\begin{aligned} \mathbf{x}_1 &= l(0, 1/\sqrt{3}) + (q_1, q_2) \\ \mathbf{x}_2 &= l(1/2, -1/2\sqrt{3}) + (q_3, q_4) \\ \mathbf{x}_3 &= l(-1/2, -1/2\sqrt{3}) + (q_5, q_6) \end{aligned}$$

so that the origin is at the centre of the triangle and the generalized coordinates q_1, …, q_6 are zero at the equilibrium position.

The kinetic energy of the system is

$$\begin{aligned} T &= \tfrac{1}{2}m(\dot{\mathbf{x}}_1.\dot{\mathbf{x}}_1 + \dot{\mathbf{x}}_2.\dot{\mathbf{x}}_2 + \dot{\mathbf{x}}_3.\dot{\mathbf{x}}_3) \\ &= \tfrac{1}{2}m(\dot{q}_1^2 + \dot{q}_2^2 + \dot{q}_3^2 + \dot{q}_4^2 + \dot{q}_5^2 + \dot{q}_6^2) \end{aligned}$$

so

$$T_{ij} = m\delta_{ij} \quad \text{or} \quad \mathbf{T} = m \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = m\mathbf{I}. \tag{1.25}$$

Thus ω^2 will be an eigenvalue of $m^{-1}\mathbf{V}$.

The potential energy is a little harder to calculate. Recall that the potential energy for a spring between two points which differ by $\mathbf{x}$ is $V(\mathbf{x}) = \frac{1}{2}k(|\mathbf{x}| - l)^2$. We need to calculate the quadratic approximation to

$$\tfrac{1}{2}k\Big[|\mathbf{x}_1 - \mathbf{x}_2| - l\Big]^2 + \text{ etc.}$$

Now

$$\mathbf{x}_1 - \mathbf{x}_2 \;=\; l(-1/2, \sqrt{3}/2) \;+\; (q_1 - q_3, q_2 - q_4)\;,$$

which is of the form $\mathbf{a} + \mathbf{v}$ where $|\mathbf{a}| = l \gg |\mathbf{v}|$, so that the length is

$$\big|\mathbf{a} + \mathbf{v}\big| \;=\; (\mathbf{a}\cdot\mathbf{a} + 2\,\mathbf{a}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{v})^{1/2} \;=\; l(1 + \mathbf{a}\cdot\mathbf{v}/l^2 + \cdots)$$

and the contribution to the potential energy is approximately

$$\tfrac{1}{2}k\Big[\mathbf{a}\cdot\mathbf{v}/l\Big]^2 \;=\; \tfrac{1}{2}k\Big[(-\tfrac{1}{2},\, \tfrac{\sqrt{3}}{2})\cdot(q_1 - q_3,\, q_2 - q_4)\Big]^2.$$

Doing the algebra gives

$$\begin{aligned} V \;\;=\;\; & \tfrac{1}{2}k\Big[-\tfrac{1}{2}(q_1 - q_3) + \tfrac{\sqrt{3}}{2}(q_2 - q_4)\Big]^2 \\ & + \tfrac{1}{2}k\Big[q_3 - q_5\Big]^2 \\ & + \tfrac{1}{2}k\Big[-\tfrac{1}{2}(q_5 - q_1) - \tfrac{\sqrt{3}}{2}(q_6 - q_2)\Big]^2 \end{aligned}$$

From this since $V = \frac{1}{2}V_{ij}q_iq_j$, we can read off the elements of $\mathbf{V}$:

$$\mathbf{V} = \frac{k}{4}\begin{pmatrix} 2 & 0 & -1 & \sqrt{3} & -1 & -\sqrt{3} \\ 0 & 6 & \sqrt{3} & -3 & -\sqrt{3} & -3 \\ -1 & \sqrt{3} & 5 & -\sqrt{3} & -4 & 0 \\ \sqrt{3} & -3 & -\sqrt{3} & 3 & 0 & 0 \\ -1 & -\sqrt{3} & -4 & 0 & 5 & \sqrt{3} \\ -\sqrt{3} & -3 & 0 & 0 & \sqrt{3} & 3 \end{pmatrix} \tag{1.26}$$

We need to find the eigenvalues of this rather unpleasant matrix. The characteristic equation

$$\det(\mathbf{V} - \omega^2\mathbf{T}) = 0$$

will be a polynomial of degree 6 in $m\omega^2/k$. We could try to simplify this, but unless something particularly nice happens we would not usually expect to extract the roots without the help of a computer.

Now comes the point of this example: because of the symmetry of the physical system, something particularly nice does happen, which allows us to obtain the normal modes and frequencies. With very little algebra, but some careful thought, we can in fact *guess* all the six normal modes vectors.

First note that there are three independent motions of the system that do not involve stretching the springs at all: the system moves as a rigid body. These are translations (linear motion at constant speed) in the x_1 direction and in the x_2 direction, and a rotation at constant angular speed about the origin.

For the translations in the x_1 direction, the x_2 coordinates of the particles remain the same so $q_2 = q_4 = q_6 = 0$ while the displacements of the particles in the x_1 direction are the same, so $q_1 = q_3 = q_5$. We therefore expect two normal mode vectors of the form

$$\mathbf{Q}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{Q}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \tag{1.27}$$

We can verify that these do indeed satisfy

$$\mathbf{VQ} = \omega^2 \mathbf{TQ}, \tag{1.28}$$

with $\omega = 0$ as expected.

For the rotation, we see, using a bit of geometry, that the particles should set off in the directions $(1, 0)$, $(-\cos 60^\circ, -\sin 60^\circ)$, $(-\cos 60^\circ, \sin 60^\circ)$, i.e. to

$$\mathbf{Q}^{(3)} = \begin{pmatrix} 1 \\ 0 \\ -1/2 \\ -\sqrt{3}/2 \\ -1/2 \\ \sqrt{3}/2 \end{pmatrix}. \tag{1.29}$$

This again corresponds to zero frequency. And again we see that $\mathbf{VQ} = 0$.

By symmetry, we expect there to be a pure expansion, a mode of oscillation in which each of the particles vibrates only along a radius vector. This gives a normal mode vector of the form

$$\mathbf{Q}^{(4)} = \begin{pmatrix} 0 \\ 1 \\ \sqrt{3}/2 \\ -1/2 \\ -\sqrt{3}/2 \\ -1/2 \end{pmatrix}. \tag{1.30}$$

Substitution into the normal mode equation (1.28) reveals that the corresponding normal frequency is given by $\omega^2 = 3k/m$.

By symmetry, again, we expect to find three modes with reflection symmetries about each of the 'heights' of the triangle. And again by symmetry these three modes will evidently have the same normal frequency. They cannot be independent, because we are only allowed a total of 6 normal modes and we have already found 4. The fact that they have the same frequency means that any linear combination will also be a normal mode, and it turns out that the three modes sum to zero. To find the mode vectors, try

$$\mathbf{Q}^{(5)} = \begin{pmatrix} 0 \\ -1 \\ \beta \\ \gamma \\ -\beta \\ \gamma \end{pmatrix} ,$$

which has reflection symmetry about the vertical. We could now find β and γ by substituting into (1.28), which would also determine ω. However, we can instead use the fact the mode vectors are orthogonal ($\mathbf{T}$ being a multiple of the unit matrix). Taking the scalar product with $\mathbf{Q}^{(2)}$ and $\mathbf{Q}^{(4)}$ gives $\gamma = 1/2$ and $\beta = \sqrt{3}/2$. Then substitution into (1.28) shows that the corresponding frequency is given by $\omega^2 = 3k/2m$.

(**Alternatively**: recall that the eigenvalues of a matrix sum to the trace of the matrix (i.e. the sum of its diagonal elements). Since the trace of $\mathbf{V}$ is $6k$ and four of its eigenvalues are 0, 0, 0 and $3k$, we see that the other two sum to $3k$. Since they are known, by symmetry, to be equal, they must each be $3k/2$.)

For the last mode vector, we can use a vector with symmetry about one of the other altitudes. Alternatively we find an orthogonal vector by trial and error. The last two vectors are then

$$\mathbf{Q}^{(5)} = \begin{pmatrix} 0 \\ -1 \\ \sqrt{3}/2 \\ 1/2 \\ -\sqrt{3}/2 \\ 1/2 \end{pmatrix} \quad \text{and} \quad \mathbf{Q}^{(6)} = \begin{pmatrix} 1 \\ 0 \\ -1/2 \\ \sqrt{3}/2 \\ -1/2 \\ -\sqrt{3}/2 \end{pmatrix} . \tag{1.31}$$

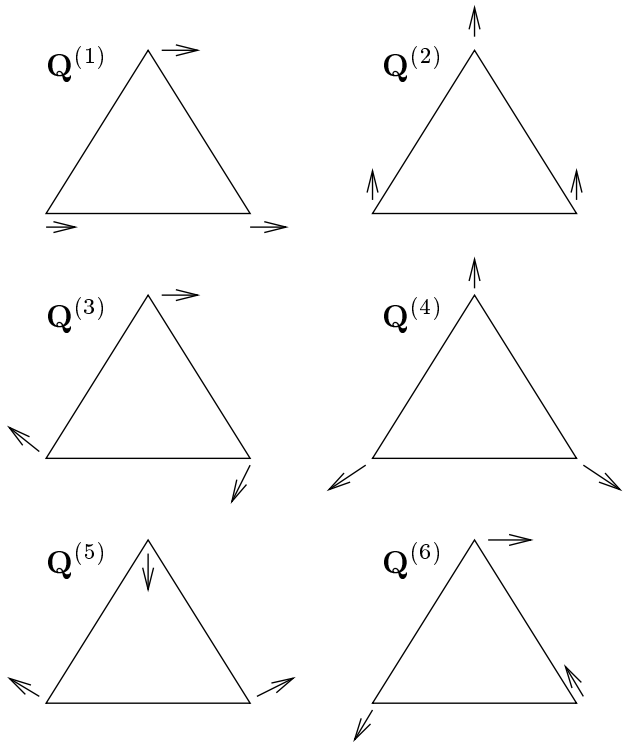


Fig. 6: Normal mode vectors.

Adiabatic invariants (non-examinable)

An extension of small-oscillations theory leads to the idea of *adiabatic invariance*, of great importance in physics and in theoretical chemistry. First we note the energy-conservation theorem for small oscillations. By multiplying Lagrange's equations in the form (1.15) by $\dot{q}_i$ and summing from 1 to N, using $T_{ij} = T_{ji}$ and $V_{ij} = V_{ji}$, we easily find $dE/dt = 0$ where E is the total energy of the oscillatory motion, $E = T + V$. For a normal mode $q_i(t) = Q_i \sin\omega(t - t_0)$ (Q_i real), we have

$$E = 2\overline{T} = 2\overline{V} = \tfrac{1}{2}\omega^2 T_{ij} Q_i Q_j = \tfrac{1}{2} V_{ij} Q_i Q_j \ ,$$

where the overbars denote time averaging. Notice the consistency with Rayleigh's formula (1.19) on page 14, and recall that $\overline{\cos^2\omega(t-t_0)} \quad = \overline{\sin^2\omega(t-t_0)} = \tfrac{1}{2}$.

Now adiabatic invariants arise when we generalize the theory to cases in which the coefficients T_{ij} and V_{ij} are no longer constant, but are allowed to be functions of time that vary slowly relative to the timescale ω^{-1}. (The word 'adiabatic' is often used in connection with such slow variation, by extension from its original meaning in thermodynamics.) The classic archetype is a simple pendulum whose length is made to change slowly, as with a mass on a string slowly pulled through a small hole in the ceiling. Then ω will increase

slowly, and so will E. (Part of the work done to pull the string goes into increasing E.) But it is approximately true that

$$\frac{d}{dt}\left(\frac{E}{\omega}\right) \;=\; 0\;. \tag{1.32}$$

E/ω is the adiabatic invariant, and is approximately constant. It changes much less than E and ω individually.

This is a general property of each normal mode of a vibrating system, as we now argue. It is easiest to demonstrate using complex exponentials (taking us a step closer to the quantum ramifications; and notice that the dimensions of E/ω are the same as those of Planck's constant.) Assume that the motion takes the form of a slowly varying normal mode that depends on a phase parameter α,

$$q_i = q_i(t,\alpha) = \mathrm{Re}\, Q_i(t)\, e^{i\phi(t)\,+\,i\alpha}\;, \tag{1.33}$$

where a real-valued function $\phi(t)$ has been introduced to represent the slowly-varying frequency $\omega(t) = d\phi(t)/dt$ and where the $Q_i(t)$ are slowly-varying functions of t, now complex-valued. The symbol Re means 'take the real part' (necessary in the classical problem though not in its quantum counterpart). The essential point is the *invariance of the oscillation problem with respect to* α. That is, α does not appear in the Lagrangian function $\mathcal{L}$ until the particular solution (1.33) is substituted into it. So if we have a solution for one value of α, then we can pick any other value of α and still have a solution.

This seemingly trivial remark leads to a neat little bit of mathematical magic. We multiply Lagrange's equations not by $\dot{q}_i$ but by $\partial q_i(t,\alpha)/\partial\alpha = q_{i,\alpha}$, say, remembering that the first term of (1.15) now needs to be rewritten as $d[T_{ij}(t)\dot{q}_j]/dt$, with the d/dt staying outside as dictated by Lagrange's equations. Then we average not with respect to t but with respect to α:

$$\frac{d}{dt}\overline{(q_{i,\alpha}T_{ij}(t)\dot{q}_j)}^{\alpha} = \overline{(\dot{q}_{i,\alpha}T_{ij}(t)\dot{q}_j)}^{\alpha} - \overline{(q_{i,\alpha}V_{ij}(t)q_j)}^{\alpha} = \tfrac{1}{2}\partial\,\overline{(\mathcal{L})}^{\alpha}/\partial\alpha = 0, \tag{1.34}$$

since $\overline{(\mathcal{L})}^{\alpha}$ is plainly independent of α. So $\overline{(q_{i,\alpha}T_{ij}(t)\dot{q}_j)}^{\alpha}$ is an *exact* constant of the motion, an important fact in its own right. We have not yet had to use the slowly-varying assumption! We finally use it twice, first to approximate $q_{i,\alpha}$ by $\dot{q}_i/\omega$, and second to equate α-averaging with t-averaging. Then the constancy of $\overline{(q_{i,\alpha}T_{ij}(t)\dot{q}_j)}^{\alpha}$ implies (1.32), the constancy of E/ω.

These remarks are not a rigorous proof of adiabatic invariance, but they are suggestive.

Chapter 2

2. Groups

This chapter contains the second set of syllabus items, together with a few items anticipating the third set:

Idea of an algebra of symmetry operations; symmetry operations on a square. Idea of conjugacy classes.

Definition of a group; group table.

Subgroups; homomorphic and isomorphic groups (including normal subgroups and cosets).

2.1 Symmetry properties and operations

Preamble (non-examinable)

Science works by fitting models to data, and *symmetry properties* powerfully constrain the model-building. It hardly needs saying that a good model of something in the real world, e.g. an electron or a CO_2 or CH_4 molecule, will have the same symmetry properties as the real thing.

It is the simplicity of symmetry properties — in many cases where models have been shown to fit data accurately — that accounts for what Eugene Wigner famously called 'the unreasonable effectiveness of mathematics' in science. The real world contains entities with simple symmetry properties. It need not be so, but according to the evidence it *is* so. Group theory is important because it gives us the tools to analyse, and classify, all possible symmetry properties.[1] Its usefulness extends beyond chemistry and physics, for

[1] Classic examples include not only fundamental particles and molecular vibrations but also the great discoveries via X-ray crystallography, from the foundations laid by W. and L. Bragg, Kathleen Lonsdale and others all the way to the discovery of DNA and protein

example to issues in probability theory, statistical inference and risk assessment.[2]

Many symmetries of physical systems are usefully thought of, to begin with, in terms of the invariance properties of simple geometrical objects: square, triangle, sphere, cylinder, hexagon, etc. For instance the shape of a square is invariant under in-plane rotation through 90°, and the shape of an equilateral triangle is invariant under in-plane rotation through $2\pi/3$ radian or 120°, a fact already made use of in the example of fig. 6. Rotation may therefore reasonably be called a *symmetry operation*. Another symmetry operation is that associated with mirror symmetry, reflection in some line or plane. *Groups are sets of mathematical entities — any mathematical entities — that behave like symmetry operations.*

(Each symmetry operation — 'symmetry' for short — involves labels, or parameters, such as the angle rotated through, and the axis of rotation. A parameter may take a continuous or a discrete range of values. The range is continuous in cases like rotation through any angle. It is discrete in cases like rotation through an integer number of right angles, or translation by an integer number of interatomic spacings in a crystal. This course will deal mainly with discrete cases.)

Symmetry operations can be combined. For instance one can rotate something and then reflect it. Several notations are in use for 'symmetry operation S_1 followed by symmetry operation S_2', e.g.,

$$S_2 \circ S_1 \qquad \text{or} \qquad S_2 * S_1 \qquad \text{or just} \qquad S_2 S_1 \ .$$

We'll use the last for simplicity. This 'composition' or 'product', so called, will also be a symmetry operation. In abstract mathematical language we say that symmetries form an **algebra**. An algebra is any set of mathematical entities equipped with a rule for combining or composing two entities, to produce a third entity that is also a member of the set.

---

structures. (There's a nice DNA diffraction tutorial in Lucas, A. A., Lambin, P., Mairesse, R., Mathot, M., 1999: Revealing the Backbone Structure of B-DNA from Laser Optical Simulations of Its X-ray Diffraction Diagram, *J. Chem. Educ.* **76**, 378–383.) For instance group theory tells us that the number of different rotation–reflection symmetry structures possible for regular crystal lattices is exactly 32, and of rotation–reflection–translation symmetry structures exactly 230: details in Leech & Newman. These symmetry structures are precisely described by what are called the crystallographic 'point groups' and 'space groups' respectively. *Point group* simply means that translations aren't included: there's always a point that stays fixed.

[2]For the relevance of group theory to these tricky topics see pp. 373–382 of Edwin T. Jaynes' book *Probability Theory: The Logic of Science* (CUP 2003).

Note the order in which the operators S_1 and S_2 are written. We are imagining that the symmetries operate on something to the right of them, so that the rightmost symmetry operates first. (Watch out though: the opposite convention is also used.)

There are three further points to be made about symmetry operations. First, there is always a trivial operation that does nothing; we call it the **identity** operation and denote it here[3] by I. (Trivially, everything is invariant under I.) Second, given any symmetry operation S there is always a reverse or **inverse** operation, denoted S^{-1}, that undoes the effect of S. In symbols,

$$S^{-1}S \;=\; I \;.$$

It is obvious that the inverse will always exist. After changing the position and orientation of any rigid geometrical object in any way, one can always move it back to its original position and orientation. By performing the first move again we see moreover that

$$SS^{-1} \;=\; I \;.$$

Third and finally, we always have **associativity**, which says that

$$S_3(S_2S_1) = (S_3S_2)S_1 \;.$$

This is true of symmetry operations because both sides evidently mean the same thing, 'S_1 followed by S_2 followed by S_3'. These rules are the defining properties of groups. That is, a *group* is an *associative algebra*[4] that has an *identity* and *inverses*. For a formal statement, see p. 42.

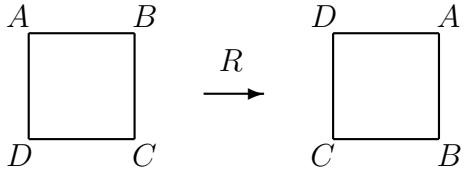
What about commutativity? The commutative rule does not, in general, hold for symmetries. That is, we generally have

$$S_2S_1 \neq S_1S_2 \;.$$

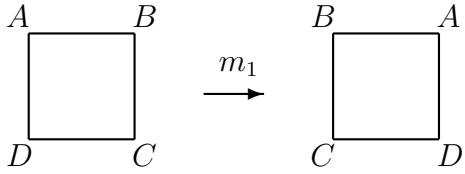
To illustrate this let $S_1 = R$, meaning clockwise rotation through 90°, applied to a square:

[3]Other symbols are in use, common choices being ι, e, or E.

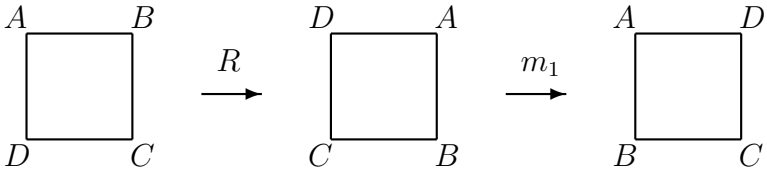
[4]There are non-associative algebras. A familiar example is the algebra of 3-dimensional physical vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ under vector multiplication: generally $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) \neq (\mathbf{a}\times\mathbf{b})\times\mathbf{c}$.



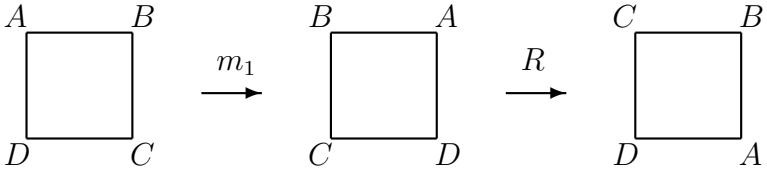
Let S_2 be another symmetry operation on the square, reflection or flipping about the vertical axis. Call this m_1:



We defined m_1R to mean first R then m_1; this produces



whereas Rm_1 means first m_1 then R,



so the result is indeed different, $m_1 R \neq R\, m_1$. Let's follow this up now by discussing the symmetries of a square more systematically.

Example 2.1 *Symmetries of the square, and some basic terminology*

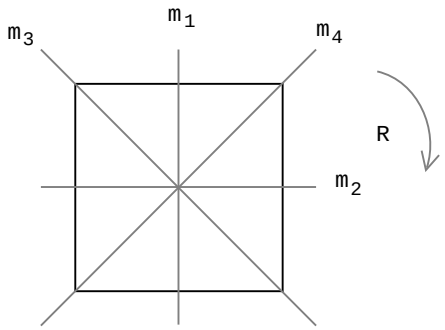


Fig. 7

Plainly the square is invariant under rotations by $0°$, $90°$, $180°$, and $270°$, denoted respectively by I, R, RR, RRR or

$$I,\ \ R,\ \ R^2,\ \ R^3\ ,$$

and invariant under reflections in the vertical and horizontal axes and in the two diagonals. We call these reflections

$$m_1,\ m_2,\ m_3,\ m_4$$

respectively, as noted in fig. 7. Each of these last four is its own inverse: ${m_1}^2 = {m_2}^2 = {m_3}^2 = {m_4}^2 = I$. Note also that $R^4 = I$.

(We can think of the m's either as true reflections within 2-dimensional space, or as $180°$ rotations or 'flips' in 3-dimensional space, about the axes shown.)

The 8 symmetries are algebraically related; for instance $m_1 m_2 = m_2 m_1 = m_3 m_4 = m_4 m_3 = R^2$. Again, $m_2 m_4 = m_4 m_1 = m_1 m_3 = m_3 m_2 = R$, and $m_3 m_1 = m_1 m_4 = m_4 m_2 = m_2 m_3 = R^{-1} = R^3$. Evidently there must be some minimal subset from which all other symmetries can be obtained by composition. We say that the complete set of symmetries — i.e., the group — is **generated** by such a subset, and the members of the subset are called **generators**. In this case we need exactly two generators.

For we need no more than one rotation R, or R^3, since all the other rotations are just multiples of R or R^3, including the inverses, e.g. $R^{-1} = R^3$. We obviously need at least one reflection (to flip the square over); and exactly one is sufficient because all the other flipped-over positions can evidently be reached by a subsequent rotation R, R^2 or R^3. For instance if we choose m_1 as a generator then we can derive $m_2 = R^2 m_1,\ \ m_3 = R^3 m_1,\ \ m_4 = R m_1$. In summary, the subset $\{R,\, m_1\}$ generates the group as follows:

$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \;=\; \{I, R, R^2, R^3, m_1, R^2 m_1, R^3 m_1, R m_1\}\ .$$

This group of symmetries is often called the **4-fold dihedral group** and denoted here by the symbol D_4. The corresponding group of symmetries of a regular polygon with n sides is similarly denoted[5] by D_n. All these groups are **point groups**, meaning that all the group operations leave one point fixed. In these examples the fixed point is the centre of the polygon.

Definition: The **order** $|G|$ of a group G is the number of elements it contains. So we have $|D_4| = 8$. It is easy to check that $|D_n| = 2n$; see q. 6 of Sheet 1.

Definition: The **order of a group element** $g \in G$ is the least integer q such that $g^q = I$. Because a group is an algebra, $\{I, g, g^2, \ldots, g^{q-1}\}$ are all elements of G. Therefore, for any finite group, $q \leqslant |G|$.

In the case of D_4, for instance, all the elements other than R, R^3 and I have order 2, while R and R^3 have order 4.

Remark: If an element g has order q, then $g^{-1} = g^{q-1}$. (So all the elements of D_4 are self-inverse except R and R^3. For the latter we have $R^{-1} = R^3$ and $(R^3)^{-1} = R$.

We can define the structure of a group by exhibiting its **group table** or **multiplication table**, so called, showing how all pairs of elements combine. For D_4 we have

[5] Also, sometimes, by Dih_n. The term 'dihedral', meaning 2-sided, comes from the 3-dimensional view in which all the symmetry operations are regarded as rotations in 3-dimensional space, with the m's as 180° (2-fold) rotations, or flips, about the in-plane symmetry axes of the polygon, and in which the polygon is therefore regarded as having distinct front and back surfaces like a real cardboard polygon. In crystallography other symbols are used. Unfortunately there is no single symbolic convention.

Identity:	I	R^2	R	R^3	m_1	m_2	m_3	m_4
180°rotation:	R^2	I	R^3	R	m_2	m_1	m_4	m_3
90°rotations:	R	R^3	R^2	I	m_4	m_3	m_1	m_2
	R^3	R	I	R^2	m_3	m_4	m_2	m_1
square flips:	m_1	m_2	m_3	m_4	I	R^2	R	R^3
	m_2	m_1	m_4	m_3	R^2	I	R^3	R
diagonal flips:	m_3	m_4	m_2	m_1	R^3	R	I	R^2
	m_4	m_3	m_1	m_2	R	R^3	R^2	I

By convention, the symbols in the first column show the first factor in the written-out product, and those in the first row show the second factor of the product, e.g. $Rm_1 = m_4$ and $m_1R = m_3$. (So **N.B.**, the symbols along the top show the **first** operation performed, and the symbols on the left the **second** operation performed.) The table incorporates the fact that multiplying by I, either from the left or from the right, has no effect. ⌟

Remark: Every row of the table is a **complete rearrangement** of every other row. Each element, without exception, has moved to another position.[6] This is true of any group table. It follows at once from the defining properties of groups. For we cannot have $g_1g = g_2g$ if g_1 and g_2 are different elements. Since every element g has an inverse g^{-1}, the equation $g_1g = g_2g$ can be multiplied on the right by g^{-1} to give $g_1 = g_2$, which would be a contradiction. (Notice how this argument uses associativity.) The same argument applies to the equation $gg_1 = gg_2$ if we multiply by g^{-1} on the left. So, not surprisingly, the same 'complete rearrangement' property is true of columns also. All this is useful to remember when checking the correctness of a group table.

Definition: A **subgroup** of a group G is a subset of G that is also a group.

The table shows that D_4 has five subgroups of order 2, namely $\{I, m_1\}$, $\{I, m_2\}$, $\{I, m_3\}$, $\{I, m_4\}$, and $\{I, R^2\}$. D_4 also has three subgroups of order 4, namely $\{I, R, R^2, R^3\}$, $\{I, R^2, m_1, m_2\}$ and $\{I, R^2, m_3, m_4\}$. The order-4 subgroup $\{I, R, R^2, R^3\}$ is for obvious reasons called a **cyclic group**. It will be denoted here[7] by C_4. The other two order-4 subgroups are examples of what's known as the 'Klein four-group' or 'Vierergruppe'.

[6] Pure mathematicians who work on combinatorics have a special term, 'derangements', for such complete rearrangements. (E.g. sudoku arrays, most of which *aren't* group tables.)

[7] Another commonly-used notation is Z_4.

Definition: If all the elements of a group G commute with each other ($g_1 g_2 = g_2 g_1$ for all $g_1, g_2 \in G$), then the group is said to be commutative or **abelian**.[8] E.g., C_4 and the Vierergruppe are abelian, but D_4 is non-abelian.

Definition: we say a mapping f between two groups (or more generally two sets) is **1–1** if $f(g_1) = f(g_2)$ implies $g_1 = g_2$. That is, f maps distinct elements to distinct elements.

Definition: we say a mapping f between two groups G, G' (or more generally two sets S, S') is **onto** if for every $g' \in G'$ there is at least one $g \in G$ such that $f(g) = g'$.

Definition: a 1–1 mapping between two groups of the same order is called an **isomorphism** if it preserves the group operations, namely composition and taking inverses.

(Note: a 1–1 mapping between two finite groups of the same order is necessarily onto, since each element of G is mapped to a different element of G' and $|G| = |G'|$.)

If such a mapping exists, then we say that the two groups are **isomorphic**, which means literally that they have the same shape or structure. More precisely, it means that they have the **same group table**. The elements of the two groups may well be different mathematical entities, and the notation used for the group elements may be different. But the isomorphism means that we may think of them, from an abstract viewpoint, as being the *same* group.

Remark: The groups $\{I, R, R^2, R^3\}$, $\{I, R^2, m_1, m_2\}$, and $\{I, R^2, m_3, m_4\}$ are all abelian, but only the last two are isomorphic.

Example 2.2 *A set of matrices isomorphic to* D_4

Define

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{R}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{R}^3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

$$\mathbf{m_1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{m_2} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{m_3} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{m_4} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

[8] after the great Norwegian mathematician Niels Henrik Abel (1802–1829).

These 8 mathematical entities — these 8 matrices — do indeed make up a group provided that composition is defined as matrix multiplication. Their group table exactly matches the table on page 34 above. In other words this set of 8 matrices forms a group that is isomorphic to D_4. *Make sure you check this.* Speed tip: remember the pattern $\left(\begin{smallmatrix} 0 & \bullet \\ \bullet & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & \bullet \\ \bullet & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} \bullet & 0 \\ 0 & \bullet \end{smallmatrix}\right)$.

Definition: If a group G is isomorphic to some set of $n \times n$ matrices that form a group under matrix multiplication, then the set of matrices is called an n-dimensional **faithful representation** of G.

Remark: Any such set of matrices must be invertible: their determinants must not vanish. (Why?)

Remark: $(g_1 g_2)^{-1} = g_2^{-1} g_1^{-1}$ for any $g_1, g_2 \in G$.

Proof: $(g_1 g_2)(g_2^{-1} g_1^{-1}) = g_1 (g_2 g_2^{-1}) g_1^{-1} = g_1 I g_1^{-1} = g_1 g_1^{-1} = I$.

(Notice how the proof depends on associativity. This property of inverses will be used repeatedly.)

Definition: Two group elements $g_1,\ g_2 \in G$ are said to be **conjugate** to each other — written $g_1 \sim g_2$ — if there exists *any* group element $g \in G$ such that

$$g_2 = g g_1 g^{-1} \ , \tag{2.1}$$

or equivalently $g_2 g = g g_1$.

Remark: Conjugacy is an equivalence relation, in the usual mathematical sense. This means that

(1) $g_1 \sim g_1$ always (because we can choose $g = I$),

(2) $g_1 \sim g_2 \ \Rightarrow \ g_2 \sim g_1$ (because we have inverses), and

(3) $g_1 \sim g_2$ and $g_2 \sim g_3 \ \Rightarrow \ g_1 \sim g_3$ (Check this!)

where as usual $\Rightarrow$ means 'implies that'.

In general, if the relation $\sim$ is defined on a set S and satisfies these properties (known respectively as reflexivity, symmetry and transitivity), then it is an equivalence relation on S, and partitions S into disjoint subsets known as equivalence classes. Any elements a and b in the same class obey $a \sim b$; any elements in distinct classes do not.

In this case, the conjugacy relation partitions any group G into disjoint classes of elements, called **conjugacy classes**. Any elements in the same class are conjugate; any elements in distinct classes are not.

Roughly speaking, in any group G, a conjugacy class consists of group elements that are structurally of the same type.

Example 2.3 *The conjugacy classes of* D_4

These correspond to the verbal categories shown at the left of the table on page 34. Thus for instance the identity I is in a conjugacy class by itself. (With $g_1 = I$ on the right-hand side of (2.1), all choices of g give $g_2 = I$.) The 180° rotation R^2 is in a class by itself. The square flips make up a different class of their own. So do the diagonal flips and the 90° rotations. ⌟

(*Important for subsequent developments*: make sure you check that the partitioning shown on page 34 is correct. To show this you must check that *none* of the 8 elements of D_4, when substituted for g on the right of (2.1), are able to eject any element g_1 from its class, for instance to give $g_2 = m_3$ when $g_1 = m_1$. For all choices of g, when $g_1 = m_1$, (2.1) must, and does, always give either $g_2 = m_1$ or $g_2 = m_2$.)

Remark: The identity of *any group* is in a class by itself. Each element of an *abelian group* is in a class by itself. A group element and its inverse, if different, may or may not be in the same class. Abelian groups illustrate the latter case. D_4 illustrates the former because of its class $\{R, R^3\}$.

Remark: For any group G of rotations in 3-dimensional space, two rotations g_1 and g_2 are in the same class if they rotate through the same angle about two different axes, *provided that* there is another rotation g available in G that can rotate the one axis into the other. ('Same angle' means same magnitude *and* same sign.)

The diagonal flips in D_4 are a case in point when viewed in 3-dimensional space. Not only are they both 180° rotations but, also, there are other elements available in D_4, namely R and R^3, that can rotate one diagonal axis into the other. So doing m_4 to the cardboard square is the same as rotating it 90° anticlockwise, doing m_3, then rotating it 90° clockwise. So (2.1) does apply with g chosen as R: specifically, $m_4 = Rm_3R^{-1}$.

Similarly, for the square flips in D_4, we have $m_2 = Rm_1R^{-1}$. And for the 90° rotations we have $R^3 = mRm^{-1} = mRm$ where m stands for any one of the four flips, since they all serve equally well to flip the rotation axis normal to the plane of the square, which turns clockwise rotation into anticlockwise rotation.

Notice that the partitioning shown on page 34 does not apply to the separate subgroups, because within a subgroup — if we exclude the trivial 'subgroup' consisting of D_4 itself — there are fewer choices of g available for use in (2.1). For instance, if we consider the diagonal flips within the Vierergruppe $\{I, R^2, m_3, m_4\}$, then each one is in a class by itself. The 90° rotations are no longer available for use in (2.1).

Definition: A **normal subgroup** H of a group G is a subgroup that consists of complete conjugacy classes of G.

(Trivially, G is a normal subgroup of itself. Similarly, the one-element group $\{I\}$ is a normal subgroup of G. To exclude these trivial cases, we sometimes speak of a **proper subgroup** or of a **proper normal subgroup** of a group G, meaning a subgroup smaller than G itself, and not just the identity element.)

(Note also that every subgroup of an abelian group is normal: you should check that this follows from the definitions.)

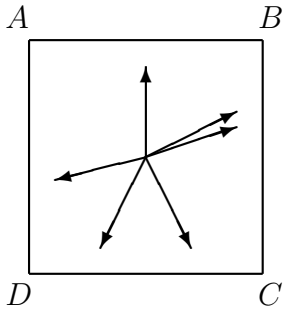
Example 2.4 D_4

The three order-4 subgroups of D_4 are normal, as is the order-2 subgroup $\{I, R^2\}$. The other four order-2 subgroups of D_4 are not normal. *However*, for instance, $\{I, m_1\}$ and $\{I, m_2\}$ *are* normal subgroups of the Vierergruppe $\{I, R^2, m_1, m_2\}$. There are fewer elements g available for use in (2.1). ⌟

2.2 The underlying vector space

There's another way to describe symmetry operations, such as those of D_4, that will prove important.

Imagine some vectors $\mathbf{x}$ drawn on the square, with their tails at the centre but otherwise arbitrary:



Representing them as column vectors $(x_1,\, x_2)^{\mathsf{T}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, we can consider matrix products such as

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;=\; \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} .$$

The right-hand side is the vector rotated clockwise through a right angle. That is, premultiplication by this particular 2×2 matrix rotates all the vectors clockwise through a right angle. (We are thinking of the coordinate axes as being fixed, while the coordinate values x_1, x_2 change.) So we can imagine the square itself being carried around with the vectors and thus rotated as well.

It's easy to see that all the symmetry operations in D_4 can be represented in this way. The isomorphism noted earlier is no accident! Premultiplication of any $\mathbf{x} = (x_1,\, x_2)^{\mathsf{T}}$ by the same set of matrices as on p. 35,

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{R}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{R}^3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

$$\mathbf{m_1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{m_2} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{m_3} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{m_4} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ,$$

performs the same symmetry operations, $I, R, R^2, R^3, m_1, m_2, m_3, m_4$ respectively, as seen within 2-dimensional space.

The action of, for instance, m_4 on $\mathbf{x} = (x_1,\, x_2)^{\mathsf{T}}$ is to turn it into $(x_2,\, x_1)^{\mathsf{T}}$. And the fact that $m_4 = Rm_3R^{-1}$, noted earlier, implies that the action of Rm_3R^{-1} on $\mathbf{x}$ must be the same. So the isomorphism says that we must have $\mathbf{m}_4\mathbf{x} = \mathbf{R}\mathbf{m}_3\mathbf{R}^{-1}\mathbf{x}$, which written out is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;=\; \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;=\; \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} ,$$

which can easily be checked to be true. [Again, remember $\left(\begin{smallmatrix} 0 & \bullet \\ \bullet & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & \bullet \\ \bullet & 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} \bullet & 0 \\ 0 & \bullet \end{smallmatrix}\right)$.]

Now consider the subsets

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{R}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{m_1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{m_2} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{R}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \mathbf{m_3} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{m_4} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

Each set of four gives a 2-dimensional faithful representation of the Vierergruppe. (Check the isomorphisms.)

Remember: One can think of the vectors $\mathbf{x}$ as elements of the 2-dimensional vector space $\mathbb{R}^2$. You should check that they satisfy the rules for vector spaces, with scalars in $\mathbb{R}$.

Notice, now, that the first representation of the Vierergruppe leaves invariant the two **subspaces** of $\mathbb{R}^2$ consisting of scalar multiples of the vectors $(1, 0)$ and $(0, 1)$. The second representation leaves invariant the two subspaces of $\mathbb{R}^2$ consisting of scalar multiples of the vectors $(1, 1)$ and $(-1, 1)$. (Check this!)

The idea of **subspaces invariant under the action of a matrix representation of a group** has far-reaching importance, as we'll see.

Example 2.5 *Worked example: the 3-fold dihedral group D_3*

Construct the group table for the dihedral group D_3, that is, the symmetries of a 2-sided equilateral triangle. Show that there are three order-2 non-normal subgroups and one order-3 normal subgroup, and show the partitioning into conjugacy classes. By considering the group actions on vectors in $\mathbb{R}^2$, or otherwise, find a 2-dimensional faithful representation of D_3. Show that within $\mathbb{R}^2$ there are subspaces invariant under the actions of each order-2 subgroup.

Answer: We take the centroid of the triangle as origin and the vertices of the triangle at $(0, 1)$, $(\sqrt{3}/2, -1/2)$ and $(-\sqrt{3}/2, -1/2)$. By analogy with the notation for D_4, denote the group elements of D_3 by $\{I, R, R^2, m_1, m_2, m_3\}$ where R means a clockwise rotation through $120°$, m_1 means a flip about the vertical or x_2 axis, and m_2 and m_3 respectively the flips about axes rotated $120°$ and $240°$ clockwise from the vertical, that is, axes with unit vectors $(\sqrt{3}/2, -1/2)$ and $(-\sqrt{3}/2, -1/2)$. The group table works out as follows:

Identity:	I	R	R^2	m_1	m_2	m_3
rotations:	R	R^2	I	m_3	m_1	m_2
	R^2	I	R	m_2	m_3	m_1
	m_1	m_2	m_3	I	R	R^2
flips:	m_2	m_3	m_1	R^2	I	R
	m_3	m_1	m_2	R	R^2	I

It is straightforward to check from (2.1), remembering to try all possible $g \in D_3$, that the three flips form one conjugacy class, and the two rotations another, as indicated in the table. As always, the identity I is in a class by itself. As for the subgroups, we see from the I's on the far diagonal of the table that the three order-2 subgroups are $\{I, m_1\}$ $\{I, m_2\}$ and $\{I, m_3\}$, all non-normal because the set $\{m_1\}$ isn't a conjugacy class in D_3, nor $\{m_2\}$ nor $\{m_3\}$. The order-3 subgroup, the only such, is the cyclic group C_3 made up of $\{I, R, R^2\}$. This is a normal subgroup of D_3 because $\{R,\ R^2\}$, $\{I\}$ *are* classes in D_3.

To find a 2-dimensional faithful representation of D_3, the easiest way is to remember that a 2-dimensional rotation matrix for clockwise rotation through an angle θ is

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and that $\cos 120^\circ = -1/2$ and $\sin 120^\circ = \sqrt{3}/2$. So taking the corresponding rotation matrices to represent R and R^2, and playing around with the signs to convert to flips (regarding them as reflections within our 2-dimensional space), we end up with the isomorphism

$$I,\ R,\ R^2 \quad \leftrightarrow \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \tag{2.2}$$

$$m_1,\ m_2,\ m_3 \quad \leftrightarrow \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}. \tag{2.3}$$

The subspaces of $\mathbb{R}^2$ invariant under the actions of the three order-2 subgroups are six in number. They are, respectively, scalar multiples of $(1, 0)$ and $(0, 1)$ for $\{I, m_1\}$, of $(1/2,\ \sqrt{3}/2)$ and $(-\sqrt{3}/2,\ 1/2)$ for $\{I, m_2\}$, and of $(\sqrt{3}/2,\ 1/2)$ and $(-1/2,\ \sqrt{3}/2)$ for $\{I, m_3\}$.

2.3 More on the general concept of 'group'

Let's now take stock and summarize the basic concepts as seen so far. This is partly because we have been a bit cavalier about one or two issues. For instance we talked about 'the' inverse, and 'the' identity, as if these elements were always uniquely defined.

However, by revisiting the rules very carefully we can see that the inverse and identity are, in fact, uniquely defined. The following is partly summary and revision, and partly pinning down the elementary concepts and making them precise, in the full generality of abstract group theory:

Definition: A **group** G is a set of elements, say

$$G = \{g_1,\ g_2,\ g_3,\ \ldots\}$$

together with a rule for combining them — we'll continue to use the words 'product', 'composition', and 'multiplication' in this connection — such that the following conditions or **axioms** are all satisfied:

1. **Closedness** or **closure**: the product $g_i g_j \in G$ (i.e., G is an algebra),
2. **Associativity**: the product satisfies $(g_i g_j)g_k = g_i(g_j g_k)$,
3. Existence of an **identity**, $I \in G$ such that $Ig = gI = g$ for all $g \in G$,
4. Existence of an **inverse**, for each $g \in G$ an element g^{-1} such that

$$g^{-1}g = gg^{-1} = I\ .$$

Remark: The identity is unique.

Proof: Suppose that there exist two elements I and I' each having the properties specified in Axiom 3. Then

$$II' \ = \ I'\,, \ \text{since } I \text{ has the properties specified}$$

and

$$II' \ = \ I\,, \ \text{since } I' \text{ has the same properties}\,.$$

The left-hand sides of these equations are the same, and so $I' = I$.

Remark: Inverses are unique.

Proof: Suppose both h and k are inverses of a given element g. Then Axiom 4 implies that, in particular,

$$gh = I \quad \text{and} \quad kg = I. \tag{2.4}$$

Then (premultiplying the first of these by k and using Axiom 3) we have

$$k(gh) = k\ ,$$

i.e. (by Axiom 2)

$$(kg)h = k\ .$$

But $kg = I$, by the second of (2.4). So using Axiom 3 again we have $h = k$.

(The two proofs just given typify the sort of proof required in abstract pure mathematics. The name of the game is to work from the axioms and from the axioms alone, and to make all the logical steps explicit, no matter how trivial. In this game, appeals to intuition aren't allowed, even if something *seems* obvious. The purpose is not to 'see' that something 'must' be true, but rather to prove it — in the mathematician's rigorous sense — so we can be sure that the proof applies in the abstract and really does include all the possibilities, not just some special cases we might have in mind, and for which we might or might not have a reliable intuition. Actually, most people with powerful scientific or mathematical intuitions have gone through harsh experiences in which intuition, always hazardous, is corrected, refined and developed by confrontation with counterintuitive facts established by meticulous experimental observation, or by rigorous mathematical proof.)

2.4 Cosets and Lagrange's theorem

2.4.1 Cosets

Definition: Given a group G, a subgroup $H = \{I, h_1, h_2, \dots\}$ of G, and an element g in G, we call the set

$$gH{=}\{g,\, gh_1,\, gh_2, \dots\}$$

a **left coset** of H in G. Similarly, we call the set

$$Hg{=}\{g,\, h_1g,\, h_2g, \dots\}$$

a **right coset** of H in G.

Remark: The left and right cosets of any subgroup H are identical if G is abelian.

Remark: The left and right cosets of any **normal** subgroup H are always identical. (Proof below).

Remark: gH and Hg each contain $|H|$ elements. (This follows from invertibility, which implies that the elements g, gh_1, $gh_2, \dots$, are distinct, and similarly g, h_1g, $h_2g, \dots$, because I, h_1, $h_2, \dots$ are distinct.)

Remark: The subgroup H and its left cosets partition G. That is, they divide G into disjoint subsets. The subgroup H and its right cosets also

partition G, though differently unless G is abelian or H normal. It is a partitioning because

(i) Two cosets are either disjoint or identical.

(ii) Two cosets Hg_1 and Hg_2 are identical if and only if $g_1 g_2^{-1} \in H$.

(iii) Every element of G is in some coset.

Proof of (i): Suppose that the cosets Hg_1 and Hg_2 have one element in common, which we can write either as $h_1 g_1$ or as $h_2 g_2$, for some h_1 and h_2 in H. Then

$$Hg_1 = Hh_1^{-1} h_2 g_2 = Hg_2,$$

since $h_1^{-1} h_2 \in H$. So if there is one element in common, the cosets are identical.

Proof of (ii): If $g_1 g_2^{-1} \in H$, then let $g_1 g_2^{-1} = h$. We have

$$Hg_1 = Hhg_2 = Hg_2$$

as required.

Conversely, suppose $Hg_1 = Hg_2$. Then $Hg_1 g_2^{-1} = H$, which means that $hg_1 g_2^{-1} \in H$ for all $h \in H$, including $h = I$, which proves the result.

Proof of (iii): Since H contains I, then for any element $g \in G$, the coset Hg contains g.

Notice the consequence that 'being in a coset' is an equivalence relation. This is not, of course, the same equivalence relation as conjugacy.

A further consequence is that we may pick out any element of a coset as its 'representative'. That is, we may specify which coset we are talking about merely by naming one of its elements. Furthermore, we may multiply such an element by any $h \in H$, and it will still represent the same coset.

Proof that the left and right cosets of any normal subgroup H are identical: Let H be normal. For any $g \in G$, we need to show that gH and Hg are the

same coset. We use the fact that, since H is normal, the conjugate ghg^{-1} of any $h \in H$ is also in H, for any $g \in G$. The proof is as follows:

Let $g \in G$ be given, and let $k \in gH$. Then $k = gh$ for some $h \in H$. Let $h_1 = ghg^{-1}$. Then $h_1 \in H$ (because H is normal), and $h = g^{-1}h_1g$. So we have $k = gh = g(g^{-1}h_1g) = (gg^{-1})h_1g = h_1g$. But this says that $k \in Hg$.

So for any given $g \in G$, $\quad k \in gH \;\Rightarrow\; k \in Hg$. A similar chain of reasoning proves that $\; k \in Hg \;\Rightarrow\; k \in gH$. So the left and right cosets of H are identical.

(Conversely, if all the left and right cosets are identical, then we must have $gH = Hg$ for all g, since $g \in gH$ and $g \in Hg$. So $gHg^{-1} = H$ for all g, and so H is normal.)

2.4.2 Lagrange's theorem

Let G be a finite group and let H be a subgroup of G (not necessarily normal). Lagrange's theorem says that $|H|$ divides $|G|$. That is, $|G| = n|H|$ where n is an integer.

This result follows immediately from the 'partitioning property' of cosets. Suppose that there are n distinct right cosets of H, counting H itself as one of them (the trivial right coset HI). Then all the right cosets are disjoint, and each has the same number of elements, namely $|H|$. Each element of G is in exactly one coset. Thus $|G| = n|H|$, as required.

(Of course we could equally well have used left cosets.)

An immediate corollary of Lagrange's theorem is the following

Remark: The order of every element of G divides $|G|$.

This follows from Lagrange's theorem and the fact that each element generates a cyclic subgroup, whose order is equal to the order of its generating element.

Example 2.6 *The group D_4 as an illustration of Lagrange's theorem*

Within D_4 we have the cyclic subgroup C_4 generated by R or by R^3. This has order 4, as do the elements R and R^3 themselves. And 4 does indeed divide $|G|$, which is 8. Again, D_4 has two copies of the Vierergruppe, whose order is 4, as subgroups.

The elements R^2, m_1, m_2, m_3, and m_4 all have order 2 and generate subgroups of order 2; and 2 also divides 8. ⌟

Example 2.7 *Worked example: all possible order-4 groups*

Prove that every order-4 group G is isomorphic either to C_4 or to the Vierergruppe.

Answer: Let $G = \{I, g_1, g_2, g_3\}$, all distinct. Lagrange's theorem and its corollary tell us that each element of G must have order 1, 2 or 4, because these are the only divisors of 4. So there are two possibilities:

(1) G contains an element of order 4. Then that element generates the whole group G. In other words, G is cyclic and has the same group table as C_4, i.e., is isomorphic to C_4.

(2) G contains no element of order 4. Then g_1, g_2, g_3 must each have order 2, since the only other possibility is an element of order 1, which is the identity I. So $g_1^2 = g_2^2 = g_3^2 = I$. Now consider g_1g_2. We cannot have $g_1g_2 = g_1$, because premultiplication by g_1 gives a contradiction, $g_2 = I$. Similarly, we cannot have $g_1g_2 = g_2$. Therefore $g_1g_2 = g_3$. Similarly, $g_2g_1 = g_3$, and $g_2g_3 = g_3g_2 = g_1$, $g_3g_1 = g_1g_3 = g_2$. So the group table is exactly that of the Vierergruppe, to which G is therefore isomorphic. ⌟

2.5 Homomorphisms and their kernels

A **homomorphism** is a map between groups that preserves group operations but is not necessarily 1–1 or onto. (As we saw earlier, if the homomorphism is 1–1 and onto, then the groups have the same order, and it's also an isomorphism.) If it's not 1–1 then there's a nontrivial subgroup of the first group that maps to the identity in the second, i.e. target, group. That subset of the first group is called the **kernel** of the homomorphism. The kernel will be shown to be a normal subgroup.

Let's fill in the details and note some examples. Let G and H be any two groups. (In this section, G and H are arbitrary groups. So H might or might not be a subgroup of G.) A *map* or *mapping* Φ from G to H is a rule

that associates an element of H, the target group, with each element of G. We write

$$\Phi : G \longrightarrow H$$

for the map, and

$$\Phi(g)$$

for the element in H to which the element g of G is mapped. We call $\Phi(g)$ the **image** of g. We are interested in maps that preserve group operations.

Definition: The map Φ is called a **homomorphism** from G to H if and only if

$$\Phi(g_1 g_2) = \Phi(g_1)\Phi(g_2) \tag{2.5}$$

for any g_1 and g_2 in G. That is, the image of the product is the product of the images.

If the map Φ is 1–1 and onto then it is also an isomorphism. The rest of this section applies to homomorphisms in general. We note next that if products are preserved then it *follows* that other group operations are preserved:

Remark: Setting $g_2 = I_G$ (the identity of the group G) in the defining property (2.5) shows that for all $g_1 \in G$

$$\Phi(g_1) = \Phi(g_1)\Phi(I_G), \quad \Rightarrow \quad \Phi(I_G) = I_H$$

(identity maps to identity); and setting $g_1 = g$ and $g_2 = g^{-1}$ in (2.5) shows that

$$\Phi(I_G) = I_H = \Phi(g)\Phi(g^{-1}), \quad \Rightarrow \quad \Big[\Phi(g)\Big]^{-1} = \Phi(g^{-1})$$

(inverses map to inverses).

Example 2.8 $\Phi : \mathbb{R} \longrightarrow U(1)$

Let $G = \mathbb{R}$, the group of real numbers x under addition, and let $H = U(1)$, the multiplicative group of unit-magnitude complex numbers. Let the map Φ be defined by

$$\Phi(x) = e^{ix}.$$

This is a homomorphism because

$$\Phi(x+y) = e^{i(x+y)} = e^{ix}e^{iy} = \Phi(x)\Phi(y).$$

Note that many elements in G map to each element in H:

$$\Phi(x + 2N\pi) = \Phi(x), \text{ where } N \text{ is any integer}.$$

Maps with this property are called **many-one**. ⌟

There are further simple examples in q. 4 of Sheet 2. They will be important for the work on representation theory.

Remark: In this course we are mostly interested in homomorphisms that are onto. For these, the image of G, i.e. the set $\{\Phi(g)\}|_{g\in G}$ is the whole target group H. For one example where this is not the case see the 2004 examination, question 9 of paper 2.

For a many-one homomorphism, more than one element (including the identity) will map to the identity.

Definition: Given a homomorphism $\Phi : G \to H$, the set of all elements $k \in G$ such that $\Phi(k) = I_H$ is called the **kernel** of Φ, denoted here by K. In other words,

$$k \in K \quad \Leftrightarrow \quad \Phi(k) = I_H \ . \tag{2.6}$$

In Example 2.8, $K = \{2N\pi\} = \{\ldots, -2\pi, 0, 2\pi, 4\pi, \ldots\}$.

Remark: The kernel K is a subgroup of G.

This is because

1. It is closed. (If $k_1, k_2 \in K$ then, by the defining property (2.5) of a homomorphism, $\Phi(k_1k_2) = \Phi(k_1)\Phi(k_2) = I_HI_H = I_H \ \Rightarrow \ k_1k_2 \in K$ as required.)

2. $I_G \in K$. ($\Phi(I_G) = I_H$ because Φ is a homomorphism).

3. If $k \in K$, then $k^{-1} \in K$. (For $\Phi(k^{-1}) = \big[\Phi(k)\big]^{-1} = I_H^{-1} = I_H$.)

Remark: K is a *normal* subgroup of G.

Proof: We need to prove that $k \in K \ \Rightarrow \ gkg^{-1} \in K$ for any $g \in G$, i.e., that $\Phi(k) = I_H \ \Rightarrow \ \Phi(gkg^{-1}) = I_H$. This follows from (2.5) and associativity, giving $\Phi(gkg^{-1}) = \Phi(g)\Phi(k)\Phi(g^{-1}) = \Phi(g)I_H\big[\Phi(g)\big]^{-1} = I_H$, as required.

Example 2.9 *Worked example: cosets in D_3*

List all the cosets of all the subgroups D_3.

Answer: In the notation of page 40, we see from the group table that the left and right cosets of the order-3 subgroup $\{I, R, R^2\}$ (in addition to the subgroup itself) are both $\{m_1, m_2, m_3\}$. (So the subgroup $\{I, R, R^2\}$ is a normal subgroup, as noted earlier.)

(Any order-3 subgroup of any order-6 group has to be normal, of course, as does an order-n subgroup of any order-$2n$ group — for instance all three of the order-4 subgroups of D_4 — simply because the cosets partition the group and in each such case there's room for only one nontrivial coset.)

Now to the order-2 subgroups of D_3. (As also noted earlier, they're not normal. So we'll find that their left and right cosets are distinct, providing different partitionings of D_3.) For instance, the two left cosets of the order-2 subgroup $\{I, m_3\}$ are

$$R\{I, m_3\} = m_2\{I, m_3\} = \{R, m_2\}\,, \qquad R^2\{I, m_3\} = m_1\{I, m_3\} = \{R^2, m_1\},$$

while the two right cosets of $\{I, m_3\}$ are

$$\{I, m_3\}R = \{I, m_3\}m_1 = \{R, m_1\}\,, \qquad \{I, m_3\}R^2 = \{I, m_3\}m_2 = \{R^2, m_2\}.$$

Similarly, we find that

the left cosets of $\{I, m_1\}$ are $\{R, m_3\}$ and $\{R^2, m_2\}$;

the right cosets of $\{I, m_1\}$ are $\{R, m_2\}$ and $\{R^2, m_3\}$;

the left cosets of $\{I, m_2\}$ are $\{R, m_1\}$ and $\{R^2, m_3\}$;

the right cosets of $\{I, m_2\}$ are $\{R, m_3\}$ and $\{R^2, m_1\}$. ⌟

Definition: The **product of two cosets** is the set of all products of two elements with one from each coset.

Example 2.10 *Products of the order-3 cosets that partition D_3*: The cosets $\{I,\ R,\ R^2\}$ and $\{m_1,\ m_2,\ m_3\}$ have a product consisting of the set $\{Im_1,\ Im_2,\ Im_3,\ Rm_1,\ Rm_2,\ Rm_3,\ R^2m_1,\ R^2m_2,\ R^2m_3\}$, which is the same as the set $\{m_1,\ m_2,\ m_3\}$. Similarly, $\{m_1,\ m_2,\ m_3\}^2 = \{I,\ R,\ R^2\}^2 = \{I,\ R,\ R^2\}$.

Remark: For any normal subgroup K of any group G, the product of any two of the cosets of K, say g_1K and g_2K ($g_1,\ g_2 \in G$), is the same as the coset $(g_1g_2)K$. (So the product is no bigger than the cosets we started with. Even though there are $|K|^2$ possible two-element products to be considered

in forming the product of the cosets, as in the case $|K| = 3$ just noted, these products are not all distinct and, in fact, yield only $|K|$ distinct elements if the subgroup K is normal.) This is easy to prove:

Proof: $g \in$ product of g_1K and $g_2K \Leftrightarrow g = g_1k_1g_2k_2$ for some $k_1,\ k_2 \in K \Leftrightarrow g \in (g_1g_2)K$. The last step makes use of the assumption that K is normal: $g = g_1k_1g_2k_2 = g_1\,(g_2g_2^{-1})\,k_1g_2k_2 = g_1g_2\,(g_2^{-1}k_1g_2)\,k_2 = (g_1g_2)(k_3k_2)$ for some $k_3 \in K$, because K is normal. And $k_3k_2 \in K$.

Remark: Any normal subgroup K and its cosets in G form a group under the definition of the product just given. This is called the **quotient group**, denoted by G/K, for which K is the identity element. (In Example 2.10, $G/K = D_3/\{I,\, R,\, R^2\}$ and consists of the two objects $\{I,\, R,\, R^2\}$ and $\{m_1,\, m_2,\, m_3\}$. Under the definition of the product they form a group isomorphic to C_2, which of course is the only possibility for a group of order 2.)

Corollary: If K is the kernel of a homomorphism $\Phi : G \to H$ and the image of Φ is all of H, then G/K is isomorphic to H.

(The proof is left as a simple exercise. We may summarize it by saying that, with the above definition of product, the quotient group 'inherits' its group operations from the 'parent' group G.)

(In Example 2.10, the mapping $D_3 \to C_2$ specified by $\{I,\, R,\, R^2,\, m_1,\, m_2,\, m_3\} \mapsto \{1,\, 1,\, 1,\, -1,\, -1,\, -1\}$, with C_2 represented as $\{1,\, -1\}$, is a homomorphism (check this!). It has kernel $K = \{I,\, R,\, R^2\}$. So the above Corollary, with $G = D_3$ and $H = C_2$, says again that D_3/K is isomorphic to C_2, or equivalently that D_3/C_3 is isomorphic to C_2.)

Remark: Products of cosets of a non-normal subgroup H need not have $|H|$ elements. (In particular, products of the order-2 cosets in D_3 need not be of order 2. For instance, $\{R, m_1\}\{R^2, m_2\} = \{RR^2,\, Rm_2,\, m_1R^2,\, m_1m_2\} = \{I,\, m_1,\, m_3,\, R\}$. So this product contains not 2, but 4 elements.)

We now turn to a fundamental category of groups traditionally called 'symmetric' groups and also, more aptly, **general permutation groups**. They will be denoted here[9] by Σ_N $(N = 1,\, 2,\, 3, \ldots)$. They are fundamental because they contain as subgroups every other finite group.

[9] Another commonly-used notation is S_N.

More precisely, every order-N finite group is isomorphic to a subgroup of Σ_N. This is sometimes called **Cayley's theorem**. It follows at once from the 'complete rearrangement' property of group tables. The elements of Σ_N consist of all possible rearrangements of N objects. Complete rearrangements are the special cases seen in group tables. And if a set of complete rearrangements comes from a group table, then it automatically respects group operations and forms not merely a subset, but a subgroup, of Σ_N.

For example, we'll see in a moment that D_4, which has order 8, is a subgroup of Σ_8 (and in fact of Σ_7, Σ_6, Σ_5, and Σ_4 also).

2.6 The general permutation group Σ_N

Consider N distinct objects occupying N prescribed positions. The four labelled corners of our square on page 31 would be one example, with $N = 4$. If we take the square away, leaving only the labels $\{A, B, C, D\}$ and consider them to be 4 separate objects in their own right, then there are still further ways to rearrange them. We may put A into any one of the 4 available positions, then B into one of the 3 remaining, etc. There are $4 \times 3 \times 2 \times 1 = 4! = 24 = |\Sigma_4|$ possible such arrangements. That is, there are $|\Sigma_4| = 24$ possible rearrangement operations, or **permutations**.

Exactly as with the original problem of the square, we can reconfirm that **permutations are group operations**. We have **closure**: two successive permutations make another permutation. There is an **identity**, a do-nothing permutation. There is an **inverse**: we can do a permutation then undo it. Successive permutations obey the **associative law**: when we do three successive permutations it doesn't matter whether we think of the first two together and then the last, or the first by itself and then the last two together. 'Three given permutations in succession' has a unique meaning regardless of whether two of them are bracketed together. The brackets are superfluous.

Remark: Σ_{N-1} is a subgroup of Σ_N for all $N \geqslant 2$. (For if we always leave one of the objects fixed, then permutations of all the others still form a group.)

Remember, Σ_N concerns any set of N *distinguishable* objects occupying N *prescribed* positions. (Quantum objects are not always distinguishable and so need to be treated differently!) A pack of cards would be another example, with $N = 53$, counting the joker. Notice that the idea of the *pack* carries with

it the idea of a set of definite positions, top, second top,..., bottom. We may therefore speak of permutation — alias shuffling — in exactly the required sense. The number of permutations is $53! \approx 4 \times 10^{69}$. Permutation groups such as Σ_N can be seriously large.[10]

Now let's zoom in on some more detail. It's convenient to specify a given permutation using the notation

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & N \\ 3 & 5 & 2 & \cdots & 1 \end{pmatrix} ,$$

which means that the object originally in the first position is moved to the third position, the object originally in the second position is moved to the fifth position, etc. [**Health warning:** these symbols are **not** matrices.] It's best to think of the moves as simultaneous: all the objects begin moving from their original positions to their new positions at the same moment, and arrive in their new positions at some later moment. (If one imagines moving

[10]Stirling's approximation tells us that $N! \sim (N/e)^N$, i.e. that the size of Σ_N grows super-exponentially in N. This is reflected in the *computational complexity* of the travelling salesman problem and many other computational problems. A brute force search to find the shortest path joining N cities requires searching all $N!$ permutations, which is infeasible for even moderately large N. Better algorithms exist, but even the best known algorithms require $\sim \exp(CN)$ steps.

Think about the number of rearrangements of the 3×10^9 base pairs in the human genome, $4^{(3\times 10^9)} > 10^{10^9}$. The suggestion that science depicts life as 'mere machinery' — with the unconscious assumption that 'machinery' means something simple, like clockwork — is profoundly wrong. That's why life can be so awesomely wonderful, varied and unpredictable. Combinatorial largeness, the largeness of quantities like 10^N or $N!$ as N gets large, seems largely unknown to governments but is crucial to understanding. Understanding, in turn, brings with it respect and humility regarding complex systems and complex problems, including IT and software ergonomics, and genetic engineering — 'wicked problems', as they're called by some systems analysts. Such respect includes respect for the combinatorially large numbers of ways for things to go wrong, reversibly or irreversibly. With large IT projects (reversible to some extent) this means only the familiar dysfunctionality, security breaches, and billion-pound cost overruns. With genetic engineering (perhaps irreversible—and even the programming *language* is largely unknown) it might or might not mean a permanently altered biosphere that makes a few billion pounds look like peanuts. This is one reason why keeping science open is so momentously important — open scientific discussion is a form of massively parallel problem solving that can at least to some extent help in dealing with complexity. Open-source software is one successful example. (One illustration of the threat that closed source software providers perceive from open source software can be found in the so-called Halloween Documents (details available from all reputable web search providers).)

the objects in succession, then there is a danger of forgetting the distinction between original positions and new positions; remember, the first row of the symbol refers only to original positions, and the second row to new positions.) The inverse operation, in this notation, can be written

$$\begin{pmatrix} 3 & 5 & 2 & \cdots & 1 \\ 1 & 2 & 3 & \cdots & N \end{pmatrix} .$$

The order in which the columns are written is immaterial. Remember that the numbers refer to positions — the fixed positions from which, and into which, the objects are moved. The numbers do not refer to the objects themselves, which will continue to be denoted by $\{A, B, C, ...\}$. For example, a group operation (permutation operation) in Σ_5 , such as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix} ,$$

applied to the five objects $\{A, B, C, D, E\}$, will rearrange them into the different sequence $\{C, A, B, E, D\}$. Again, the product written on the left-hand side of

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 2 & 3 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 3 \end{pmatrix}$$

(in which, remember, the permutation written second acts first) will rearrange $\{A, B, C, D, E\}$ into $\{C, A, B, E, D\}$ and then into $\{A, B, E, C, D\}$. You should check that the permutation on the right-hand side also takes $\{A, B, C, D, E\}$ into $\{A, B, E, C, D\}$.

The identity permutation can be written, in our notation, as

$$I = \begin{pmatrix} 1 & 2 & 3 & \cdots & N \\ 1 & 2 & 3 & \cdots & N \end{pmatrix} ,$$

though to save writing we'll just continue to call it I. It's natural to save writing in some other ways as well:

First shorthand convention: A permutation that rearranges only some of the objects $\{A, B, ...\}$, such as

$$\begin{pmatrix} 2 & 3 & 1 & 5 & 4 \\ 3 & 1 & 2 & 5 & 4 \end{pmatrix} ,$$

is often written in the shorthand form

$$\begin{pmatrix} 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix} , \tag{2.7}$$

whose meaning is clear provided we remember we're talking about permutations in Σ_5 . For a general permutation in Σ_N , the convention says that any positions not shown contain objects that stay where they are.

Definition: An ***n*-cycle** in Σ_N $(n \leqslant N)$ is a permutation that acts only on the positions p_r $(r = 1, 2, ..., n)$ — i.e. acts only on a subset of the full set of positions $1, 2, ..., N$ — and has the special *tail-biting form*

$$\begin{pmatrix} p_1 & p_2 & p_3 & \cdots & p_n \\ p_2 & p_3 & p_4 & \cdots & p_1 \end{pmatrix} .$$

Second shorthand convention: The notation

$$(p_1\ p_2\ p_3\ \cdots p_n)$$

means the same n-cycle. The order in which the positions are shown now matters. By convention, the objects are moved, from each position shown, to the next position *on the right* except that the object in the last position shown moves to the first position shown. (A better notation would be to display the positions in a circle, but that would waste paper.)

Notice that the symbol in the display (2.7) above depicts a 3-cycle. With our conventions, it could equally well be depicted as

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{or} \quad (1\ 2\ 3) \quad \text{or} \quad (2\ 3\ 1)\ . \tag{2.8}$$

This particular 3-cycle takes $\{A, B, C, D, E\}$ into $\{C, A, B, D, E\}$.

Remark: Any permutation can be decomposed uniquely into disjoint cycles.

For consider the general permutation in Σ_N, fully written out as, say,

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & N \\ p_1 & p_2 & p_3 & \cdots & p_N \end{pmatrix} . \tag{2.9}$$

In this notation we can rearrange the N columns in any way whatever, without changing the meaning. So, starting from the left, we can rewrite the same permutation as

$$\begin{pmatrix} 1 & p_1 & p_r & \cdots \\ p_1 & p_r & p_s & \cdots \end{pmatrix} ,$$

continuing until there is a subset of columns on the left that has been rearranged into tail-biting form, with position 1 at the bottom of the last column in the subset. If this last column is the N^{th} then we are finished: the subset is the complete set and we have shown that (2.9) is an N-cycle. If not, then the process is repeated until the complete set of columns has been partitioned into successive tail-biting subsets — that is, partitioned into disjoint cycles — as was to be shown. Apart from the order in which the subsets of columns are written, which is irrelevant, this partitioning is evidently unique.

Example 2.11 *A permutation in* Σ_6

The following permutation partitions (uniquely) into a 2-cycle and a 4-cycle:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 6 & 1 & 4 & 3 \end{pmatrix} \;=\; (1\;2\;5\;4)\,(3\;6) \;=\; (3\;6)\,(1\;2\;5\;4)\;.$$

(The disjointness of the two sets of positions involved, positions 1, 2, 4 and 5 on the one hand, and positions 3 and 6 on the other, obviously means that the two cycles commute.) (Notice also that this example happens to be a complete rearrangement.)

A 2-cycle is also called a *transposition*, because it just exchanges two objects and leaves the rest alone. Plainly a 2-cycle is its own inverse. Notice that Σ_2 consists of I and just one 2-cycle. It is isomorphic to — from an abstract viewpoint is the same group as — the cyclic group C_2.

Remark: An n-cycle can be decomposed into $(n-1)$ 2-cycles, for instance via

$$\begin{aligned} (p_1\;p_2\;\ldots p_n) &= (p_1\;p_n)\,(p_1\;p_2\;\ldots p_{n-1}) \\ &= (p_1\;p_n)(p_1\;p_{n-1})\ldots(p_1\,p_3)\,(p_1\,p_2)\,. \end{aligned}$$

(Never forget that the p_r denote positions and not objects.) These 2-cycles are **not disjoint**, and we lose uniqueness in general. There are many other such decompositions. Recalling the previous remark, we now see that **any permutation can be decomposed** (though not uniquely) **into 2-cycles or transpositions**.

Definition: A permutation is said to be *odd* or *even* if, respectively, it is a product of an odd or even number of 2-cycles or transpositions.

Example 2.12 *Worked example: Show that oddness and evenness in this sense are well-defined, despite the non-uniqueness.*

Partial answer: This is because one cannot replace a single transposition by two transpositions. Call the two transpositions S and T. There are only three types of possibility: first, that S and T act on the same pair of positions, in which case $S = T$ and $ST = I$, plainly not a transposition; second, that S and T act on disjoint pairs of positions, in which case ST moves 4 objects and is plainly not a transposition; and third, that S and T act on pairs p_1, p_2 and p_2, p_3 with just one position p_2 in common ($p_1 \neq p_3$), in which case $ST = (p_1\, p_2)(p_2\, p_3) = (p_1\, p_2\, p_3)$, which is a 3-cycle and not a transposition.

Remark: n-cycles are even if n is odd, and odd if n is even.

Example 2.13 *Worked example: The alternating group A_N as a subgroup of Σ_N*

Show that the subset A_N of Σ_N consisting of **all even permutations** forms a group, but that the subset consisting of all odd permutations does not. Show that A_N is a *normal* subgroup of Σ_N and that $|A_N| = \frac{1}{2}N!$ (A_N is called the **alternating group** on N letters). Show that the dihedral group D_3 is isomorphic to Σ_3, and that the cyclic group C_3 is isomorphic to A_3. Show that the symmetries of a tetrahedron in 3-dimensional space, including reflections (mirror images), form a group isomorphic to Σ_4 and that the same without reflections (the *rigid rotations* of a tetrahedron) is isomorphic to A_4. Show that Σ_3, Σ_4 and A_4 are all non-abelian.

Answer: First, A_N is closed because each even permutation can be expressed as the product of an even number of transpositions. Two such products, when multiplied together, will give a product consisting of a larger, but still even, number of transpositions. Second, A_N includes the identity I, because $I = T^2$ where T is any transposition. Third, A_N includes inverses because each transposition T is its own inverse, and a product $T_1T_2T_3...T_n$ of transpositions has the inverse $T_n^{-1}T_{n-1}^{-1}T_{n-2}^{-1}...T_1^{-1} = T_nT_{n-1}T_{n-2}...T_1$. A_N is therefore a subgroup of Σ_N. It is a *normal* subgroup because the inverse of an even permutation is even, and similarly for an odd permutation; so conjugation by any element $g \in \Sigma_N$ preserves evenness. (That is, with $h \in A_N$, i.e. h even, we have ghg^{-1} even, regardless of whether or not g is even. **Or**, A_N is the kernel of the homomorphism from Σ_N to $C_2 = \{1, -1\}$ defined by even $\mapsto 1$, odd $\mapsto -1$ (this does preserve products). So there is a quotient group Σ_N/A_N isomorphic to C_2 and therefore of order 2, so $|A_N| = \frac{1}{2}|\Sigma_N| = \frac{1}{2}N!$ (reminding us that there is no room within Σ_N for more than one nontrivial coset.)

(Here's an independent proof of $|A_N| = \frac{1}{2}N!$ — that is, exactly half the elements of Σ_N are even. Let o and e be the number of distinct odd and even permutations respectively (so that $o + e = N!$). If g_1 and g_2 are any two distinct odd permutations, then $(1\,2)g_1$ and $(1\,2)g_2$ are distinct even permutations, distinct because $(1\,2)$ has an inverse (itself). So, extending this to all the odd permutations in Σ_N, we see that to all the o distinct odd permutations there must correspond o distinct even permutations; so $e \geqslant o$. But we can equally well start with even permutations and show that $o \geqslant e$. Therefore $e = o$, as required.)

The isomorphisms between D_3 and Σ_3, and between C_3 and A_3, follow from the facts that rotating a triangle in 3-dimensional space can permute its three vertices in any manner whatever, and rotating a triangle in 2-dimensional space permutes its three vertices cyclically — remember that n-cycles are even if n is odd, as with the 3-cycle (1 2 3).

The isomorphisms between the tetrahedron symmetries and Σ_4 and A_4 can be shown by constructing the group tables, or again from considering the geometry, as follows.

If we start with the vertices coincident with fixed spatial positions labelled 1, 2, 3, 4, then the even permutations are (1 2)(3 4) and two similar permutations, each of which rotates the tetrahedron by 180° about the perpendiculars bisecting opposite pairs of edges, together with (1 2 3), (1 3 2) and three similar pairs, eight 3-cycles in all, which rotate the triangular faces of the tetrahedron in their own planes about their centroids by 120° and 240°. This exhausts A_4 since $|A_4| = \frac{1}{2}4! = 12 = 1+3+8$. And if mirror images are allowed, then each of the foregoing even permutations acquires a distinct mirror-image counterpart, which is odd. For instance, taking the reflection plane lying in one edge, and perpendicular to the edge opposite, exchanges two vertices while leaving the other two alone. This is a transposition and therefore an odd permutation. So every odd permutation in Σ_4 produces a mirror image of the tetrahedron rotated somehow.

(These symmetries are important to understanding the small oscillations and other properties of a methane molecule, CH_4.)

Σ_3 is non-abelian since D_3 is non-abelian. Since Σ_3 is a subgroup of Σ_N for $N > 3$ (remark on page 51), Σ_4 is non-abelian. A_4 is non-abelian since, for example, (1 2 3)(1 3 4) = (2 3 4) $\neq$ (1 3 4)(1 2 3) = (1 2 4). ⌟

Definition: The **cycle shape** of an element of Σ_N is the set of numbers $\{n_2, n_3, \ldots\}$ specifying the number of 2-cycles, 3-cycles, ... in the unique decomposition of that element into disjoint cycles. In simple cases such as $\{2, 1, 0, 0, \ldots\}$ we can use the alternative notation $(\cdot\,\cdot)(\cdot\,\cdot)(\cdot\,\cdot\,\cdot)$ to show the cycle shape. This means the same thing as $(\cdot\,\cdot\,\cdot)(\cdot\,\cdot)(\cdot\,\cdot)$, or $(\cdot\,\cdot)(\cdot\,\cdot\,\cdot)(\cdot\,\cdot)$, because the decomposition is unique and disjoint.

Thus, for instance, the example in the display (2.7) above has cycle shape $\{0, 1, 0, 0, \ldots\}$, equivalently $(\cdot\,\cdot\,\cdot)$, since it is simply a 3-cycle. The example (3 6) (1 2 5 4) on p. 55 has cycle shape $\{1, 0, 1, 0, \ldots\}$, equivalently $(\cdot\,\cdot)(\cdot\,\cdot\,\cdot\,\cdot)$.

Warning: the decomposition into 2-cycles is irrelevant here, because in that case the 2-cycles are neither disjoint nor unique.

Remark: Within Σ_N (though not within A_N), the conjugacy classes are exactly the cycle shapes.

The proof is easy, and left as an exercise. Essentially, within Σ_N all possible permutations are available as conjugating elements for use in (2.1). So any permutation with a given cycle shape applies to any set of elements of Σ_N displayed in any order. That is, conjugation allows us to substitute any elements whatever, for the dots in any of the cycle-shape patterns such as $(\cdot\,\cdot)(\cdot\,\cdot\,\cdot)$.

Now do Sheet 2 q. 3. Note also that D_3 also illustrates the point, because it's isomorphic to Σ_3. The cycle shapes are $(\cdot\,\cdot\,\cdot)$ for the conjugacy class $\{R,\, R^2\}$, and $(\cdot\,\cdot)$ for the conjugacy class $\{m_1,\, m_2,\, m_3\}$.

A_4 is interesting for another reason, besides tetrahedron rotations. It's the smallest group showing that the converse of Lagrange's theorem is false. We have $|A_4| = \frac{1}{2}4! = 12$ but there's no subgroup of order 6. The subgroups of A_4 are all of order 2, 3 or 4. In fact they consist of the identity with $\{(123), (132)\}$, $\{(124), (142)\}$, $\{(134), (143)\}$, and $\{(234), (243)\}$ — i.e., the four copies of C_3 corresponding to the $\pm 120^\circ$ in-plane rotations of the four faces of the tetrahedron — together with the three flips $\{(12)(34)\}$, $\{(13)(24)\}$, and $\{(14)(23)\}$. So there are three copies of C_2 making up one copy of the Vierergruppe. This last is the only normal subgroup of A_4. The nontrivial conjugacy classes consist of the three flips, the four $+120^\circ$ rotations, and the four -120° rotations. The flips can interchange the 120° rotation axes but can't convert between clockwise and anticlockwise.

Definition: If a group G has two normal subgroups H and J, then we say that G is the **direct product** of H and J provided that

1. G is generated by the elements of H and J alone, and
2. H and J are not only normal, but are also disjoint apart from the identity.

(That is, H and J share no element other than the identity.)

Notation: the direct product is usually written $G = H \times J = J \times H$.

Remark: If G is the direct product of H and J, then each element of H commutes with each element of J.

Proof: If $h \in H$ and $j \in J$, then $hj(jh)^{-1} = hjh^{-1}j^{-1} = (hjh^{-1})j^{-1}$. But this belongs to J because J is normal. Similarly, $hjh^{-1}j^{-1}$ belongs to H because H is normal. So by condition 2 above, $hj(jh)^{-1}$ must be the identity. So $hj = jh$.

Example: The Vierergruppe is the direct product of any two of its subgroups, so has the structure $C_2 \times C_2$. By contrast, C_4 does *not* have that structure; recall Example 2.7, and note that C_4 has only one distinct, albeit normal, C_2 subgroup.

Remark: If $G = H \times J$ then the quotient groups G/H and G/J are isomorphic to J and H respectively. (Both quotient groups exist because of the fact that H and J are both normal in G, as is implied as soon as we write $G = H \times J$.)

Proof: Consider for instance G/H. By condition 1, we have $G/H = \{jH | j \in J\}$. The mapping from J to G/H defined by $j \mapsto jH$ is a homomorphism, because H is normal, $\Rightarrow j_1 j_2 H = (j_1 H)(j_2 H)$. To prove that the mapping is 1–1, and therefore an isomorphism, it's enough to prove that for any $j_1, j_2 \in J$ such that $j_1 \neq j_2$ we must have $j_1 H \neq j_2 H$, i.e., the cosets $j_1 H$ and $j_2 H$ are distinct. Suppose, on the contrary, that $j_1 H = j_2 H$. Then we would have $j_2^{-1} j_1 H = H$, implying that $j_2^{-1} j_1 \in H$. But $j_2^{-1} j_1 \in J$. So condition 2 now implies that $j_2^{-1} j_1 = I$ and therefore that $j_1 = j_2$, which is a contradiction.

There are useful lists of small abelian and non-abelian groups and their properties, including their direct-product decompositions if any, at

`http://en.wikipedia.org/wiki/List_of_small_groups`

— note that Σ_N is denoted there by S_N, and the Vierergruppe by Dih_2.

Chapter 3

3. Representation theory

This chapter contains the third and final set of syllabus items:

Representation of groups; reducible and irreducible representations; basic theorems of representation theory.

Classes and characters. Examples of character tables of point groups. *Applications in Molecular Physics*.

Remember, 'point groups' means finite symmetry groups, such as D_3 and D_4, that leave a point fixed in space. (Groups that leave no point fixed, such as translation groups, are also of interest especially in crystallography, but are not treated here.) Some detailed applications to molecules are included for completeness but, as indicated by the asterisks, are non-examinable.

3.1 Motivation

The representation of finite symmetry groups by nonsingular matrices has important and well-established applications in chemistry and physics. Examples are the small oscillations of molecules with symmetries, such as CO_2, H_2O, NH_3, CH_4, etc., especially when it comes to quantum rather than classical models. Other examples include the behaviour of crystal lattices.

With simple small-oscillation problems such as the equilateral triangle problem of Example 1.5, the symmetry considerations are intuitively obvious. However, representation theory comes into its own when we are confronted by the more complicated problems involved in building models of things in

the real world. These include crystal lattices as well as real molecules.[1] For many of these problems, intuition is liable to fail.

Representation theory is the tool that relates the intrinsically non-linear algebra of groups to linear problems, like classical small oscillations, linear wave equations, and the Schrödinger equation in quantum mechanics. One must determine how the group acts on the vector space underlying the linear problem. If the vector space is finite dimensional, the action is by matrices.

Representation theory investigates systematically the possible forms that can be taken by the matrices representing a given group. We already know that some finite symmetry groups can be represented by matrices. This was evident from the examples, which brought in the concept of the **underlying vector space**. We saw on page 40 that reorienting or reflecting objects in physical space can be thought of as linear operations on arbitrary vectors $\mathbf{x}$ within the vector spaces $\mathbb{R}^2$ or $\mathbb{R}^3$.

But the real power of representation theory lies in two facts. The first is that, as will be proved next, *any finite group whatever* can always be represented by matrices.[2] The second is that the underlying vector space can be any vector space in *any number of dimensions, with real or complex coordinates*.

An example is the real vector space $\{\mathbf{Q}^{(m)}\}$ spanned by the generalized eigenvectors of a classical small-oscillations problem. So representation theory is a natural tool for discerning the qualitative features of such problems that arise from symmetries. Most importantly, symmetries can give rise to repeated eigenvalues, or 'degeneracy', as it is called — i.e., to sets of normal modes sharing the same frequency. Representation theory shows how to find these in a completely general way. Representation theory doesn't directly determine the eigenvalues themselves, although it often simplifies their calculation.

[1] For instance CO_2, H_2O, and CH_4 are all important greenhouse gases in the atmosphere. Unlike O_2 and N_2, they are not transparent to infrared radiation. This is because of the way some of their small oscillations interact with infrared photons.

[2] Not all infinite groups can be so represented, though many can. Important examples that can are SO(3), the group of continuous rigid rotations in three dimensions — that is, rotations through any angle about any axis — and SU(2), the group of continuous unitary transformations with determinant 1, or 'complex rigid rotations' in two complex dimensions. SU(2) arises in elementary quantum mechanics, for spin-$\frac{1}{2}$ particles such as electrons. It has a quotient group $SU(2)/C_2$ isomorphic to SO(3).

It turns out that finding the possible forms of the matrices is closely related to finding the **invariant subspaces** of the underlying vector space, where, as always, 'invariant' means invariant under group actions.

3.2 The regular representation

We next prove that any finite group whatever can always be represented by matrices. The idea is to use the 'complete rearrangement' property of group tables.

The proof is by construction. Suppose we are given the group table for any finite group G. Let $|G| = N$. Consider the $N \times N$ unit matrix $\mathbf{I}$. From $\mathbf{I}$ we may form a set of $(N-1)$ other matrices, simply by subjecting its rows to a set of complete rearrangements corresponding to those in the group table. Denote by $\mathbf{D}(g)$ the rearranged matrix corresponding to the row in the table that starts with the group element g^{-1}. The resulting matrices $\mathbf{D}(g)$ provide a faithful representation of G called the **regular representation**. How it works is easiest to see from an example.

Example 3.1 *Worked example: Write down the regular representation of D_3 and Σ_3.*
Answer: In the group table for D_3 seen on page 40 the first row is I, R, R^2, m_1, m_2, m_3. The row starting with the group element $R^{-1} = R^2$ is the third row, R^2, I, R, m_2, m_3, m_1. We can think of this third row as resulting from a complete rearrangement of the first row; so for instance the first element of the first row becomes the second element of the third row. The rule for constructing the matrix $\mathbf{D}(R)$ says that, correspondingly, the first row of $\mathbf{I}$ becomes the second row of $\mathbf{D}(R)$. Similarly, the second row of $\mathbf{I}$ becomes the third row of $\mathbf{D}(R)$, the third row of $\mathbf{I}$ becomes the first row of $\mathbf{D}(R)$, the fourth row of $\mathbf{I}$ becomes the last row of $\mathbf{D}(R)$, and so on. After doing the same for the other group elements, reading off the rearrangements from the second, fourth, fifth and sixth rows of the table on page 40, we have, with dots in place of zeros for legibility,

$$\mathbf{D}(I)=\begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}, \quad \mathbf{D}(R)=\begin{pmatrix} \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \end{pmatrix}, \quad \mathbf{D}(R^2)=\begin{pmatrix} \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \end{pmatrix},$$

$$\mathbf{D}(m_1)=\begin{pmatrix} \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \end{pmatrix}, \mathbf{D}(m_2)=\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \mathbf{D}(m_3)=\begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

It is easy to check that the entire group table is reproduced by multiplication among these matrices. That is, the correspondence between the matrices and the original group elements is an isomorphism. So the matrices provide a faithful representation of D_3 and therefore of Σ_3, which we already know is isomorphic to D_3. ⌟

Remark: The traces all vanish except for $\mathrm{Tr}\,(\mathbf{D}(I)) = \mathrm{Tr}\,(\mathbf{I}) = 6$. The determinants are ± 1 with the first three positive and the last three negative. This is a consequence of the evenness of R, R^2 regarded as elements of Σ_3, i.e. as the 3-cycles (1 2 3) and (3 2 1), and the oddness of m_1, m_2, m_3 regarded similarly as the 2-cycles (2 3), (3 1) and (1 2).

Remark: We have now seen *three* different faithful matrix representations of D_3 and Σ_3 — of sizes 2×2, 3×3, and 6×6. The first two appeared in q. 1 of Sheet 2. Make sure you've done that question.

Remark: The construction of $\mathbf{D}(g)$ in the regular representation is read off from the group-table row starting with g^{-1}, rather than g. In case this seems surprising, the reason is essentially the distinction noted in Section 2.6 between rearranging *positions* and rearranging *objects*. A general proof that the construction works is as follows.

Proof (non-examinable) that the regular representation faithfully represents any G: With $|G| = N$, consider the first row of the group table, consisting of the elements of G laid out in some definite order beginning with $I = g_1$, say. Consider also the row of the table that results from multiplying the first row by some element $g^{-1} \neq I$. The two rows are

$$G \;=\; \{g_1,\, g_2, ...,\, g_N\} \qquad \text{and} \qquad g^{-1}G \;=\; \{g_{p_g(1)},\, g_{p_g(2)}, ...,\, g_{p_g(N)}\}\,, \tag{3.1}$$

where an integer-valued function $p_g(i)$ is introduced in order to describe the rearrangement of the group elements. By definition, the group element originally located at the $p_g(i)^{\text{th}}$ position in the first row of the table migrates to the i^{th} position in the row $g^{-1}G$. In particular, since $g_1 = I$ we have $g_{p_g(1)} = g^{-1}$. Completeness of the rearrangement implies that $p_g(i) \neq i$ for all i. Also, as with any other permutation the function $p_g(i)$ is invertible, with inverse function $p_g^{-1}(i)$ say.

The elements of $\mathbf{D}(g)$ can now be specified using the Kronecker delta. The unit matrix $\mathbf{I}$ has entries $I_{ij} = \delta_{ij}$, so applying the complete rearrangement to its rows produces

$$D_{ij}(g) \;=\; \delta_{\,p_g(i)\;j} \;=\; \delta_{\,i\;p_g^{-1}(j)}\,, \tag{3.2}$$

which takes the value 1 when $p_g(i) = j$, equivalently when $i = p_g^{-1}(j)$, and 0 otherwise. Thus defined, the matrices multiply exactly as required in order for them to provide a faithful representation of G. For we have

$$D_{ij}(g)D_{jk}(h) \;=\; \delta_{\,p_g(i)\;j}\;\delta_{\,j\;p_h^{-1}(k)} \;=\; \delta_{\,p_g(i)\;p_h^{-1}(k)} \;=\; \delta_{\,p_h\{p_g(i)\}\;k} \;=\; D_{ik}(gh)\,,$$

where the summation convention applies with indices running from 1 to N. Remember that $(gh)^{-1} = h^{-1}g^{-1}$ and note that, in terms of rows in the group table labelled by their

first element, $h^{-1}g^{-1}\{g\text{-row}\} = \{h^{-1}\text{-row}\}$, i.e., $h^{-1}g^{-1}\{g_{p_g^{-1}(i)}\}_{i=1,..N} = \{g_{p_h(i)}\}_{i=1,..N}$ so that, replacing i by $p_g(i)$ — which merely rearranges the ordering of the elements on both sides — we finally have $h^{-1}g^{-1}\{I\text{-row}\} = \{g_{p_h\{p_g(i)\}}\}_{i=1,..N}$ as required. ⌟

Of course the regular representation may well consist of very large matrices indeed, for instance $53! \times 53!$ for Σ_{53}, the group of permutations of a pack of cards.[3]

Remark: From the proof just given, the trace $\mathrm{Tr}\,\mathbf{D}(g) = D_{ii}(g) = 0$ unless $g = I$, in which case we have $\mathrm{Tr}\,\mathbf{D}(I) = D_{ii}(I)) = |G| = N$, when $\mathbf{D}(g)$ is the regular representation of G.

(Notice that the traces exhibit *different* patterns in the other faithful representations that we have seen. For instance, the 3-dimensional representation in q. 1 of Sheet 2 has traces 3, 1, 1, 1, 0, 0 (taking $\mathbf{I}$ as the first matrix).

Remark: For the regular representation, for any $g \in G$ the matrix product of $\mathbf{D}(g)$ with itself, repeated N times, must give the unit matrix: $\mathbf{D}(g)^N = \mathbf{I}$. That fact would hardly be obvious by inspection. But it follows immediately from the isomorphism $g \mapsto \mathbf{D}(g)$ and the fact that the N^{th} power $g^N = I$ for every g in G. This in turn restricts the determinants $|\mathbf{D}(g)|$ of the regular representation. For groups of odd order, all the determinants $|\mathbf{D}(g)| = 1$. For groups of even order, $|\mathbf{D}(g)| = \pm 1$, with sign according to the evenness or oddness of the group-table rearrangements.

[3] Another example, of somewhat smaller order than $|\Sigma_{53}|$ but theoretically intriguing, and famous among group theorists, is the 'Monster Group' discovered in 1980. The 'Monster Group', also known as the 'Fischer-Griess Monster, or the 'Friendly Giant', has order just under 10^{54}. The order is exactly $2^{46}.3^{20}.5^9.7^6.11^2.13^3.17.19.23.29.31.41.47.59.71$ $= 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000$. So, for the Monster Group, the 'very large matrices indeed' that make up its regular representation are 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000 × 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000 matrices. It is known that the Monster has a smaller faithful representation, whose matrices have entries that are complex numbers. These matrices are only $196,883 \times 196,883$. (If one replaces the complex numbers by the elements of the "order-2 field", namely 0 or 1 (mod 2), then one can obtain a $196,882 \times 196,882$ faithful representation.) The Monster Group is *simple*, where 'simple' has a technical meaning (!) — namely, that it has no nontrivial proper normal subgroups. In fact, it is the largest of the *sporadic* finite simple groups, i.e. of the finite simple groups that do not belong to an infinite family (such as the families D_N or Σ_N).

3.3 Equivalence and inequivalence

It is one thing to know that any finite group can be represented by a set of matrices. It is another to find a set that's efficient, in the sense of being no bigger than necessary, and to make practical use of it. The key to doing this, in a systematic way, is to consider how a matrix representation changes under **similarity transformations**. These correspond to changes of basis in the underlying vector space.

Definition: Two sets of matrices $\{\mathbf{D}\}$, $\{\mathbf{E}\}$ are called **equivalent** if they are in 1–1 correspondence under a single similarity transformation, i.e. if

$$\mathbf{D} \;=\; \mathbf{S}\mathbf{E}\mathbf{S}^{-1} \tag{3.3}$$

for a *single* invertible matrix $\mathbf{S}$ and for *all* the matrices $\mathbf{D}$ and $\mathbf{E}$. If no such $\mathbf{S}$ exists, then the sets $\{\mathbf{D}\}$ and $\{\mathbf{E}\}$ are called **inequivalent**.

Remark: Equivalence $\Rightarrow$ isomorphism. That is, if two equivalent sets of matrices faithfully represent groups, then the groups are isomorphic: effectively they are the same group. A similarity transformation cannot change the group table. (For $\mathbf{D}_l\mathbf{D}_m = \mathbf{D}_n \;\Leftrightarrow\; \mathbf{S}\mathbf{D}_l\mathbf{S}^{-1}\mathbf{S}\mathbf{D}_m\mathbf{S}^{-1} = \mathbf{S}\mathbf{D}_n\mathbf{S}^{-1}$.) But the converse is false: it's possible to have inequivalent sets of matrices faithfully representing the same group, as will be seen in Example 3.4 below. First, two examples of equivalence:

Example 3.2 *Two faithful representations of the Vierergruppe*
These are the two examples on p. 40; we give them new names,

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{D}_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D}_4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{E}_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{E}_3 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{E}_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

These two sets are **equivalent** because they are related by (3.3) with, for example,

$$\mathbf{S} \;=\; \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \qquad \mathbf{S}^{-1} \;=\; \tfrac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

(Of course it's easiest to check that $\mathbf{DS} = \mathbf{SE}$ for each pair of matrices. Notice that it makes no difference if we multiply $\mathbf{S}$ by any number.)

We could have anticipated this result from the remark about invariant subspaces on page 40. The similarity transformation corresponds to rotating the axes by 45° (also rescaling, which is irrelevant) to bring them parallel to the invariant subspaces of $\{\mathbf{D}\}$ or of $\{\mathbf{E}\}$, as the case may be. ⌟

Example 3.3 *Two faithful representations of C_4:*

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{D}_3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{D}_4 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{E}_2 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \mathbf{E}_3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{E}_4 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

These two sets are equivalent because they are related by (3.3) with, for example,

$$\mathbf{S} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, \qquad \mathbf{S}^{-1} = \tfrac{1}{2}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

Notice how **complex numbers** enter naturally, and necessarily. In fact, the most general possible choice can easily be shown to be

$$\mathbf{S} = \begin{pmatrix} a & b \\ ia & -ib \end{pmatrix}$$

for any nonzero numbers a, b. Notice that the two sets are both faithful representations of C_4. Notice also that the determinant $|\mathbf{S}| = -2iab \neq 0$. ⌟

Example 3.4 *C_4 again:*

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{D}_3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{D}_4 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{E}_2 = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}, \quad \mathbf{E}_3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathbf{E}_4 = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}$$

both faithfully represent C_4. But these two sets are **inequivalent**. This is easy to show by trying to find

$$\mathbf{S} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}; \tag{3.4}$$

it is a simple exercise to show that no choice of the numbers a, b, c, d will suffice. One can make $\mathbf{DS} = \mathbf{SE}$ for each pair of matrices, but there is no choice of a, b, c, d that gives a nonzero determinant $|\mathbf{S}| \neq 0$. (To show this it is enough to calculate $\mathbf{D}_2\mathbf{S}$ and $\mathbf{SE}_2$.) ⌟

Example 3.5 *The quaternion group Q*

Q is an order-8 group having some features in common with D_4, though not all. Like D_4 it can be faithfully represented by 2×2 matrices, though with complex entries. The elements of Q can be denoted $\{\pm I, \pm\mathcal{I}, \pm\mathcal{J}, \pm\mathcal{K}\}$ where $\mathbf{I}$ is the unit matrix and where $\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$ can be defined by

$$\mathcal{I} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathcal{J} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \mathcal{K} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

which are related to the Pauli spin matrices.[4] You should check (a) that Q has three order-4 subgroups each isomorphic to C_4 (whereas D_4 has only one such), (b) that Q has only one order-2 subgroup (whereas D_4 has five), and (c) that all the subgroups of Q are normal (whereas only the order-4 subgroups of D_4 are normal).

The neatest way to verify all this is to show that I, $\mathcal{I}$, $\mathcal{J}$, $\mathcal{K}$ satisfy Hamilton's relations $\mathcal{I}^2 = \mathcal{J}^2 = \mathcal{K}^2 = \mathcal{I}\mathcal{J}\mathcal{K} = -I$ (with matrix multiplication understood, as usual). From these one may show further that $\mathcal{I}\mathcal{J} = \mathcal{K} = -\mathcal{J}\mathcal{I}$ and three similar relations. Like D_4, Q is not abelian!

The order-2 subgroup of Q is $\{\pm I\}$, with cosets $\{\pm\mathcal{I}\}$, $\{\pm\mathcal{J}\}$, and $\{\pm\mathcal{K}\}$ (left *and* right). An order-4 subgroup is $\{\pm I, \pm\mathcal{I}\}$ (isomorphic to C_4) with coset $\{\pm\mathcal{J}, \pm\mathcal{K}\}$. Similarly, we have a subgroup $\{\pm I, \pm\mathcal{J}\}$ with coset $\{\pm\mathcal{K}, \pm\mathcal{I}\}$, and a subgroup $\{\pm I, \pm\mathcal{K}\}$ with coset $\{\pm\mathcal{I}, \pm\mathcal{J}\}$.

Remark: It is easy to show, by seeking a similarity transformation as in (3.4), that these eight matrices are **inequivalent** to the eight matrices representing D_4 on page 35. Indeed, the result follows, with no need for calculation, simply from the fact that Q and D_4 are not isomorphic and therefore have different group tables. (As noted above, a similarity transformation cannot change the group table.)

It is crucial to remember that equivalence always refers to the complete set of matrices, not just to one or to any subset of them. ⌟

Remark: Generalizing the notation $U(1)$ above, in Example 2.8, we use the notation $U(n)$ to denote the group of unitary $n \times n$ matrices. Recall that a unitary matrix $\mathbf{D}$ has complex entries and that its inverse is its transposed complex conjugate, implying that the determinant $|\mathbf{D}| = \pm 1$. Notice that $I, \mathcal{I}, \mathcal{J}, \mathcal{K} \in U(2)$.

[4] In terms of the standard notation σ_1, σ_2, σ_3 for the spin matrices (e.g. Jones chapter 8, p. 140) we have $\mathcal{I} = i\sigma_2$, $\mathcal{J} = i\sigma_1$, and $\mathcal{K} = i\sigma_3$. (Remember, footnotes aren't examinable.)

Remark: $U(n)$ is a subgroup of the group $GL(n, \mathbb{C})$ of *all* nonsingular $n \times n$ matrices with complex entries.

(Make sure you check that $GL(n, \mathbb{C})$ and $U(n)$ are indeed groups.)

The letters GL stand for 'General Linear', referring to the way that the matrices act on vectors as linear transformations, within the underlying vector space which is of course $\mathbb{C}^n$, **the *n*-dimensional vector space with complex numbers as scalars.** This means that matrix entries and vector components are all complex numbers.

$GL(n, \mathbb{C})$ is the basic scaffolding for building representation theory. As Example 3.3 on page 66 demonstrates, **complex scalars arise naturally and inevitably**, even though there are particular cases where attention can be confined to real scalars. (Then we sometimes consider $GL(n, \mathbb{R})$, the real counterpart of $GL(n, \mathbb{C})$, and a subgroup of $GL(n, \mathbb{C})$. So then we have real scalars, real matrix entries and real vector components.)

3.4 The character of a representation

Definition: The **character** of a representation is the **set of traces** of its matrices.

The traces can be real or complex numbers. They form important patterns, or fingerprints, both of group structure and of particular representions of that structure. By convention, the first element of a character is taken from the unit matrix. Then one can read off the dimension of the representation as the first element of its character. The characters of the representations encountered so far in this course are as follows (recall that $D_3 \sim \Sigma_3$):

Regular representation of any finite group G (page 63):	$\{\|G\|, 0, 0, 0, 0, 0, 0, 0, \ldots\ 0\}$
Representations of D_3 and Σ_3 in Sheet 2 q. 1:	$\{3, 1, 1, 1, 0, 0\}$
Representation of D_3 and Σ_3 in Example 2.5	$\{2, -1, -1, 0, 0, 0\}$
Representation of D_4 in Example 2.2	$\{2, 0, -2, 0, 0, 0, 0, 0\}$
Representation of Q in Example 3.5:	$\{2, -2, 0, 0, 0, 0, 0, 0\}$
Rep'ns of the Vierergruppe in Example 3.2, **D**'s *and* **E**'s:	$\{2, -2, 0, 0\}$
Representations of C_4 in Example 3.3, **D**'s *and* **E**'s:	$\{2, 0, -2, 0\}$
Representation of C_4 in Example 3.4, **D**'s only:	$\{2, 0, -2, 0\}$
Representation of C_4 in Example 3.4, **E**'s only:	$\{2, 2i, -2, -2i\}$

Notice that the elements of each character *add to zero*, except in the first two examples. The reasons for this will be discussed in the next section.

Notice also that the characters are the same for all the pairs of equivalent representions, and different for inequivalent representions. For instance the characters differ for D_4 and Q, and for the **D**'s and **E**'s in Example 3.4. This is not accidental. We will note next

that characters are invariant under similarity transformations — one reason why they're important fingerprints. (That invariance was illustrated in Sheet 2 q. 1.)

Remark: Characters are **invariant under similarity transformations**.

This is because $\mathrm{Tr}\,(\mathbf{ABC}) = A_{ij}B_{jk}C_{ki} = \mathrm{Tr}\,(\mathbf{BCA})$, the cyclic property of traces, implies that

$$\mathrm{Tr}\,(\mathbf{S}\,\mathbf{D}\,\mathbf{S}^{-1}) \;=\; \mathrm{Tr}\,(\mathbf{D}\,\mathbf{S}^{-1}\,\mathbf{S}) \;=\; \mathrm{Tr}\,(\mathbf{D})\ . \tag{3.5}$$

Here's another reason:

Remark: Characters are **invariant within a conjugacy class**.

For if two elements are conjugate — recall the definition (2.1) — then in any representation their matrices, $\mathbf{D}_1$ and $\mathbf{D}_2$, say, must be related by

$$\mathbf{D}_2 \;=\; \mathbf{D}\mathbf{D}_1\mathbf{D}^{-1} \tag{3.6}$$

where $\mathbf{D}$ is some other matrix in the representation. But the cyclic property of traces then implies, in exactly the same way as above, that

$$\mathrm{Tr}\,(\mathbf{D}_2) \;=\; \mathrm{Tr}\,(\mathbf{D}\,\mathbf{D}_1\,\mathbf{D}^{-1}) \;=\; \mathrm{Tr}\,(\mathbf{D}_1\,\mathbf{D}^{-1}\,\mathbf{D}) \;=\; \mathrm{Tr}\,(\mathbf{D}_1)\ . \tag{3.7}$$

3.5 Reducibility and irreducibility

Examples 3.2 and 3.3 illustrate **reducibility**.[5] In 2 dimensions, reducibility means simply that **all** the matrices in a representation can be diagonalized by a **single** similarity transformation. Thus, in Example 3.3 the $\mathbf{D}$ matrices get reduced to the purely diagonal $\mathbf{E}$ matrices, and in Example 3.2 it's the $\mathbf{E}$ matrices that are reduced to the purely diagonal $\mathbf{D}$ matrices.

Such reduction is far more than the ordinary diagonalization of a single matrix. Again the idea is to consider all the matrices together, because again we want insight into the structure of the group they represent. And remember, yet again, that the (single) similarity transformation can't change that structure because it can't change the group multiplication table:

$$\mathbf{D}_l\mathbf{D}_m = \mathbf{D}_n \quad\Rightarrow\quad \mathbf{S}\mathbf{D}_l\mathbf{S}^{-1}\mathbf{S}\mathbf{D}_m\mathbf{S}^{-1} = \mathbf{S}\mathbf{D}_n\mathbf{S}^{-1}\ . \tag{3.8}$$

(That is, equivalent representations are isomorphic.)

[5] In some texts reducibility, as defined here, is called 'complete reducibility'.

A representation that cannot be reduced to smaller blocks is called irreducible. The irreducible representations are the building blocks of all the representations of a given group, and for a finite group there are only a finite number of inequivalent irreducible representations. Irreducible representations are often called IR's or **irreps** for short.

Remark: A 2-dimensional **non-abelian** representation is irreducible.

For if it were reducible, then its group table would be the same as when the matrices are all diagonal. And diagonal matrices commute with each other, which would be a contradiction. So our 2-dimensional matrix representations of D_3, D_4 and Q **are all irreducible**.

What does reducibility mean when the dimensionality is 3 or greater? By definition it means *block-diagonalization*, understood to include elementwise diagonalization:

Definition: For any number of dimensions, **reducibility** means reducibility by a similarity transformation **to block-diagonal form** such that every nonzero matrix element is confined within the same block-diagonal shape.

The 3-dimensional example of Sheet 2 q. 1 illustrates this. The last set of matrices (third display on the example sheet) have their nonzero elements confined within a single block-diagonal shape, which in this case has 1×1 matrices at upper left and 2×2 matrices at lower right. In this example, we know that further reduction to three 1×1 matrices on the diagonal, i.e., elementwise diagonalization, is not possible. By the Remark above, the 2×2 matrices at lower right are irreducible because the group is non-abelian.

Notice that the elements of each character **add to zero** for each 2-dimensional representation listed on page 68. In particular, they add to zero for the 2-dimensional representations of D_3, D_4 and Q, which are all irreducible as already remarked. *This turns out to be a general property of irreducible representations of any dimension* — indeed, a special case of what I'll call their **'gridlock orthogonality properties'**, §§3.8ff.

By contrast, the character of the regular representation has elements that never add up to zero. We saw that they add up to $|G|$. That will allow us to prove that the regular representation must always be reducible. Also reducible is the 3-dimensional case of Sheet 2 q. 1, whose character is {3, 1, 1, 1, 0, 0}. And, sure enough, the elements add up to a nonzero value, 6.

Remark: **Reducibility means that the underlying vector space has invariant subspaces.**

If the underlying vector space is 2-dimensional, and if the 2×2 matrices of a representation are all in diagonal form, then plainly they leave invariant just two 1-dimensional subspaces. These are the two subspaces spanned by column vectors $(1,\, 0)^{\mathsf{T}}$, $(0,\, 1)^{\mathsf{T}}$ — multiplied, of course, by any scalar we please, which includes **complex scalars** when we are working within $GL(n, \mathbb{C})$.

In the case of Example 3.2, those subspaces are real 1-dimensional subspaces of the vector space $\mathbb{R}^2$. In the case of Example 3.3, the invariant subspaces of the $\mathbf{D}$'s are complex 1-dimensional subspaces of the vector space $\mathbb{C}^2$. These particular subspaces happen to be invisible within $\mathbb{R}^2$: we can't simultaneously diagonalize the $\mathbf{D}$'s by real matrices.

If the underlying vector space has more than 2 dimensions, then plainly it can have invariant subspaces of more than 1 dimension. The dimensions of such subspaces match the sizes of the matrices on the diagonal after the matrices have been reduced, i.e., put into block-diagonal form.

Remark: In Examples 3.2 and 3.3, the invariant subspaces are mutually orthogonal *provided that* we define orthogonality of two vectors $\mathbf{x}$, $\mathbf{y}$ using a scalar or inner product $(\mathbf{x},\, \mathbf{y})$ of Hermitian type, namely

$$(\mathbf{x},\, \mathbf{y}) \;=\; x_j^*\, y_j \;=\; \mathbf{x}^\dagger\, \mathbf{y}\ , \tag{3.9}$$

with summation convention over the components. The asterisk denotes the complex conjugate and the dagger the transposed complex conjugate.

Orthogonality has, of course, no meaning unless one says with respect to what inner or scalar product. Two vectors are orthogonal, by definition, if their inner product vanishes. You should check that this holds in both the examples mentioned — Example 3.3 as well as 3.2 — provided that one remembers the complex conjugate in (3.9). Without it, the invariant subspaces of the $\mathbf{E}$'s in Example 3.3 won't transform to orthogonal invariant subspaces of the $\mathbf{D}$'s. (Notice that in this example $\mathbf{S} = \sqrt{2}$ times a unitary matrix.)

3.6 The group-invariant inner product $[\mathbf{x},\, \mathbf{y}]$

This is also called the group-invariant scalar product, or just the invariant inner product or invariant scalar product. It is key to some of the most beautiful theorems of representation theory and generalizes the ordinary Hermitian inner product as follows:

Definition: Let $(\mathbf{x}, \mathbf{y})$ be the ordinary Hermitian inner product defined by (3.9). Let the matrices in a representation of a given group G be the set $\{\mathbf{D}(g)\}$. Define

$$[\mathbf{x}, \mathbf{y}] \;=\; \sum_{g\in G} \big(\mathbf{D}(g)\mathbf{x},\, \mathbf{D}(g)\mathbf{y}\big)\;, \tag{3.10}$$

where the round brackets mean the same as in (3.9) and where **the summation is over all the group elements** g.

Remark (the invariance property): For any fixed $h \in G$, we have

$$[\mathbf{D}(h)\mathbf{x}, \mathbf{D}(h)\mathbf{y}] \;=\; [\mathbf{x}, \mathbf{y}]\;. \tag{3.11}$$

This result follows at once from the complete rearrangement property of group tables, after substitution into (3.10). The sum over group elements doesn't care about the order in which the elements are taken. (This device of summing over group elements will come up repeatedly.)

Implication: Any given representation of a given group G **becomes unitary** if the inner product is redefined as in (3.10) (and one changes basis in the underlying vector space to be orthonormal with respect to (3.10) — another similarity transformation, of course). (This construction is sometimes referred to as *Weyl's unitary trick*.)

So *all* finite groups can be seen as groups of **generalized rotations and reflections within a complex vector space**, representable by **unitary matrices**. Generalized rotations and reflections mean group actions, on the underlying vector space, that preserve the **lengths** and mutual orthogonality of vectors.[6] 'Length' can be defined as $(\mathbf{x}, \mathbf{x})^{1/2}$ or $[\mathbf{x}, \mathbf{x}]^{1/2}$. If we take it as $[\mathbf{x}, \mathbf{x}]^{1/2}$ — which has all the standard properties of inner or scalar products — then the whole of representation theory for finite groups can be developed within $U(n)$, with similarity transformations and group actions **all described by unitary matrices**. (Example 3.3 illustrates this.)

(The standard properties are positive definiteness, $[\mathbf{x}, \mathbf{x}] > 0$ when $\mathbf{x} \neq 0$, Hermitian symmetry, $[\mathbf{x}, \mathbf{y}] = [\mathbf{y}, \mathbf{x}]^*$, and linearity $[\mathbf{x}, a\mathbf{y} + b\mathbf{z}] \;=\; a[\mathbf{x}, \mathbf{y}] + b[\mathbf{x}, \mathbf{z}]$, all easily verified from (3.10). When (3.10) is used to define 'length' and 'orthogonality', all the standard derivations of matrix eigenvector-orthogonality, etc., go through in a way that closely parallels the case of real generalized inner products recalled on page 16.)

[6] Mathematicians call these 'isometries', meaning operations that preserve lengths or distances, in any sense that satisfies the triangle inequality.

In particular, when (3.10) is used to define 'length' and 'orthogonality', then *every invariant subspace has an orthogonal complement, and all invariant subspaces can be taken to be mutually orthogonal* — not just particular cases like those of Examples 3.2 and 3.3.

Summary: **Reducibility ⇔ mutually orthogonal invariant subspaces ⇔ reduction possible** as defined here, **to block-diagonal form**.

Of the faithful representations we've seen, it's only the *irreducible* representations whose characters have elements *guaranteed* to add up to zero. The representations in Example 3.4 have the characters $\{2,\, 0, -2,\, 0\}$ and $\{2,\, 2i, -2, -2i\}$, whose elements do both add up to zero. The first does so essentially because it's irreducible within $GL(n, \mathbb{R})$ even though not within $GL(n, \mathbb{C})$. The second does so only because it's made up of two identical copies of a 1-dimensional irreducible representation.

3.7 Unfaithful representations

So far we have dealt only with faithful representations, which by definition are isomorphic to the group G they represent. However, we may also consider sets of matrices, including 1×1 matrices, that are **merely homomorphic** to G. For the moment I'll call these 'unfaithful representations'.

You already know some relevant examples, from q. 4 of Sheet 2. The first is a homomorphism from D_4 to the trivial group consisting of the identity only — let's call it C_1 — represented by a single 1×1 matrix (1), the same thing as the ordinary integer 1:

$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1\}$$

(So the kernel of this homomorphism is the whole of D_4.) Then there are three nontrivial homomorphisms, mapping D_4, in three different ways, to a faithful representation of C_2 consisting of the two 1×1 matrices (1) and (-1):

$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{1,\ 1,\ 1,\ 1,\ -1,\ -1,\ -1,\ -1\}$$
$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{1,\ -1,\ 1,\ -1,\ 1,\ 1,\ -1,\ -1\}$$
$$\{I, R, R^2, R^3, m_1, m_2, m_3, m_4\} \mapsto \{1,\ -1,\ 1,\ -1,\ -1,\ -1,\ 1,\ 1\}$$

You will already have checked that these mappings preserve group multiplication and are therefore homomorphisms. (Their kernels are the three normal subgroups of D_4, namely C_4 and two copies of the Vierergruppe.)

But what's the point in considering representations of C_2, let alone of C_1, as 'representations' of a much bigger group such as D_4? They represent only some aspects of D_4, so we are throwing away information! **How can this be useful?**

It's like a chess gambit.[7] One throws away information but gets back much more in return. The payoff comes from considering **all the irreducible representations together**, faithful and unfaithful. **So by convention they're all called 'representations'**, regardless of how much or how little information each one contains. So, in particular, the trivial representation $\{(1)\}$, which contains no information at all, and faithfully represents only the trivial group C_1, is nevertheless called a 'representation' of, for instance, D_4, or indeed of any finite group whatever.

What do **equivalence** and **irreducibility** mean for 1-dimensional representations? The key is to notice that 1-dimensional similarity transformations are powerless to do anything: $\mathbf{SDS}^{-1} = \mathbf{D}$ if the matrices are 1×1. (In a 1-dimensional vector space there can be no changes of basis beyond rescaling.) It follows that

All 1-dimensional representations are irreducible, and that

Any two distinct 1-dimensional representations are inequivalent.

Remark: The **character** of a 1-dimensional representation is the same thing as the representation itself, because the traces of 1-dimensional matrices are trivially the single entries in such matrices. Therefore, the images of the above homomorphisms are *also*, of course, the characters of these four 1-dimensional representations of D_4.

The representations above are the *only* 1-dimensional representations of D_4. (You can take this on faith, or easily prove it. You need only show that no other choices of ± 1 qualify as homomorphisms, nor do any other sets of numbers, real or complex.)

[7] The great mathematician J. E. Littlewood called it 'impudence' in mathematics. See *Littlewood's Miscellany*, ed. B. Bollobás, 1986, Cambridge University Press, 200 pp.

Similarly, the only 1-dimensional representations of D_3 are

$$\{I, R, R^2, m_1, m_2, m_3\} \mapsto \{1,\ 1,\ 1,\ 1,\ 1,\ 1\} \tag{3.12}$$
$$\{I, R, R^2, m_1, m_2, m_3\} \mapsto \{1,\ 1,\ 1, -1, -1, -1\}\ . \tag{3.13}$$

The trivial representation (3.12) has already been given a sneak preview, as the top left entries in the matrices defining the second 3-dimensional representation of D_3 or Σ_3 in Sheet 2 q. 1.

The general theorems below prove that the complete set of inequivalent irreps for D_4 and D_3 consists of the 1-dimensional irreps just displayed together with the 2-dimensional faithful irreps already found, in Example 2.2 and in Example 2.5, Eqs.(2.2, 2.3). In both cases there are, indeed, **no more inequivalent irreps**.

Associated with the complete set of inequivalent irreps for every finite group are certain patterns implied by the general theorems, the patterns associated with 'gridlock orthogonality'. It's easy to see what these are like, from simple examples. A good way to begin to see the patterns is to summarize the characters of all the irreps in a '**character table**'.

The character tables for D_3 and D_4 are enough to convey the idea. I'll show them, therefore, before discussing the general theorems.

3.8 The character tables for D_3 and D_4

For D_3, we denote the trivial irrep by $d^{(1)}$, meaning the whole set of 1×1 matrices given by (3.12). Similarly, the next 1-dimensional irrep, given by (3.13), is denoted by $d^{(2)}$, and the remaining irrep, the 2×2 faithful irrep given in Example 2.5, is denoted by $d^{(3)}$. The corresponding characters will be denoted by $\chi_{d^{(1)}}$ $\chi_{d^{(2)}}$, and $\chi_{d^{(3)}}$:

Example 3.6 *Character table for D_3*

	I	R	R^2	m_1	m_2	m_3
$\chi_{d^{(1)}}$	1	1	1	1	1	1
$\chi_{d^{(2)}}$	1	1	1	-1	-1	-1
$\chi_{d^{(3)}}$	2	-1	-1	0	0	0

The last row, showing the character $\chi_{d^{(3)}}$, whose elements are the traces of the 2×2 matrices in Eqs.(2.2, 2.3), is laid out in the same order as in those equations and as on pages 68 and 41. The partitioning into conjugacy classes is shown by vertical rules. Notice the invariance of character elements within conjugacy classes, as dictated by (3.7). ⌟

Similar conventions are used for D_4. We have the four 1-dimensional irreps shown on page 73, and will denote them by $d^{(1)}, \ldots, d^{(4)}$. We have the one faithful 2-dimensional irrep shown on page 35 (Example 2.2), and will denote it by $d^{(5)}$. Its traces appear in the bottom row of the table:

Example 3.7 *Character table for D_4*

	I	R^2	R	R^3	m_1	m_2	m_3	m_4
$\chi_{d^{(1)}}$	1	1	1	1	1	1	1	1
$\chi_{d^{(2)}}$	1	1	1	1	−1	−1	−1	−1
$\chi_{d^{(3)}}$	1	1	−1	−1	1	1	−1	−1
$\chi_{d^{(4)}}$	1	1	−1	−1	−1	−1	1	1
$\chi_{d^{(5)}}$	2	−2	0	0	0	0	0	0

⌟

Not only do all the horizontal rows add to zero except the first, as noted earlier, but **every row**, regarded as a $|G|$-dimensional vector, **is orthogonal to every other row**. Moreover, **every column is orthogonal to every other column that belongs to a different conjugacy class**. (Therefore character tables are often shown with the conjugacy classes lumped together into single columns; then every column is orthogonal to every other column, as suggested by the word '**gridlock**'.)

It can be proved that these are completely general properties of the character table of any finite group whatever, provided only that in cases where complex values arise (a) one uses a **unitary representation**, as (3.11) shows we can if we wish, and (b) interprets orthogonality in the **Hermitian sense**.

And yet more is true: the gridlock is even tighter. Not only does it constrain the character table, but also the **individual matrix elements** of the irreps of dimension 2 or greater. Details in the next section. Once again

the key device is **summation over all the group elements**, as in (3.10), taking note of the **complete rearrangement property**.

And from the gridlock come some remarkable constraints on the number and sizes of irreps, as follows. **For any given finite group** G we can prove:

Theorem 1: The number of inequivalent irreps — including the trivial irrep — equals the number of conjugacy classes. (Both the above character tables illustrate that fact.)

Theorem 2: If the dimensions of the inequivalent irreps are denoted by $n_1(=1)$ and $n_2, \ldots, n_N$, then

$$\sum_{\alpha=1}^{N} n_\alpha^2 = |G| \ .$$

Notice how this pins down the number of inequivalent irreps! For D_3, we have

$$1^2 + 1^2 + 2^2 = 6 \, ;$$

and there are indeed no other possibilities. We really have found all the inequivalent irreps, and the character table on page 75 really is complete.

Similarly for D_4, we have

$$1^2 + 1^2 + 1^2 + 1^2 + 2^2 = 8 \, ;$$

so its character table on page 76 is also complete. There can be no more inequivalent irreps of D_4.

For small groups one can often determine the **entire character table** from the foregoing constraints alone, assuming that one has previously determined the irrep or irreps having more than one dimension.[8] The latter can usually be found from geometric intuition as we did for D_3 and D_4.

Corollary of Theorem 2:

Every irrep has dimension $\leqslant \big(|G| - 1\big)^{1/2}$, for $|G| > 1$.

Thus for instance no group of order less than 5 can have a 2-dimensional or higher-dimensional irrep, so no group of order less than 5 can be non-abelian. In fact the smallest

[8] An example is the quaternion group Q (p. 67). Using Theorem 2 we can easily show that its character table must be the same as for D_4. (Note that Q has a homomorphism onto the Vierergruppe whose 2-dimensional matrices are, however, *reducible* as we saw in Example 3.2, so can't enter the character table.) Another nice example is Σ_4 (p. 90).

such is of order 6, namely D_3 and its isomorph Σ_3. (Because 5 is prime, the only order-5 group is C_5.) For Σ_{53}, there can be no irrep with dimension greater than $(53! - 1)^{1/2}$.

We also note for completeness

Theorem 3: The dimension n_α of each irrep divides $|G|$.

The proof of this last theorem is well beyond our present scope.[9] What's usually most important in practice, though — for instance in analysing the **small oscillations of symmetric molecules** — is the use of Theorems 1 and 2 to pin down the nature of the irreps of small groups. It turns out for instance that the **numbers of degenerate normal modes** (distinct modes with the same normal frequency) are just the **dimensions of the irreps**.

The proofs of Theorems 1 and 2 are relatively easy, but too long for a 10-lecture Easter-term course. However, I'll give them in the non-examinable sections below, after showing where gridlock orthogonality comes from.

3.9 The orthogonality theorems for irreps

This is the fundamental theorem from which Theorems 1 and 2 are derived. It depends on the following remark, called **Schur's lemma**:

Remark: Given any finite group G, consider two of its inequivalent irreps with underlying vector spaces V_1 and V_2. Let $f : V_1 \to V_2$ be a linear mapping from V_1 to V_2 that preserves not only all the *vector-space* operations — as any such linear mapping must — but also all the *group* operations. Then

$$f \text{ must be } \textbf{either} \text{ zero } \textbf{or} \text{ a vector-space isomorphism} \tag{3.14}$$

(meaning a 1–1 mapping that preserves all vector-space operations). Furthermore if V_1 and V_2 are isomorphic, then

$$f \text{ must be } \textbf{a scalar multiple of the identity map.} \tag{3.15}$$

Here's a sketch of the proof (non-examinable). Because f preserves group operations, its kernel (the subspace of V_1 that maps to the zero vector in V_2) is a (group-)invariant

[9] The proof depends on recognizing that the underlying vector space can be replaced by a structure known in linear algebra as a 'module over a commutative ring'. This exposes some consequences of the fact that, in finite group representations, the inverse matrices can be obtained by matrix multiplication and that therefore no numerical (arithmetical) divisions are required — contrary to what one might think from the standard formulae for matrix inverses.

subspace. Similarly, its image in V_2 is an invariant subspace of V_2. But **irreducibility** says that there can be no invariant subspaces other than the zero subspace or the whole space, and (3.14) follows.

If V_1 and V_2 are isomorphic, the second case of (3.14), then we can regard them as the same vector space. Then (3.15) follows from the fact that f cannot have an eigenspace other than the whole vector space. Recall that the eigenspace of a linear mapping is the subspace spanned by a degenerate set of eigenvectors, i.e. all with the same eigenvalue. (Recall 'Gram–Schmidt orthogonalization' etc.) But, under our assumptions, such an eigenspace is another **(group-)invariant subspace**. So irreducibility says that it can only be the whole vector space, with only one eigenvalue, the scalar multiple referred to in (3.15). (The theory of eigenspaces is simplest for orthogonal and unitary transformations, and in view of (3.11) this is sufficient, though strictly speaking not needed.) ⌟

These results translate into matrix form as follows. If the linear mapping f is described by some matrix $\mathbf{T}$, then group operations are preserved **if and only if T commutes with the irreps**, in an appropriate sense. More precisely:

Given any finite group G, let two of its inequivalent irreps be denoted by $\mathbf{d}^{(\alpha)}(g)$ and $\mathbf{d}^{(\beta)}(g)$, with dimensions respectively n_α and n_β. Suppose also that $\mathbf{T}$ is an $n_\alpha \times n_\beta$ matrix independent of $g \in G$ and such that

$$d^{(\alpha)}_{ij}(g)\, T_{jl} \;=\; T_{ik}\, d^{(\beta)}_{kl}(g) \qquad \text{for all } \; g \in G\,.$$

(No other assumption is made about $\mathbf{T}$; for instance its elements could all be zero.) Then **Schur's lemma** says that

$$\mathbf{T} = \mathbf{0} \qquad \text{if } \; \alpha \neq \beta \;\; \text{(i.e. different irreps)} \tag{3.16}$$

and

$$\mathbf{T} = \lambda\, \mathbf{I} \qquad \text{if } \; \alpha = \beta \;\; \text{(i.e. same irrep, \& } \therefore \text{ same dim.)} \tag{3.17}$$

for some scalar number λ, real or complex.
These two remarks are sometimes called Schur's first and second lemmas, in some order. From them we can derive the fundamental orthogonality theorem, which I'll label 'Theorem 0' because it's the foundation-stone of the whole 'gridlock edifice':

Theorem 0: Let G be a given finite group with $|G|$ elements. Let $d^{(1)}$, …, $d^{(N)}$ be the inequivalent irreps of G, with dimensions n_1, …,n_N, respectively. Then for any two of these, say $d^{(\alpha)}$ and $d^{(\beta)}$, the matrix elements satisfy

$$\sum_g d^{(\alpha)}_{ij}(g)\, d^{(\beta)}_{kl}(g^{-1}) \;=\; \frac{|G|}{n_\alpha}\, \delta_{\alpha\beta}\, \delta_{il}\, \delta_{jk} \tag{3.18}$$

(no summation over α).

The key to proving this and subsequent results is the fact previously made use of to get (3.11), that the **summation over all group elements $g \in G$ doesn't care about the order in which the elements are taken**. That is, for instance, for any function $f(\cdot)$ and any fixed $g_1 \in G$ we have $\sum_g f(g) = \sum_g f(g^{-1}) = \sum_g f(g_1 g) = \sum_g f(g_1^{-1} g)$, etc.

Proofs from here on are non-examinable.

Proof of the fundamental orthogonality theorem: For convenience let $S_{il,kj}$ denote the array on the left of (3.18):

$$S_{il,kj} \;=\; \sum_g d^{(\alpha)}_{ij}(g)\, d^{(\beta)}_{kl}(g^{-1}) \;.$$

Because of the summation over group elements, this is independent of g. Initially we regard j, k as fixed and consider i, l as matrix indices. Because $d^{(\alpha)}_{ii'}(g_1)\, d^{(\alpha)}_{i'j}(g) = d^{(\alpha)}_{ij}(g_1 g)$ (homomorphism preserves group multiplication) we have for any fixed $g_1 \in G$

$$\begin{aligned}
d^{(\alpha)}_{ii'}(g_1)\, S_{i'l,kj} &= \sum_g d^{(\alpha)}_{ij}(g_1 g)\, d^{(\beta)}_{kl}(g^{-1}) \\
&= \sum_g d^{(\alpha)}_{ij}(g)\, d^{(\beta)}_{kl}((g_1{}^{-1} g)^{-1}) \\
&= \sum_g d^{(\alpha)}_{ij}(g)\, d^{(\beta)}_{kl}(g^{-1} g_1) \\
&= \sum_g d^{(\alpha)}_{ij}(g)\, d^{(\beta)}_{kl'}(g^{-1})\, d^{(\beta)}_{l'l}(g_1) \\
&= S_{il',kj}\, d^{(\beta)}_{l'l}(g_1) \;.
\end{aligned}$$

The second line uses the complete rearrangement property — invariance when the order of summation is changed — and the fourth line again uses the homomorphism property.

Since the foregoing holds for all $g_1 \in G$, Schur's lemma implies that

$$S_{il,kj} \propto \delta_{il} \,, \qquad \alpha = \beta \,,$$

$$S_{il,kj} = 0, \qquad \alpha \neq \beta \,.$$

By a similar argument

$$d^{(\beta)}_{kk'}(g_1)\, S_{il,k'j} \;=\; S_{il,kj'}\, d^{(\alpha)}_{j'j}(g_1) \;,$$

giving

$$S_{il,kj} \propto \delta_{jk} \quad \text{when} \quad \alpha = \beta \,.$$

Altogether we have, therefore, for some constant C,

$$S_{il,kj} = \sum_g d^{(\alpha)}_{ij}(g) d^{(\beta)}_{kl}(g^{-1}) = C\,\delta_{\alpha\beta}\,\delta_{il}\delta_{jk}\ .$$

For $\alpha = \beta$, summing over $j = k$ gives, since $d^{(\alpha)}_{il}(gg^{-1}) = d^{(\alpha)}_{il}(I) = \delta_{il}$,

$$|G|\,\delta_{il} = C\,\delta_{il}\,\delta_{kk} \quad \Rightarrow \quad C = |G|\,/n_\alpha$$

since δ_{kk} sums to n_α. This gives the orthogonality theorem (3.18).

Remark: If the representation $d^{(\beta)}$ is *unitary*, we may take

$$d^{(\beta)}_{kl}(g^{-1}) \;=\; d^{(\beta)}_{kl}(g)^{-1} = \; d^{(\beta)}_{lk}(g)^* \tag{3.19}$$

where $*$ denotes the complex conjugate.

Corollary of Theorem 0 (**character 'row orthogonality'**, or 'gridlock mark 1'): If we set $i = j$ and $k = l$ in the orthogonality relation (3.18) and sum over the repeated indices using $\delta_{jl}\delta_{jl} = \delta_{jj} = n_\alpha$, the result is

$$\sum_g \chi_{d^{(\alpha)}}(g)\,\chi_{d^{(\beta)}}(g^{-1}) \;=\; |G|\,\delta_{\alpha\beta}\ , \tag{3.20}$$

or equally well (because, yet again, the sum over group elements doesn't care about the order in which the elements are taken)

$$\sum_g \chi_{d^{(\alpha)}}(g^{-1})\,\chi_{d^{(\beta)}}(g) \;=\; |G|\,\delta_{\alpha\beta}\ . \tag{3.21}$$

In the unitary case these become, respectively,

$$\sum_g \chi_{d^{(\alpha)}}(g)\chi^*_{d^{(\beta)}}(g) \;=\; |G|\,\delta_{\alpha\beta} \tag{3.22}$$

using the trace of (3.19) (note that the right-hand side is real), and

$$\sum_g \chi^*_{d^{(\alpha)}}(g)\chi_{d^{(\beta)}}(g) \;=\; |G|\,\delta_{\alpha\beta}\ , \tag{3.23}$$

which is the same thing rearranged (and, in any case, a real quantity).

A particular case of the orthogonality relation (3.20) occurs if we choose $d^{(\beta)}$ to be the trivial irrep, and $d^{(\alpha)}$ to be any distinct irrep. Then $\chi_{d^{(\beta)}}(g^{-1}) = 1$ for all g, so $\sum_g \chi_{d^{(\alpha)}}(g) = 0$, that is, the characters add up to zero, a result illustrated several times in the examples above.

Since the character values are the same for each group element in a given conjugacy class, we can rewrite these orthogonality relations in the form

$$\sum_{C_i} |C_i|\ \chi_{d^{(\alpha)}}(g_i)\chi_{d^{(\beta)}}(g_i^{-1}) \ = \ |G|\,\delta_{\alpha\beta}\ , \tag{3.24}$$

the sum now being over the classes C_i and $|C_i|$ the number of elements in the class C_i, with g_i any representative of C_i, i.e. any *single* element in C_i.

Remark: One may think of a 'representation' (faithful or unfaithful, reducible or irreducible) in two slightly different ways. The first way is, as above, to think of it simply as the set of matrices, such as $\mathbf{D}(g)$ or $\mathbf{d}(g)$ (lower-case for irreps) with components $D_{ij}(g)$ or $d_{ij}(g)$, to which a group G is isomorphic or homomorphic. For any one representation there are of course $|G|$ of these matrices, as g runs over the whole group G. The second way[10] is to think of the 'representation' as the mapping itself, the homomorphism, between G and the set of matrices. The mapping is often denoted by $D(g)$ or $d(g)$.

Please note: *Everything* from here on is non-examinable — including the section on molecular oscillations, as per schedule.

In order to prove Theorem 2, we need to consider first the

3.9.1 Decomposition of a reducible representation

Given any finite group G, let us suppose that we know its irreps, say homomorphisms $d^{(1)}$, $d^{(2)}$, ..., $d^{(N)}$, and matrices $\mathbf{d}^{(1)}$, $\mathbf{d}^{(2)}$, ..., $\mathbf{d}^{(N)}$ in some basis. Let D be any reducible representation of G. This means that D must be decomposable as a block-diagonal 'direct sum'

$$m_1 d^{(1)} \oplus m_2 d^{(2)} \oplus \cdots \oplus m_N d^{(N)}\ ,$$

where the m_k are integers called the multiplicities. The term **direct sum** and the symbol $\oplus$ simply mean, by definition, the concatenation of irreps down the main diagonal of each block-diagonal matrix of D, produced by some similarity transformation $\mathbf{S}$. The multiplicities m_k are the number of times each irrep is repeated in the direct sum, i.e., the number of copies of each irrep that appears on the main diagonal.

Symbolically, we may say the same thing as follows. Let $\mathbf{D}(g)$ be the matrix corresponding to some group element $g \in G$ in an *arbitrary* basis. Then the reduction can be written symbolically as

$$\mathbf{S}\,\mathbf{D}(g)\,\mathbf{S}^{-1} \ = \ \mathbf{I}^{(m_1)} \otimes \mathbf{d}^{(1)}(g) \ \oplus \ \mathbf{I}^{(m_2)} \otimes \mathbf{d}^{(2)}(g) \ \oplus \ \cdots \ \oplus \ \mathbf{I}^{(m_N)} \otimes \mathbf{d}^{(N)}(g)\,, \tag{3.25}$$

[10] This second way of thinking tends to be favoured by pure mathematicians.

where the symbol $\otimes$ is shorthand for block matrix multiplication, with $\mathbf{I}^{(m_1)}$, the $m_1 \times m_1$ unit matrix, acting on $\mathbf{d}^{(1)}$ as if it were a scalar and thus creating m_1 copies of it concatenated down the diagonal, and similarly with $\mathbf{d}^{(2)}$, ..., $\mathbf{d}^{(N)}$.

Now remember that the character value $\chi_D(g)$ of the representation matrix $\mathbf{D}(g)$ is invariant under similarity transformations. Therefore, taking the trace of (3.25), we have

$$\chi_D(g) \;=\; m_1\chi_{d^{(1)}}(g) + m_2\chi_{d^{(2)}}(g) + \cdots + m_N\chi_{d^{(N)}}(g)\;. \tag{3.26}$$

We can use this result to find m_1, m_2, ..., m_N without having to find the similarity transformation $\mathbf{S}$. Multiply (3.26) by $\chi_{d^{(\alpha)}}(g^{-1})$, sum over all classes in G, and use the orthogonality relation (3.24):

$$m_\alpha \;=\; \frac{1}{|G|}\sum_{C_i} |C_i|\,\chi_D(g_i)\chi_{d^{(\alpha)}}(g_i^{-1})\;, \tag{3.27}$$

and similarly for m_2, ..., m_N.

3.9.2 Decomposition of the regular representation, and proof of Theorem 2

Given any finite group G, we denote its regular representation by $D^{(R)}$, and the individual matrix corresponding to group element $g \in G$ by $\mathbf{D}^{(R)}(g)$. Recall first that the character of $D^{(R)}$ has only one nonzero character element, that corresponding to the identity:

$$\chi_{D^{(R)}}(I) = |G|\,, \qquad \chi_{D^{(R)}}(g) = 0\,, \quad g \neq I\,. \tag{3.28}$$

(This was shown on page 63 to follow from the complete rearrangement property, which holds for the particular row-permutations of $\mathbf{I}$ that make up the matrices $\mathbf{D}^{(R)}(g)$.)

To decompose $D^{(R)}$ into the irreps of G we use the orthogonality relation for characters, (3.24), and its consequence (3.27). For any irrep we have $\chi_{d^{(\alpha)}}(I) = n_\alpha$, with n_α its dimension, so, using (3.28) in (3.27) and noting that $|C_i| = 1$ when $g_i = I$,

$$\text{RHS (3.27)} \;=\; \frac{1}{|G|}\sum_{C_i} |C_i|\,\chi_{D^{(R)}}(I)\;\chi_{d^{(\alpha)}}(I) \;=\; n_\alpha\;.$$

Thus (3.27) tells us that the dimension, n_α, of each irrep is equal to the multiplicity, m_α, of its occurrence in the regular representation $D^{(R)}$. So

$$D^{(R)} = n_1 d^{(1)} \oplus n_2 d^{(2)} \oplus \cdots \oplus n_N d^{(N)}\;.$$

This means that for some similarity transformation $\mathbf{S}$ the individual matrices $\mathbf{D}^{(R)}(g)$ of the representation must all satisfy

$$\mathbf{S}\mathbf{D}^{(R)}(g)\mathbf{S}^{-1} = \mathbf{I}^{(n_1)} \otimes \mathbf{d}^{(1)}(g) \oplus \mathbf{I}^{(n_2)} \otimes \mathbf{d}^{(2)}(g) \oplus \cdots \oplus \mathbf{I}^{(n_N)} \otimes \mathbf{d}^{(N)}(g)\;. \tag{3.29}$$

Now substitute $g = I$, the identity element. By counting down the diagonal we immediately get

$$\sum_\alpha n_\alpha^2 \;=\; |G| \;,$$

which is Theorem 2.

3.9.3 Proof of character 'column orthogonality' or 'gridlock mark 2'

This finally leads to the proof of Theorem 1. And now, for the first time, it's **essential to use unitarity**, which (3.10)ff. assures us is always possible (by choosing basis vectors that are orthonormal with respect to the group-invariant inner product).

First we evaluate (3.29) for two group elements $g = g_1$ and $g = g_2^{-1}$, multiply the results together, and take the trace. For the left-hand side we have

$$\big(\mathbf{S}\mathbf{D}^{(R)}(g_1)\mathbf{S}^{-1}\mathbf{S}\mathbf{D}^{(R)}(g_2^{-1})\mathbf{S}^{-1}\big)_{ii} = \big(\mathbf{S}\mathbf{D}^{(R)}(g_1)\mathbf{D}^{(R)}(g_2^{-1})\mathbf{S}^{-1}\big)_{ii} = \big(\mathbf{S}\mathbf{D}^{(R)}(g_1 g_2^{-1})\mathbf{S}^{-1}\big)_{ii}$$

$$= S_{ij}D^{(R)}_{jk}(g_1 g_2^{-1})S^{-1}_{ki} = D^{(R)}_{jk}(g_1 g_2^{-1})\delta_{jk}$$

$$= D^{(R)}_{jj}(g_1 g_2^{-1}) = |G| \;\text{ or }\; 0 \;\text{ according as }\; g_1 = g_2 \text{ or } g_1 \neq g_2) \;.$$

by (3.28). Denote this by $|G|\,\delta_{g_1,g_2}$, extending the Kronecker delta notation in an obvious way.

The right-hand side is the trace of a block-diagonal matrix product in which each term of the form $\mathrm{Tr}\left(\mathbf{d}^{(\alpha)}(g_1)\,\mathbf{d}^{(\alpha)}(g_2^{-1})\right)$ occurs with multiplicity n_α. So altogether we have

$$\begin{aligned} |G|\,\delta_{g_1,\,g_2} &= \sum_\alpha n_\alpha \mathrm{Tr}\left(\mathbf{d}^{(\alpha)}(g_1)\mathbf{d}^{(\alpha)}(g_2{}^{-1})\right) \\ &= \sum_\alpha n_\alpha d^{(\alpha)}_{ij}(g_1) d^{(\alpha)}_{ji}(g_2{}^{-1}) \;\; \text{(summed over } i,j) \\ &= \sum_\alpha n_\alpha d^{(\alpha)}_{ij}(g_1)\, d^{(\alpha)}_{ij}(g_2)^* \;\; \text{(by unitarity)} \;. \end{aligned} \tag{3.30}$$

Now for any fixed i, j, α one can regard the matrix elements

$$d^{(\alpha)}_{ij}(g_1),\;\; d^{(\alpha)}_{ij}(g_2),\;\; d^{(\alpha)}_{ij}(g_3),\;\; \ldots$$

as the components of a $|G|$-dimensional vector. The orthogonality theorem (3.18) says that such vectors coming from different irreps and also from different positions in the matrix (different ij) are orthogonal, and that each vector has positive length. Hence these vectors **form a complete basis** in the $|G|$-dimensional vector space.

Similarly, if we fix g, we can regard the elements

$$d^{(\alpha)}_{ij}(g)$$

for varying i, j, α as the components of a $|G|$-dimensional vector. (Theorem 2 shows there are $|G|$ components.) The result (3.30) shows that these vectors **also form a complete basis** in the $|G|$-dimensional vector space.

In both cases, we use that given a basis of orthonormal vectors $\mathbf{v}_r$ ($r = 1, \dots n$) in $\mathbb{C}^n$, satisfying $(\mathbf{v}_r\,,\,\mathbf{v}_s) = \mathbf{v}_r{}^\dagger \mathbf{v}_s = \delta_{rs}$, there's a **completeness relation** that takes the form $\sum_r \mathbf{v}_r\,\mathbf{v}_r{}^\dagger = \mathbf{I}^{(n)}$, saying that the vectors span all of $\mathbb{C}^n$.

And now a similar completeness relation may be obtained for characters. The character orthogonality relation (3.24) corresponds to orthogonality of N c-dimensional vectors where c is the number of conjugacy classes and N is the number of irreps. The associated completeness relation will then require that $N = c$, giving Theorem 1.

To derive it we let $g_1 \to g g_1 g^{-1}$ in (3.30), second-bottom line, then fix g_1 and g_2 and sum over all $g \in G$ as well as over α, i and j,

$$\begin{aligned}
|G| \sum_g \delta_{g g_1 g^{-1},\, g_2} &= \sum_g \sum_\alpha n_\alpha d^{(\alpha)}_{ij}(g g_1 g^{-1})\, d^{(\alpha)}_{ji}({g_2}^{-1}) \\
&= \sum_\alpha n_\alpha \sum_g d^{(\alpha)}_{ik}(g) d^{(\alpha)}_{kl}(g_1) d^{(\alpha)}_{lj}(g^{-1}) d^{(\alpha)}_{ji}({g_2}^{-1}) \\
&= |G| \sum_\alpha \delta_{ij}\,\delta_{kl}\, d^{(\alpha)}_{kl}(g_1) d^{(\alpha)}_{ji}({g_2}^{-1}) \\
&= |G| \sum_\alpha \chi_{d^{(\alpha)}}(g_1) \chi_{d^{(\alpha)}}({g_2}^{-1}) \\
&= |G| \sum_\alpha \chi_{d^{(\alpha)}}(g_1) \chi_{d^{(\alpha)}}(g_2)^* \,,
\end{aligned}$$

where the third step uses Theorem 0, the full orthogonality relation (3.18). Now as g runs over the whole group G, with g_1 and g_2 fixed, $g g_1 g^{-1}$ runs (more than once) over the conjugacy class C_1 containing g_1, whose size $|C_1| < |G|$. If C_1 differs from the class C_2 containing g_2, then the delta function on the left is always zero. If $C_1 = C_2$ then some of the terms on the left are nonzero, because conjugation by some element or elements of G must send g_1 to g_2. There must be at least one such element, $\tilde{g}$ say. Then $g_2 = \tilde{g} g_1 \tilde{g}^{-1}$. If there is any other such element, $\tilde{g}'$ say, then it must satisfy $g_2 = \tilde{g}' g_1 \tilde{g}'^{-1} = \tilde{g} g_1 \tilde{g}^{-1}$, which can be rewritten as $h g_1 h^{-1} = g_1$ where $h = \tilde{g}^{-1}\tilde{g}'$. But the set of all h with $h g_1 h^{-1} = g_1$, for fixed g_1, is a subgroup of G, say H_{g_1} (Sheet 2 q. 6) because if $k \in H_{g_1}$ then $k g_1 k^{-1} = g_1$, $\Rightarrow$ $k^{-1} k g_1 k^{-1} k = k^{-1} g_1 k$ $\Rightarrow$ $g_1 = k^{-1} g_1 k$, so $k^{-1} \in H_{g_1}$; and, again, if $k, \ell \in H_{g_1}$ then $(k\ell) g_1 (k\ell)^{-1} = k(\ell g_1 \ell^{-1}) k^{-1} = k g_1 k^{-1} = g_1$, so $k\ell \in H_{g_1}$. So the set of elements $\{\tilde{g},\, \tilde{g}',\, \tilde{g}'', \dots\}$ that send g_1 to g_2 is in 1–1 correspondence with the set $\tilde{g}^{-1}\{\tilde{g},\, \tilde{g}',\, \tilde{g}'', \dots\}$, which is just H_{g_1}. So the (co)set $\{\tilde{g},\, \tilde{g}',\, \tilde{g}'', \dots\}$ has size $|H_{g_1}|$. Now g_2 can be *any* element of C_1, including g_1 itself (which is sent to itself by conjugation with any $h \in H_{g_1}$). So G can be partitioned into $|C_1|$ disjoint sets each of size $|H_{g_1}|$. So $|C_1||H_{g_1}| = |G|$. Therefore $\sum_g \delta_{g g_1 g^{-1},\, g_2} = |H_{g_1}| \delta_{C_1,C_2} = \big(|G|/|C_1|\big)\,\delta_{C_1,C_2}$ hence, finally, **column orthogonality for characters**, in turn implying Theorem 1:

$$\sum_\alpha \chi_{d^{(\alpha)}}(g_1) \chi_{d^{(\alpha)}}(g_2)^* = \frac{|G|}{|C_1|}\,\delta_{C_1,C_2}\,. \tag{3.31}$$

3.10 Applications to normal modes

An important application of representations and their decompositions is to simplify the analysis of normal modes and their associated normal frequencies. We may also determine the **degeneracy**. The degeneracy of a normal frequency is the number of distinct modes that share this frequency.

Following Chapter 1 we assume a general Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\,\dot{\mathbf{q}}^{\mathsf{T}}\mathbf{T}\dot{\mathbf{q}} - \tfrac{1}{2}\,\mathbf{q}^{\mathsf{T}}\mathbf{V}\mathbf{q}\,.$$

If the system has a symmetry then the action of the symmetry transformation g on the generalized coordinates $\mathbf{q}$ is

$$\mathbf{q} \longrightarrow \mathbf{D}(g)\mathbf{q}\,,$$

where, if there are $\mathcal{N}$ degrees of freedom so that $\mathbf{q}$ is an $\mathcal{N}$-component column vector, $\mathbf{D}(g)$ is a $\mathcal{N}\times\mathcal{N}$ matrix. For a symmetry, $\mathcal{L}$ must be invariant and this requires

$$\mathbf{D}(g)^{\mathsf{T}}\,\mathbf{T}\mathbf{D}(g) = \mathbf{T}\,, \qquad \mathbf{D}(g)^{\mathsf{T}}\,\mathbf{V}\mathbf{D}(g) = \mathbf{V}\,.$$

For real coordinates we may take the matrices $\mathbf{D}(g)$ to be *orthogonal*, or $\mathbf{D}(g)^{\mathsf{T}}\mathbf{D}(g) = \mathbf{I}$, so that this becomes[11]

$$\mathbf{D}(g)^{-1}\mathbf{T}\mathbf{D}(g) = \mathbf{T}\,, \qquad \mathbf{D}(g)^{-1}\mathbf{V}\mathbf{D}(g) = \mathbf{V}\,. \tag{3.32}$$

The representation provided by the matrices $\mathbf{D}(g)$ may be decomposed into irreps. For simplicity we'll assume at first the multiplicities are all 1. Then after a similarity transformation, $\mathbf{S}\mathbf{D}(g)\mathbf{S}^{-1}$ for some nonsingular $\mathbf{S}$, we have

$$\mathbf{S}\mathbf{D}(g)\mathbf{S}^{-1} = \left(\begin{array}{c|c|c|c} \mathbf{d}^{(1)}(g) & & & \\ \hline & \mathbf{d}^{(2)}(g) & & \\ \hline & & \ddots & \\ \hline & & & \mathbf{d}^{(N)}(g) \end{array}\right) \tag{3.33}$$

with all irreps different. Since (3.32) must be true for all group elements g we must then have, by Schur's lemma,

$$\mathbf{S}\mathbf{T}\mathbf{S}^{-1} = \left(\begin{array}{c|c|c|c} t_1\mathbf{I}^{(n_1)} & & & \\ \hline & t_2\mathbf{I}^{(n_2)} & & \\ \hline & & \ddots & \\ \hline & & & t_N\mathbf{I}^{(n_N)} \end{array}\right) \tag{3.34}$$

[11] The normal mode frequencies are given by the eigenvalues of $\mathbf{T}^{-1}\mathbf{V}$ and without requiring $\mathbf{D}(g)$ to be orthogonal we have $\mathbf{D}(g)^{-1}\mathbf{T}^{-1}\mathbf{V}\mathbf{D}(g) = \mathbf{T}^{-1}\mathbf{V}$.

and

$$\mathbf{SVS}^{-1} = \left(\begin{array}{c|c|c|c} v_1\mathbf{I}^{(n_1)} & & & \\ \hline & v_2\mathbf{I}^{(n_2)} & & \\ \hline & & \ddots & \\ \hline & & & v_N\mathbf{I}^{(n_N)} \end{array}\right) \tag{3.35}$$

where the t_α and the v_α are scalar numbers, and where $\mathbf{I}^{(n_\alpha)}$ is the $n_\alpha \times n_\alpha$ unit matrix, n_α being the dimension of the irrep $d^{(\alpha)}$. Note that

$$\mathbf{I} = \left(\begin{array}{c|c|c|c} \mathbf{I}^{(n_1)} & & & \\ \hline & \mathbf{I}^{(n_2)} & & \\ \hline & & \ddots & \\ \hline & & & \mathbf{I}^{(n_N)} \end{array}\right) . \tag{3.36}$$

With the diagonal form for $\mathbf{T}$ and $\mathbf{V}$ given by (3.34) and (3.35) finding the normal frequencies is easy; they are just ${\omega_\alpha}^2 = v_\alpha/t_\alpha$ with degeneracy n_α, the dimension of the irrep.

The normal mode generalized eigenvectors just span the invariant subspace (of the underlying vector space) acted on by the relevant irrep of the symmetry group.

More generally, the same irrep may occur more than once in the decomposition of D. This is the situation where multiplicities can be 2 or greater.

If the irrep $d^{(\alpha)}$ occurs with multiplicity m_α, then the scalar numbers t_α and v_α are replaced by $m_\alpha \times m_\alpha$ matrices $\mathbf{T}^{(m_\alpha)}, \mathbf{V}^{(m_\alpha)}$. Schur's lemma now gives

$$\begin{aligned} \mathbf{STS}^{-1} &= \mathbf{T}^{(m_1)} \otimes \mathbf{I}^{(n_1)} \oplus \mathbf{T}^{(m_2)} \otimes \mathbf{I}^{(n_2)} \oplus \cdots \oplus \mathbf{T}^{(m_N)} \otimes \mathbf{I}^{(n_N)} , \\ \mathbf{SVS}^{-1} &= \mathbf{V}^{(m_1)} \otimes \mathbf{I}^{(n_1)} \oplus \mathbf{V}^{(m_2)} \otimes \mathbf{I}^{(n_2)} \oplus \cdots \oplus \mathbf{V}^{(m_N)} \otimes \mathbf{I}^{(n_N)} , \end{aligned}$$

In this case $\mathbf{T}$ and $\mathbf{V}$ are not completely diagonal but finding the normal frequencies reduces to solving, for each α, $\det(\omega^2\mathbf{T}^{(m_\alpha)} - \mathbf{V}^{(m_\alpha)}) = 0$ and the m_α frequencies obtained from this equation each have degeneracy n_α.

The degeneracy of any normal frequency is then the dimension of the associated irrep of the symmetry group. (There may be extra degeneracy, which is either accidental (boring), or due to a hitherto unsuspected symmetry, perhaps an exciting new discovery.)

Example 3.8 *The* CO_2 *molecule*

There are three particles on a line, and the system is symmetric under reflections, so the symmetry group is

$$G = \{I, m\},$$

where $m^2 = I$. The action of the group is given by a representation D, where

$$\mathbf{D}(m)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_3 \\ -x_2 \\ -x_1 \end{pmatrix} \qquad \text{so} \qquad \mathbf{D}(m) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

The character of this representation has the values

$$\chi_D(I) = 3, \qquad \chi_D(m) = -1.$$

Since the order of the group is 2, there can only be 2 irreps, each of dimension 1 (so that $1^2 + 1^2 = 2$). To be consistent with equation (3.27), we must have

$$D = d^{(1)} \oplus 2d^{(2)}.$$

(We need $\chi_D(I) = m_1 \times 1 + m_2 \times 1$ and $\chi_D(m) = m_1 \times 1 + m_2 \times (-1)$.) Mode 1 is symmetric ($m$ is represented by $+1$) with $x_2 = 0$, $x_1 = -x_3$, and modes 2 and 3 are antisymmetric (m represented by -1) with $x_1 = x_3$.

⌟

Example 3.9 *Oscillations of an equilateral triangle*

The symmetry group is Σ_3. Choosing the coordinates as in Example 1.5, the permutation (23), for example, acts on the coordinates according to

$$q_1 \to -q_1,\ q_2 \to q_2,\ q_3 \leftrightarrow -q_5,\ q_4 \leftrightarrow q_6,$$

so

$$\mathbf{D}((23)) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

The permutation (123) corresponds to a rotation when (q_3, q_4) is given by (q_1, q_2) rotated through 120° and similarly for (q_5, q_6) and (q_1, q_2). This is given by the matrix

$$\mathbf{D}((123)) = \begin{pmatrix} 0 & 0 & 0 & 0 & -1/2 & \sqrt{3}/2 \\ 0 & 0 & 0 & 0 & -\sqrt{3}/2 & -1/2 \\ -1/2 & \sqrt{3}/2 & 0 & 0 & 0 & 0 \\ -\sqrt{3}/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1/2 & \sqrt{3}/2 & 0 & 0 \\ 0 & 0 & -\sqrt{3}/2 & -1/2 & 0 & 0 \end{pmatrix}.$$

Exercise: with $\mathbf{V}$ given by (1.26) show that $\mathbf{D}((23))\mathbf{V} = \mathbf{V}\mathbf{D}((23))$ and $\mathbf{D}((123))\mathbf{V} = \mathbf{V}\mathbf{D}((123))$.

With the above results $\chi_D(I) = 6$, $\chi_D((23)) = 0$, $\chi_D((123)) = 0$. Applying the orthogonality theorem shows that

$$D = d^{(1)} \oplus d^{(2)} \oplus 2d^{(3)},$$

using the notation of page 75 for the irreps of Σ_3.

The two translations of the system account for one of the $d^{(3)}$'s (which have degeneracy 2), and the rigid rotation accounts for the $d^{(2)}$ (since the mode is unchanged by cyclic permutation but its sign is changed by reflections). The remaining $d^{(3)}$ represents a pair of degenerate nonzero modes and $d^{(1)}$ (non-degenerate) represents the dilation mode.

This is demonstrated by the transformation of the normal mode eigenvectors found in Chapter 1. For the two translational modes given by (1.27)

$$\begin{aligned}
\mathbf{D}((23))\begin{pmatrix}\mathbf{Q}^{(1)} & \mathbf{Q}^{(2)}\end{pmatrix} &= \begin{pmatrix}\mathbf{Q}^{(1)} & \mathbf{Q}^{(2)}\end{pmatrix}\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\\
\mathbf{D}((123))\begin{pmatrix}\mathbf{Q}^{(1)} & \mathbf{Q}^{(2)}\end{pmatrix} &= \begin{pmatrix}\mathbf{Q}^{(1)} & \mathbf{Q}^{(2)}\end{pmatrix}\begin{pmatrix}-1/2 & \sqrt{3}/2\\ -\sqrt{3}/2 & -1/2\end{pmatrix},
\end{aligned}$$

generating a two dimensional representation. For the rotation and dilation modes given by (1.29) and (1.30) we have

$$\begin{aligned}
\mathbf{D}((23))\mathbf{Q}^{(3)} &= -\mathbf{Q}^{(3)}, \quad \mathbf{D}((123))\mathbf{Q}^{(3)} = \mathbf{Q}^{(3)},\\
\mathbf{D}((23))\mathbf{Q}^{(4)} &= \mathbf{Q}^{(4)}, \qquad \mathbf{D}((123))\mathbf{Q}^{(4)} = \mathbf{Q}^{(4)}.
\end{aligned}$$

The remaining modes, as given in (1.31), $\begin{pmatrix}\mathbf{Q}^{(6)} & \mathbf{Q}^{(5)}\end{pmatrix}$ transform exactly as $\begin{pmatrix}\mathbf{Q}^{(1)} & \mathbf{Q}^{(2)}\end{pmatrix}$.

Example 3.10 *The* CH_4 *molecule*

CH_4 has hydrogen atoms at the vertices of a tetrahedron, so the symmetry group is Σ_4. This group acts on the 5×3 coordinates and the corresponding 15-dimensional representation D satisfies

$$\chi_D(I) = 15,\ \chi_D((12)) = 3,\ \chi_D((123)) = 0,$$

$$\chi_D((1234)) = -1,\ \chi_D((12)(34)) = -1.$$

This reduces to $d^{(1)} \oplus d^{(3)} \oplus 3d^{(4)} \oplus d^{(5)}$. The three rigid translations and three rigid rotations account for one $d^{(4)}$ and the $d^{(5)}$, leaving 9 modes with degeneracies 1, 2, 3, and 3. I'll show how to work out the full character table for Σ_4, but leave the rest as an exercise since these notes are getting FAR too long!!

There are 5 conjugacy classes with typical elements, in disjoint cycle notation,

$$I,\ (12),\ (123),\ (1234),\ (12)(34).$$

As remarked on page 57, the cycle shapes determine the conjugacy classes since this is a full 'symmetric group' or permutation group, being the symmetric group on 4 letters. Every other element can be written in one of these shapes. The numbers $|C_i|$ $(i = 1, \ldots, 5)$ of elements in each of these conjugacy classes are respectively

$$1,\ 6,\ 8,\ 6,\ 3.$$

By Theorem 2, the dimensions of the irreps satisfy

$$n_1^2 + n_2^2 + n_3^2 + n_4^2 + n_5^2 = 24$$

the only solution of which is

$$n_1 = n_2 = 1,\ n_3 = 2,\ n_4 = n_5 = 3.$$

The two 1-dimensional irreps are easy to identify: $d^{(1)}$ is (as always) the trivial irrep, and $d^{(2)}$ is -1 for odd permutations and $+1$ for even permutations. (This obviously gives a representation, i.e. qualifies as a homomorphism, and since we know that there are no other 1-dimensional irreps we need look no further.) At this stage, we can fill in the first column of the character table, and the first two rows below the $|C_i|$ (see below), that is, the rows for $d^{(1)}$ and $d^{(2)}$.

One of the 3-dimensional irreps, $d^{(4)}$ say, corresponds to rotation matrices in $\mathbb{R}^3$ that preserve the tetrahedron. The matrix corresponding to (12) is just a reflection, so $\chi_{d^{(4)}}((12)) = +1$ (because the eigenvalues of a reflection matrix are $+1$, $+1$ and -1, and the trace is the sum of the eigenvalues). The cycle (123) represents a rotation by $2\pi/3$, so $\chi_{d^{(4)}}((123)) = 0$, because the trace of a rotation matrix is equal to $1 + 2\cos\theta$, where θ is the angle of rotation.

And now we can use orthogonality to complete the $d^{(4)}$ row of the table. (We could nearly have done it without finding $\chi_{d^{(4)}}((123))$, using the orthogonality theorem (3.24) with $\alpha = \beta$. But because this gives a quadratic equation for the unknown character values, there are two solutions, only one of which is the required one.)

Surprisingly, the $d^{(3)}$ row is then completely determined by the orthogonality theorem. (The final equation is quadratic, but has one repeated root.) The $d^{(5)}$ row is then similarly determined.

So we can finally pin down the entire character table to be the following. To save space, in the conventional way, it's laid out with the conjugacy classes lumped into single columns. The numbers in the second row are *not* character elements, but the sizes $|C_i|$ of the classes:

	I	(12)	(123)	(1234)	(12)(34)
$\lvert C_i \rvert$	1	6	8	6	3
$\chi_{d^{(1)}}$	1	1	1	1	1
$\chi_{d^{(2)}}$	1	-1	1	-1	1
$\chi_{d^{(3)}}$	2	0	-1	0	2
$\chi_{d^{(4)}}$	3	1	0	-1	-1
$\chi_{d^{(5)}}$	3	-1	0	1	-1

$\lrcorner$

$\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim$ **END OF NOTES** $\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim$