Linear Transformations and Group Representations

These notes are intended to follow the "Groups, Fields, and Vector Spaces" notes from 2024-2025.

Eigenvectors and Eigenvalues

The terms "linear transformation of (or on) *V*", "linear operator on *V*", and "member of $Hom(V,V)$ will be used interchangeably.

Definitions

Above we defined the determinant of a linear transformation *A* on *V*, and (by doing this in a coordinate-free manner) showed that it is an intrinsic property of *A*, i.e., one that is independent of the choice of basis. Here we use the determinant to find some other intrinsic properties of *A*.

For a linear transformation *A* in a vector space *V*, an eigenvector is *v* is, by definition, a nonzero vector that satisfies $Av = \lambda v$ for some scalar (field element) λ. λ is called the eigenvalue for *A* associated with *v*. λ is allowed to be 0, but *v* must be nonzero. Note that eigenvalues and eigenvectors are defined in a coordinate-free fashion, so they are intrinsic properties of *A*.

Typically, a linear transformation has a whole a *set* of eigenvalues λ_j and associated eigenvectors v_j, satisfying $Av_j = \lambda_j v_j$.

We will initially work in a finite-dimensional vector space. This allows us to see the algebraic structure clearly, but it also uses some tools that do not apply in the infinite-dimensional case. (And, without further assumptions, many of the results for finite dimensions do not hold in the infinite-dimensional case.) But later we will add one more piece of structure – the "inner product" – which (a) restricts the infinite-dimensional spaces we can consider, but (b) ensures that the key results for finite-dimensional spaces will apply. Essentially, this happens because the "inner product" yields a notion of distance, and the notion of distance allows us to restrict attention to vectors that can be well-approximated by vectors in a finite-dimensional space.

For a finite-dimensional vector space *V*, (say, of dimension *n*) we can find the eigenvalues of *A* by solving the "characteristic equation" of *A*, namely, $\det(zI - A) = 0$. (Here, *I* is the identity transformation on *V*). This works because if $\det(zI - A) = 0$ is solved by $z = \lambda$, then $\lambda I - A$ is an operator that transforms a basis set into a set whose span is of at most dimension $n-1$. So some linear combination of the basis set (say, *v*) must be mapped to 0 by $\lambda I - A$. And if $(\lambda I - A)v = 0$, then $\lambda I v = Av$, i.e., $\lambda v = Av$.

If the dimension of *V* is *n*, the characteristic equation is a polynomial of degree *n*, i.e.,

$$\det(zI - A) = z^n - a_1 z^{n-1} + a_{n-2} z^{n-2} - \ldots + (-1)^n a_n .$$

The characteristic equation leads to intrinsic descriptions

Note that since we didn't use coordinates to define the determinant, the characteristic equation is an intrinsic property of A. This means that each of its roots – i.e., each of the eigenvalues – are intrinsic properties of A.

One of our goals: our setup is that V is a space of signals and $Hom(V,V)$ a space of linear transformations (e.g., from input to output), and we are studying a particular A in $Hom(V,V)$. We would like to describe A in terms of its intrinsic properties.

Each of the coefficients of the characteristic equation are also intrinsic properties of A. The coefficient a_n is the determinant of A (set $z=0$ in the above), but the other coefficients yield other intrinsic properties.

To find out what these are, we need to think about what the definition of the determinant means in terms of coordinates. The determinant of a linear transformation L is a sum of terms, each of which is an n-fold product of the entries in L. A typical term, say, $L_{1,j_1}L_{2,j_2}...L_{n,j_n}$, corresponds to letting L act on an elementary tensor product $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ of distinct basis vectors, and determining the coordinate of $v_{j_1} \otimes v_{j_2} \otimes ... \otimes v_{j_n}$ in the result – where in the result, we choose the j_k th basis element of V from the k th copy of V. So there will be one term in the determinant that is a product of the diagonal elements of L, i.e., $L_{1,1}L_{2,2}...L_{n,n}$, corresponding to choosing the first basis element from the first copy of V, the second basis element from the second copy of V, etc. In all the other terms (as well as in this term), all of the j_k 's must be distinct, since otherwise the term would be annihilated by the antisymmetrization process. This means that there are no terms that contain exactly one term in which the subscripts of L do not match, i.e., there are no terms that contain only one off-diagonal element of L.

Applying this to $L = zI - A$ shows that the coefficient of z^{n-1} in $\det(zI - A)$, namely a_1, is the sum of the elements on the diagonal of A. This is known as the "trace" of A, $\mathrm{tr}(A)$.

If the base field is algebraically closed, then the characteristic equation will have a full set of solutions (roots), which we denote $\lambda_1,...,\lambda_n$. Some of these may be duplicates. But in any case, the characteristic equation can be factored completely:
$\det(zI - A) = (z-\lambda_1)(z-\lambda_1)\cdot...\cdot(z-\lambda_n)$. This consequence is why we choose $k = \mathbb{C}$: the complex numbers are algebraically closed. So the characteristic equation always has a full set of roots.)

Equating coefficients with the characteristic equation shows that $\mathrm{tr}(A)$ is the sum of the eigenvalues, and $\det(A)$ is the product of the eigenvalues, and also gives meaning to the other coefficients of the characteristic equation – for example, a_2 is the sum of all pairwise products of the eigenvalues, $\lambda_1\lambda_2 + \lambda_1\lambda_3 + ... + \lambda_{n-1}\lambda_n$.

Because the trace is the sum of the eigenvalues, it has another important property that we will use below: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. To see this, write $C = BA$, so $AB = ACA^{-1}$. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ is thus equivalent to $\mathrm{tr}(ACA^{-1}) = \mathrm{tr}(C)$. Three ways to see this: (i) recognize that ACA^{-1} is the same transformation as *C*, written in a different basis set. (ii) The trace is the highest-power term in the characteristic equation, and ACA^{-1} has the same characteristic equation as *C*: $\det(\lambda I - ACA^{-1}) = \det(A(\lambda I - C)A^{-1}) = \det(\lambda I - C))$. (iii) Any eigenvector/eigenvalue pair for C, say, (λ, *v*) corresponds to an eigenvector/eigenvalue pair for ACA^{-1}, namely (λ, Av) (and vice-versa): $ACA^{-1}(Av) = ACv = A\lambda v = \lambda(Av)$. (These arguments only holds if *A* is invertible, but they are is readily extended to the case when *A* is not.)

To emphasize: the eigenvalues and eigenvalues of *A* do not depend on the coordinates chosen for *V* – so they form a coordinate-independent description of *A*. (Of course to *communicate* the eigenvectors v_j, one typically does need to choose coordinates.)

Eigenvalues define subspaces

Eigenvectors corresponding to the same eigenvalue form a subspace. To see this, suppose *v* and *w* are both eigenvectors of *A* with the same eigenvalue λ. Then any linear combination of *v* and *w* also is an eigenvector of *A* with the eigenvalue λ.
$A(av + bw) = aAv + bAw = a\lambda v + b\lambda w = \lambda(av + bw)$.

So we can talk about the eigenspace associated with an eigenvalue λ, namely, the set of all eigenvectors. This forms a subspace of the original space *V*.

Conversely, eigenvectors corresponding to different eigenvalues lie in different subspaces. Suppose instead that *v* is an eigenvector of *A* with the eigenvalue λ, and that *W* is a subspace of *V* with a basis set of eigenvectors w_m all of whose eigenvalues λ_m are distinct from λ. Then *v* cannot be in *W*. For if *v* were in *W*, then we could write $v = \sum a_m w_m$. On the one hand, $Av = \lambda v$ so $Av = \sum \lambda a_m w_m$. On the other hand, we could write $Av = A\left(\sum a_m w_m\right) = \sum a_m A(w_m) = \sum \lambda_m a_m w_m$. Since the w_m are a basis set, they are linearly independent, so the coefficients of the w_m must match in these two expansions of *Av*. That is, for each *m*, we would need to have $(\lambda - \lambda_m)a_m = 0$. Since we have assumed that for all *m,* $\lambda_m \neq \lambda$, it follows that all the a_m must be 0 – so *v* is not an eigenvector.

The above comment guarantees that eigenvectors corresponding to distinct eigenvalues are linearly independent.

While there is no guarantee that the eigenvectors span *V*, there are many circumstances when this is the case. One case is that the characteristic equation has all distinct roots. Others are mentioned below.

When the eigenvectors form a basis

Say there is a special linear transformation T (specified by the problem at hand), with all of its eigenvalues λ_j distinct. Then its eigenvectors v_j form a basis that is singled out by T. It is also a basis in which the action of T on any vector $v \in V$ is simple to specify: since $v = \sum a_j v_j$ for some set of coefficients a_j, then $T(v) = T\left(\sum a_j v_j\right) = \sum T\left(a_j v_j\right) = \sum a_j T\left(v_j\right) = \sum a_j \lambda_j v_j$.

Another way of looking at this is that if you use the eigenvectors v_j as the basis set, then the matrix representation of T is $T = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$.

Note also that if the eigenvalues of T are λ_j, then the eigenvalues of aT are $a\lambda_j$, the eigenvalues of T^2 are λ_j^2, etc., and, we can even interpret $f(T)$ as a transformation with eigenvalues $f(\lambda_j)$, for any function f.

Shared eigenvectors and commuting operators

Say A and B are linear transformations, and $AB = BA$. If v is an eigenvector of B with eigenvalue λ, then Av is also an eigenvector of B with eigenvalue λ. This is because

$$B(Av) = (BA)v = (AB)v = A(Bv) = A(\lambda v) = \lambda(Av).$$

Now further suppose that the eigenspace of B corresponding to eigenvalue λ has dimension 1, i.e., this eigenspace consists of the scalar multiples of v. It follows that v is also an eigenvector of A. This is because (under the dimension-1 hypothesis) Av and v are both in the same one-dimensional eigenspace of B, so it must be that Av is a multiple of v, i.e., $Av = \mu v$, i.e., v is an eigenvector of A.

So if all of the eigenspaces of B have dimension 1, it follows that *every* eigenvector of B is also an eigenvector of A. This means that the eigenvector basis set that diagonalizes B also diagonalizes A.

Thus, even if all of the eigenvalues of A are not distinct, the fact that it commutes with an operator B whose eigenvalues are distinct means that the eigenvectors for B form a natural basis for A, in which A is diagonal. Diagonal matrices are exceptionally easy to add and multiply.

How this applies to signals and systems

Before proceeding further with the abstract development, it is useful to see how what we already have is applied to signals and systems.

V is a vector space of functions of time. Linear transformations on V arise as filters, as input-output relations, as descriptors of spiking processes, etc. We want to find invariant descriptors for linear transformations on V, and, if possible, a preferred basis set. We will use the fact that all of these linear transformations commute with time-translation, which is also a linear transformation. Additionally, the time-translation operation has the nice property that all of its eigenspaces have dimension 1.

Example: linear filters

The transformation $w = Lv$, with

$$w(t) = \int_0^\infty L(\tau)v(t-\tau)d\tau \tag{1}$$

is a linear transformation on V. View $v(t)$ as an input to a linear filter, $w(t)$ as an output. Here, $L(t)$, which describes L, is called the "impulse response": $L(t)$ is the response $w = Lv$ when $v(t) = \delta(t)$, the delta-function impulse, since $L\delta(t) = \int_0^\infty L(\tau)\delta(t-\tau)d\tau = L(t)$. (This is the basic property of the delta-function.)

Example: smoothing

Smoothing transformations are also linear transformations $w = Lv$, with

$$w(t) = \int_{-\infty}^\infty L(\tau)v(t-\tau)d\tau . \tag{2}$$

For example, take $L(t) = 1/(2h)$ if $|t| < h$, 0 otherwise -- "boxcar smoothing". Or take $L(t)$ to be a Gaussian. L in this context is often called the "smoothing kernel."

Below we show that time translation is an operator that commutes with the above L's. We will then use this to determine a natural basis for V, in which it is simple to describe the action of the L's, and to see how they combine.

Other examples that benefit from this setup will arise when we discuss point processes.

Time-translation invariance

In the above examples, the transformation L is "time-translation invariant" -- independent of absolute clock time. This crucial property can be formulated algebraically as a statement that time translation commutes with these operators.

To do this, we define the time-shift operator D_T on V as follows:

$$(D_T v)(t) = v(t+T) . \tag{3}$$

That is, D_T advances time by T units. Note that this is a linear transformation.

This is equivalent to an expression of the form (2), if we allow L to be a generalized function, $L(\tau)=\delta(\tau+T)$.

($\delta(x)$ is a generalized function that satisfies $\int\limits_{-\infty}^{\infty}\delta(x-a)g(x)dx=g(a)$. It can be thought of as the limiting case of a "blip" of width Δ and height $1/\Delta$, as $\Delta\to 0$). The "limit" is not a limit in the usual sense, but integrals of the delta-function have a limit, which is all we need. And it is also in keeping with our policy that if something makes sense with arbitrarily fine discretizations of time, then there should be an extension of it that makes sense in the continuum limit.)

With $L(\tau)=\delta(\tau+T)$, eq. (2) becomes $\int\limits_{-\infty}^{\infty}\delta(\tau+T)v(t-\tau)d\tau=v(t+T)=\left(D_Tv\right)(t)$, since the only contribution to the integral is when the argument of the delta-function is zero, i.e., when $\tau+T=0$, i.e., $\tau=-T$.

Time-translation invariance of a linear operator A means that A has the same effect if the absolute clock time is unchanged. That is, $AD_T=D_TA$. The left-hand side means, first shift absolute time and then apply A; the right-hand side means, first apply A and then shift absolute time.

To show that operators defined by eq. (1) or (2) are time-translation invariant (it suffices to consider the case of eq. (2)), we need $LD_Tv=D_TLv$:

$$\left(LD_Tv\right)(t)=\int\limits_{-\infty}^{\infty}L(\tau)\left(D_Tv(t-\tau)\right)d\tau=\int\limits_{-\infty}^{\infty}L(\tau)\left(v(t-\tau+T)\right)d\tau=\int\limits_{-\infty}^{\infty}L(\tau'+T)v(t-\tau')d\tau'$$

where the last equality follows by substituting $\tau'=\tau-T$. Consequently,

$$\int\limits_{-\infty}^{\infty}L(\tau'+T)v(t-\tau')d\tau'=\int\limits_{-\infty}^{\infty}D_TL(\tau')v(t-\tau')d\tau'=\left(D_TLv\right)(t).$$

Since D_T itself (for any T) is of the form (2), this means that $D_{T'}D_T=D_TD_{T'}$, i.e., any two D_T's commute. Thus, we should expect to find a set of vectors that are eigenvectors for all of the D_T's. These will turn out to be all distinct, and to span V.

Then, these eigenvectors must also be eigenvectors for any time-translation invariant operator L (i.e., an L for which $LD_T=D_TL$). Expressed in this basis, L (including all transformations of the form (1) or (2)) are diagonal – and thus, easy to manipulate.

We will first find the eigenvectors and eigenvalues of D_T "by hand." Then, we see how their properties arise because of the way that the time-translation group acts on the domain of the functions of V.

What are the eigenvectors and eigenvalues of D_T?

Let's find the vectors v that are simultaneous eigenvectors of all the D_T's.

First, observe that $D_S\left(D_T v\right)(t) = v(t+T+S) = D_{T+S}(v)$, so that $D_S D_T = D_{T+S}$. Intuitively, translating in time by T, and then by S, is the same as translating in time by T+S. Abstractly, the mapping $W : T \to D_T$ is a homomorphism of groups. It maps elements T of the group of the real numbers under addition (time translation) to some isomorphisms D_T of V.

Say $v(t)$ is an eigenvector for all of the D_T's. We next see how the eigenvalue corresponding $v(t)$ depends on T. Say the eigenvalue associated with $v(t)$ for D_T is $\lambda(T)$. Since $D_S D_T = D_{T+S}$, $v(t)$ is an eigenvector of D_{T+S}, with eigenvalue $\lambda(T+S) = \lambda(T)\lambda(S)$. So the dependence of the eigenvalue on T must satisfy $\lambda(T+S) = \lambda(T)\lambda(S)$. Equivalently, $\log\lambda(T+S) = \log\lambda(T) + \log\lambda(S)$. That is, $\log\lambda(T)$ must be proportional to T. Choose a proportionality constant c. $\log\lambda(T) = cT$ implies that $\lambda(T) = e^{cT}$, for some constant c.

This determines $v(t)$: This is because $v(t+T) = \left(D_T v\right)(t) = \lambda(T)v(t) = e^{cT}v(t)$. Choosing $t = 0$ now yields $v(T) = v(0)e^{cT}$, so these are the candidates for the simultaneous eigenvectors of all of the D_T's.

If we choose a value of c that has a positive real part, then $v(T)$ gets infinitely large as $T \to \infty$. But if we choose a value of c that has negative real part, then $v(T)$ gets infinitely large as $T \to -\infty$. So the only way that we can keep $v(T)$ bounded for all T is to choose c to be pure imaginary. With $c = i\omega$, $v(T) = e^{i\omega T}$.

The above elementary calculation found all the eigenvalues and eigenvectors of the translation operator, but it did not guarantee that the eigenvectors span the space (i.e., form a basis for it). This is also true, and it follows from some very general results about how groups (in this case, the translation group) act on vector spaces (in this case, functions on the line). We'll get a look at this general result below. But also, this result – that these eigenvectors are a basis – can be derived for periodic functions in discrete time simply by counting dimensions. In this case, the domain of the functions is $\mathbb{Z}_n$, time translation corresponds to rotating the n-gon, and the eigenvectors are $v_k(t) = e^{\frac{2\pi i}{n}kt}$.

Thus, the set of $v_\omega(T) = e^{i\omega T}$ (for all ω) not only form the complete set of eigenvectors of each of the D_T's, but also form a basis for a vector space of complex-valued functions of time. They thus constitute natural coordinates for this vector space, in which time-translation-invariant linear operators are all diagonal. Fourier analysis is simply the re-expression of functions of time in these coordinates. This is also why Fourier analysis is useful. Because linear operators are diagonal when expressed in these new coordinates, the actions of filters can be carried out by coordinate-by-coordinate multiplication, rather than integrals (such as eq. (1)).

Hilbert spaces

To get started with this general result, we need to add one more piece of structure to vector spaces: the inner product. An inner product (or "dot-product"), essentially, adds the notion of distance. A vector space with an inner product, and in which all inner products are finite, is known as a Hilbert space. In a Hilbert space, it is possible to make general statements about what kinds of linear transformations have a set of eigenvectors that form a basis.

Some preliminary comments:

We can always make a finite-dimensional vector space into a Hilbert space, since we are guaranteed a basis (choose a set of coordinate axes), and we can choose the standard dot-product in that basis. This determines a notion of distance, and hence, which vectors are "unit vectors", i.e., what are the spheres. Had we chosen a different set of coordinate axes, say, ones that are oblique (in the basis of the first set), then we would have defined a different dot-product. But we could always find a linear transformation from the vector space to itself that transforms one dot-product into the other – this is the linear transformation that changes the first basis set into the second one. It would turn spheres into ellipsoids, and vice-versa. Thus, while adding Hilbert space structure to a finite-dimensional vector space does add a notion of "geometry", it doesn't allow us to prove things that we couldn't prove before – since Hilbert space structure was guaranteed, and all we are doing is choosing one example from an infinite set of possibilities.

The situation is very different for infinite-dimensional vector spaces, such as function spaces. Here, when we add a dot-product (and insist that it has a finite value), we actually need to exclude some functions from the space. As in the finite-dimensional case, adding the dot-product gives a notion of "geometry." But it does something even more important: by excluding some functions from the vector space, it allows many it allows our intuitions from finite-dimensional vector spaces to generalize.

Definition of an inner product

An inner product (or "dot-product") on a vector space V over the reals or complex numbers is a function from pairs of vectors to the base field, typically denoted $\langle v, w \rangle$ or $v \bullet w$. It must satisfy the following properties (where a is an element of the base field):

Symmetry: $\langle v, w \rangle = \langle w, v \rangle$ for $k = \mathbb{R}$, and $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for $k = \mathbb{C}$ "conjugate symmetry"). (Here and below, we denote the complex conjugate of a field element a by $\overline{a}$).

Linearity: $\langle av_1 + bv_2, w\rangle = a\langle v_1, w\rangle + b\langle v_2, w\rangle$, and $\langle v, aw_1 + bw_2\rangle = \overline{a}\langle v, w_1\rangle + \overline{b}\langle v, w_2\rangle$
The second equality follows from the first one by applying symmetry.

Positive-definiteness: $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ only for $v = 0$.

The quantity $\langle v, v\rangle = \|v\|^2$ can be regarded as the square of the size of v, i.e., the square of its distance from the origin.

Implicit in the above definition is that $\langle v, w\rangle$ is finite. This does not mean that there is a universal upper limit to it that applies to all members of V, just that for any pair of vectors, $\langle v, w\rangle$ is a finite number.

Note that $\langle av, bw\rangle = a\overline{b}\langle v, w\rangle$. The necessity for the complex-conjugation is apparent if one considers $\langle iv, iv\rangle$. With complex-conjugation of the "b", we find $\langle iv, iv\rangle = i\overline{i}\langle v, v\rangle = i(-i)\langle v, v\rangle = \langle v, v\rangle$, which is "good" – multiplication of v by a unit (i) does not change its length. But without complex conjugation, we'd find that $\langle iv, iv\rangle$ would equal $-\langle v, v\rangle$, i.e., positive-definiteness would be violated.

If $\langle v, w\rangle = 0$, v and w are said to be orthogonal.

The inner product, distances, triangle inequality, Cauchy-Schwartz, and angles

The quantity specified by $d(v, w) = \|v - w\| = \sqrt{\langle v - w, v - w\rangle}$ qualifies as a "metric" (i.e., a distance), because it is (a) symmetric, (b) non-negative, and (c) satisfies the triangle inequality $d(u, w) \leq d(u, v) + d(v, w)$. But demonstrating the triangle inequality is slightly harder than one might guess. The triangle inequality follows from positive-definiteness of the inner product.

We want to show that $d(v, w) \leq d(v, x) + d(x, w)$, i.e.,. $\|v - w\| \leq \|v - x\| + \|x - w\|$; with $y = v - x$ and $z = x - w$ this is equivalent to $\|y + z\| \leq \|y\| + \|z\|$, and to $\|y + z\|^2 \leq (\|y\| + \|z\|)^2 = \|y\|^2 + \|z\|^2 + 2\|y\|\|z\|$.

Since $\|y\|^2 = \langle y, y\rangle$, $\|z\|^2 = \langle z, z\rangle$, and $\|y + z\|^2 = \langle y + z, y + z\rangle = \langle y, y\rangle + \langle z, z\rangle + \langle y, z\rangle + \langle z, y\rangle$, the latter is equivalent to $\langle y, z\rangle + \langle z, y\rangle \leq 2\|y\|\|z\|$, which is implied by $\left|\langle y, z\rangle + \langle z, y\rangle\right|^2 \leq 4\|y\|^2\|z\|^2$. Since $\langle y, z\rangle + \langle z, y\rangle = \langle y, z\rangle + \overline{\langle y, z\rangle} = 2\,\mathrm{Re}\langle y, z\rangle \leq 2\left|\langle y, z\rangle\right|$, it suffices to show that $\left|\langle y, z\rangle\right| \leq \|y\|\|z\|$.

This final inequality is a form of the Cauchy-Schwartz inequality. It follows applying the positive-definiteness property to the difference between y and its projection on z (see section below on projections). We first find the projection of y on z: $p = z\frac{\langle y,z\rangle}{\langle z,z\rangle}$. The difference between this and y is given by $q = y - z\frac{\langle y,z\rangle}{\langle z,z\rangle}$. So $y = p + q$, with p proportional to z and p orthogonal to q (since $\langle q,p\rangle = \left\langle y - z\frac{\langle y,z\rangle}{\langle z,z\rangle}, z\frac{\langle y,z\rangle}{\langle z,z\rangle}\right\rangle = \langle y,z\rangle\frac{\langle y,z\rangle}{\langle z,z\rangle} - \langle z,z\rangle\left(\frac{\langle y,z\rangle}{\langle z,z\rangle}\right)^2 = 0$). That is, y is the hypotenuse of a right triangle, p lying along z, and q being the perpendicular to the line of z. Now we can calculate:

$$\langle y,y\rangle = \langle p+q,p+q\rangle = \langle p,p\rangle + \langle q,q\rangle = \left\langle z\frac{\langle y,z\rangle}{\langle z,z\rangle}, z\frac{\langle y,z\rangle}{\langle z,z\rangle}\right\rangle + \langle q,q\rangle$$

$$= \langle z,z\rangle\left|\frac{\langle y,z\rangle}{\langle z,z\rangle}\right|^2 + \langle q,q\rangle = \frac{\left|\langle y,z\rangle\right|^2}{\langle z,z\rangle} + \langle q,q\rangle$$

From this (and the non-negativity of $\langle q,q\rangle$), it follows that $\langle y,y\rangle \geq \frac{\left|\langle y,z\rangle\right|^2}{\langle z,z\rangle}$, which (since $\langle z,z\rangle$ is non-negative – if it were zero, the Cauchy-Schwartz inequality would have been trivial) implies the desired inequality $\langle y,y\rangle\langle z,z\rangle \geq \left|\langle y,z\rangle\right|^2$.

The Cauchy-Schwartz inequality enables us to interpret the quantity $\frac{\left|\langle y,z\rangle\right|}{\sqrt{\langle y,y\rangle\langle z,z\rangle}}$ as the cosine of the angle between the vectors y and z, as it is zero if the vectors are orthogonal, and has a maximal value of 1 if y and z are proportional to each other. The Cauchy-Schwartz inequality guarantees that it is always a real number in the range $[0,1]$.

Examples of the inner product

For a vector space of n-tuples of complex numbers, the standard inner product is

$$\langle u,v\rangle = \sum_{n=1}^{N} u_n \overline{v_n}. \tag{4}$$

Note that although we used coordinates to define these inner product, defining an inner product is not the same as specifying coordinates. As we will see below, we can choose alternate sets of coordinates that lead to exactly the same inner product. This is because the inner product only fixes a notion of distance, while the coordinates specify individual directions.

For functions of time, the standard inner product is

$$\langle f,g\rangle = \int_{-\infty}^{\infty} f(t)\overline{g(t)}dt. \qquad (5)$$

We cannot consider all functions of time to form a Hilbert space with the inner product given by eq. (5), since this is not guaranteed to be finite. However, we can take V to be all functions of time for which the integral for $\langle f,f\rangle$ exists and is finite, i.e., that $\|f\|^2 = \langle f,f\rangle = \int_{-\infty}^{\infty} f(t)\overline{f(t)}dt = \int_{-\infty}^{\infty} |f(t)|^2 dt$ is finite. This guarantees (not obvious – Cauchy's inequality) that eq. (5) is finite as well, and makes this space (the "square-integrable functions of time") to be a Hilbert space. It's easy to find an example of a function that is perfectly well-defined, but for which the above integral does not exist – for example, a function that has any constant but nonzero value.

The inner product and the dual

An inner product specifies a correspondence between a vector space V and its dual V^*. (Remember, this was guaranteed in the finite-dimensional case, since the dimension of V and its dual are the same, but it is not guaranteed for the infinite-dimensional case.) That is, for each element v in V, the inner product provides a member φ_v of V^*, whose action is defined by $\varphi_v(u) = \langle u,v\rangle$. This correspondence is conjugate-linear (not linear), because $\varphi_{av} = \overline{a}\varphi_v$. Note, though, that not every conjugate-linear correspondence between V and V^* corresponds to an inner product, since the inner product also has to be positive-definite.

Some special kinds of linear operators

In a manner somewhat analogous to the above mapping between vectors and their duals, the inner product also specifies a mapping from an operator to its "adjoint": the adjoint of an operator A is the operator A^* (sometimes written $A^\dagger$) for which $\langle Au,v\rangle = \langle u,A^*v\rangle$, for all u and v.

A few basic properties. First, the adjoint of the adjoint is the original operator. $\left(A^*\right)^* = A$: Since, by definition, $\left(A^*\right)^*$ is defined as the operator for which $\langle A^*u,v\rangle = \left\langle u,\left(A^*\right)^* v\right\rangle$, we need to show $\langle A^*u,v\rangle = \langle u,Av\rangle$. This holds because $\langle A^*u,v\rangle = \overline{\langle v,A^*u\rangle} = \overline{\langle Av,u\rangle} = \langle u,Av\rangle$. (The middle equality is the definition of the adjoint, the first and third equalities are the conjugate-symmetry of the inner product.)

Second, the adjoint of a product is the product of the adjoints, in reverse order. $B^*A^* = (AB)^*$, since $\left\langle u, B^*A^*v \right\rangle = \left\langle Bu, A^*v \right\rangle = \left\langle ABu, v \right\rangle$.

Third, the adjoint of the inverse is the inverse of the adjoint. That is, $(A^{-1})^* = (A^*)^{-1}$, since taking $B = A^{-1}$ in the above yields $(A^{-1})^* A^* = (AA^{-1})^* = I$, so $(A^{-1})^*$ is the inverse of A^*.

To see what the adjoint means in terms of coordinates: We write out $\left\langle Au, v \right\rangle = \left\langle u, A^*v \right\rangle$ and $\left\langle Au, v \right\rangle = \left\langle u, A^*v \right\rangle$, and force them to be equal, and look at the consequences for A and A^*. Choose e_k (the vectors whose coordinates have a 1 in the unit position, and 0 elsewhere) as the basis elements and writes A, u and v in coordinates, i.e., $u = \sum_k u_k e_k$ and $v = \sum_k v_k e_k$, so that,

$Au = \sum_j \left(\sum_k A_{jk} u_k \right) e_j$, and $\left\langle Au, v \right\rangle = \sum_j \left(\sum_k A_{jk} u_k \right) \overline{v}_j = \sum_k \left(\sum_j A_{kj} u_j \right) \overline{v}_k$.

Similarly, $A^* v = \sum_j \left(\sum_k A^*{}_{jk} v_k \right) e_j$ and $\left\langle u, A^*v \right\rangle = \sum_j \left(\sum_k \overline{A^*{}_{jk} v_k} \right) u_j$. Since this must be true for all u_j and v_k, it follows that $A_{kj}\overline{v}_k = \overline{A^*{}_{jk} v_k}$, i.e., that $A_{kj} = \overline{A^*{}_{jk}}$. Thus, in coordinates, to find the adjoint, you (a) transpose the matrix (exchange rows with columns), and (b) take the complex-conjugate of its entries. And in the case of $k = \mathbb{R}$, the adjoint is the same as the transpose.

For the translation operator D_T acting on functions of the line, $\left(D_T v \right)(t) = v(t+T)$, the adjoint is $\left(D_T \right)^* = D_{-T}$, i.e., $\left(D_{-T} v \right)(t) = v(t-T)$, since
$\left\langle D_T u, v \right\rangle = \int u(t+T)\overline{v(t)}dt = \int u(t')\overline{v(t'-T)}dt = \left\langle u, D_{-T} v \right\rangle$, where we've made the substitution $t' = t + T$.

The adjoint allows us to define several special kinds of operators. These classes are intrinsic properties of $Hom(V,V)$ for a Hilbert space, i.e., they are defined in a coordinate-free manner but do require the specification of the inner product.

Self-adjoint operators

A "self-adjoint" operator A is an operator for which $A = A^*$. Self-adjoint operators have real eigenvalues, and, to some extent, can be thought of as analogous to real numbers. The fact that self-adjoint operators have real eigenvalues follows from noting that if $Av = \lambda v$, then $\lambda \left\langle v, v \right\rangle = \left\langle \lambda v, v \right\rangle = \left\langle Av, v \right\rangle = \left\langle v, A^*v \right\rangle = \left\langle v, Av \right\rangle = \left\langle v, \lambda v \right\rangle = \overline{\lambda} \left\langle v, v \right\rangle$, so $\lambda = \overline{\lambda}$.

For self-adjoint operators, eigenvectors with different eigenvalues are orthogonal.
Say $Av = \lambda v$ and $Aw = \mu w$, with $\lambda \neq \mu$. Then

$\lambda\langle v,w\rangle = \langle \lambda v,w\rangle = \langle Av,w\rangle = \langle v,A^*w\rangle = \langle v,Aw\rangle = \langle v,\mu w\rangle = \overline{\mu}\langle v,w\rangle$, so $\lambda = \overline{\mu}$ or $\langle v,w\rangle = 0$. Since both λ and μ are real, and they are assumed to be unequal, it follows that $\langle v,w\rangle = 0$.

Unitary operators

A "unitary" operator A is an operator for which $AA^* = A^*A = I$, i.e., their adjoint is equal to their inverse. Unitary operators have eigenvalues whose magnitude is 1, and, to some extent, can be thought of as analogous to rotations, or to complex numbers of magnitude 1. The fact that unitary operators have eigenvalues of magnitude 1 follows from noting that if $Av = \lambda v$, then $|\lambda|^2\langle v,v\rangle = \lambda\overline{\lambda}\langle v,v\rangle = \langle \lambda v,\lambda v\rangle = \langle Av,Av\rangle = \langle v,A^*Av\rangle = \langle v,v\rangle$, so $|\lambda|^2 = 1$.
If the base field is $\mathbb{R}$, then a unitary operator is also a called an orthogonal operator.

For unitary operators, eigenvectors with different eigenvalues are orthogonal.
Say $Av = \lambda v$ and $Aw = \mu w$, with $\lambda \neq \mu$. Then
$\langle v,w\rangle = \langle Av,Aw\rangle = \langle \lambda v,\mu w\rangle = \lambda\overline{\mu}\langle v,w\rangle = \dfrac{\lambda}{\mu}\langle v,w\rangle$ (with the last equality because $\mu\overline{\mu} = |\mu|^2 = 1$). So if $\lambda \neq \mu$, then $\langle v,w\rangle = 0$. Conversely, a self-adjoint operator that has an inverse, and for which all eigenvalues have magnitude unity, is necessarily unitary.

Note that the time-translation operator D_T is unitary, since its adjoint is D_{-T}, which is also its inverse.

Note also that the unitary operators in $Hom(V,V)$ form a group. It is closed under multiplication since $\left((AB)^*\right)^{-1} = \left(B^*A^*\right)^{-1} = \left(A^*\right)^{-1}\left(B^*\right)^{-1} = AB$ (if *A* and *B* have the property that their adjoint is their inverse, then so does *AB*). Inverses are present because if A is unitary, i.e., $A^* = A^{-1}$, then $\left(A^{-1}\right)^* = A^{**} = A = \left(A^{-1}\right)^{-1}$, i.e., A^{-1} is also unitary.

Projection operators

A "projection" operator is a self-adjoint operator P for which $P^2 = P$. One can think of *P* as a (geometric) projection onto a subspace – the subspace that is the range of *P*. It is also natural to consider the complementary projection, $Q = I - P$, as the projection onto the perpendicular (orthogonal) subspace. To see that *Q* is a projection, note $Q^2 = (I-P)^2 = (I-P)(I-P) = I - IP - PI + P^2 = I - P - P + P = I - P = Q$. Also $PQ = P(I-P) = P - P^2 = 0$. Also, the eigenvalues of a projection operator must be 0 or 1. This is because if $Pv = \lambda v$, then $Pv = P^2v = P(Pv) = P(\lambda v) = \lambda^2 v$ also, so $\lambda^2 = \lambda$, which solves only for 0 or 1.

A vector can be decomposed into a component that is in the range of P, and a component that is in the range of Q, and these components are orthogonal.
$v = Iv = (P+Q)v = Pv + Qv$, and $\langle Pv, Qv \rangle = \langle v, PQv \rangle = 0$ -- justifying the interpretation of P and Q as projections onto orthogonal subspaces.

Projections onto one-dimensional subspaces are easy to write. The projection onto the subspace determined by a vector u is the operator $P_u(v) = u\dfrac{\langle v,u \rangle}{\langle u,u \rangle}$.

To see that P_u is self-adjoint, note that $\langle P_u(v), w \rangle = \langle u, w \rangle \dfrac{\langle v,u \rangle}{\langle u,u \rangle} = \dfrac{\langle v,u \rangle \overline{\langle w,u \rangle}}{\langle u,u \rangle}$ but also

$\langle v, P_u(w) \rangle = \left\langle v, u\dfrac{\langle w,u \rangle}{\langle u,u \rangle} \right\rangle = \dfrac{\langle v,u \rangle \overline{\langle w,u \rangle}}{\langle u,u \rangle}$, where the last equality follows because the denominator must be real.

To see that $P_u^2 = P_u$, calculate

$$P_u^2(v) = u\frac{\langle P_u(v), u \rangle}{\langle u,u \rangle} = u\frac{\left\langle u\frac{\langle v,u \rangle}{\langle u,u \rangle}, u \right\rangle}{\langle u,u \rangle} = u\frac{\frac{\langle v,u \rangle}{\langle u,u \rangle}\langle u,u \rangle}{\langle u,u \rangle} = u\frac{\langle v,u \rangle}{\langle u,u \rangle}.$$

This construction can be extended to find projections onto multidimensional subspaces, specified by the range of an operator B. This is the heart of linear regression, and it will be useful for principal components analysis. Assuming that B^*B has an inverse, the projection can be written:

$P_B = B(B^*B)^{-1}B^*$. There's an important piece of fine print here, in that the "inverse" of B^*B is only computed within the range of B.

A comment on how this definition of projection corresponds to the "usual" notion of projection as it applies to images. In the imaging context, one might consider a projection of an image $I(x,y,z)$ onto, say, the (X,Z)-plane by taking an average over all y. In our terms, this is a mapping from a function on a 3-d array of pixels (x,y,z), to a function on a 2-d array of pixels (x,z), i.e., a mapping between two vector spaces – and hence, might seem not to be a projection. But it is, in fact, a projection in our sense too. The functions on the 2-d array of pixels can also be regarded as functions on a 3-d array, but with exactly the same image "slice" for each value of y. The "projection" in our terms is to map $I = I(x,y,z)$ to PI, where the projection is defined by $(PI)(x,y,z) = \dfrac{1}{N_Y}\sum_{y'=1}^{N_Y} I(x,y',z)$.

Normal operators

A "normal" operator is an operator that commutes with its adjoint. Self-adjoint and unitary operators are normal. The only normal operators we will deal with here are either self-adjoint or unitary.

Idempotent operators

An "idempotent" operator is one whose square is itself, i.e., $A^2 = A$. It follows that all eigenvalues of an idempotent operator are 0 or 1, just like for a projection – but operators that are idempotent need not be projections (because idempotent operators can be defined in a space without an inner product.)

Spectral theorem

Statement of theorem: in a Hilbert space, the eigenvectors of a normal operator form a basis. More specifically, the operator A can be written as

$$A = \sum_{\lambda} \lambda P_{\lambda} \tag{6}$$

where P_{λ} is the projection onto the subspace spanned by the eigenvectors of A with eigenvalue λ.

So this guarantees that the eigenvectors $v(t) = e^{i\omega t}$ of D_T form a basis, since D_T is unitary (and therefore, normal). It also tells us why we shouldn't consider (possible) eigenvectors like $v(t) = e^{ct}$ for real c, since they are not in the Hilbert space. It also tells us that we can decompose vectors by their projections onto $v(t) = e^{i\omega t}$ (since they form a basis), and why representing operators in this basis (eq. (6)) results in a simple description of their actions.

For finite-dimensional vector spaces, "typical" operators (i.e., operators corresponding t matrices with randomly-chosen entries), whether they are normal or not, have a full set of eigenvalues and a full set of eigenvectors – just by counting up the roots of the characteristic equation. We can view the spectral theorem and the concept of "normal operators" as a way to guarantee circumstances in which this nice situation applies in the infinite-dimensional case.

But was it "luck" that D_T turned out to be unitary? Was it "luck" that, when the full set of operators was considered together, they had a common set of eigenvectors $v(t) = e^{i\omega t}$, and that there was one for each eigenvalue? Short answer: no, this is because the operators D_T expressed a symmetry of the problem.

The spectral theorem will also help us in another context, matters related to principal components analysis, which also hinges on self-adjoint operators. In contrast to time series analysis (and its generalizations) in which unitary operators arise from *a priori* symmetry considerations, in principal components analysis, self-adjoint operators arise from the data itself.

Group representations

To understand why operators that express symmetries are unitary, and why they have common eigenvectors, and why they (often) have eigenspaces of dimension 1, we need to take a look at "group representation theory." The basic setup is that vector spaces are often functions on a set of points, and if a group acts on a set of points, this induces transformations of the vector space. The transformations in the vector space "represent" the group. It turns out to be not too hard to find all possible representations of a group, and to write them in terms of "prime" (irreducible) representations. An (almost) elementary argument will show that because these representations are "prime", they lead to a way to divide up the vector space, so that in each piece, the group acts in a simple way.

Unitary representations: definition and simple example

A unitary representation U of a group G is a structure-preserving mapping from the group to $Hom(V,V)$. More precisely, it is a group homomorphism from elements g of G into unitary operators U_g in $Hom(V,V)$. If U is an isomorphism, it is called a "faithful" representation.

Note that since U is a group homomorphism, $U_{gh} = U_g U_h$, where gh on the left is interpreted as multiplication in G, and $U_g U_h$ on the right is interpreted as composition in $Hom(V,V)$.

It's worth looking at examples of group representations, since it makes it more impressive to find out that we can write out *all* the representations of a group. (The examples below don't show this; they just show examples of the variety of representations that are possible.)

Example: cyclic groups

Consider the group $\mathbb{Z}_n$ of addition (mod n), and let $V = \mathbb{C}$, i.e., V is the one-dimensional vector space of the complex numbers over itself. Then $U_p = e^{\frac{2\pi i}{n}p}$ is a representation of $\mathbb{Z}_n$. To check that it is a homomorphism, note that $U_p U_q = e^{\frac{2\pi i}{n}p} e^{\frac{2\pi i}{n}q} = e^{\frac{2\pi i}{n}(p+q)} = U_{p+q}$. U_p is an isomorphism provided that $p \neq 0 \pmod{n}$.

Example: the translation group on the line

Consider (again) time-shifts D_T acting on functions on the line by $(D_T v)(t) = v(t+T)$. This is a unitary representation, of the group of shifts on a line, in the vector space V of functions on the line.

Example: the dihedral group

The dihedral group D_n is the set of rotations and reflections of a regular n-gon. We can write out each of these rotations as a 2-d matrix, and obtain a 2-dimensional representation of the group.

Example: permutation groups

If a group is presented as a permutation group, we can form unitary representations another way. We can write these permutations as permutation matrices, i.e., matrices that are mostly 0's, with a 1 in position (j,k) if element in position j is moved to position k. This is a representation too, as composing the permutations is equivalent to composing the matrices.

For the above cases (dihedral groups and permutation groups), we did not have to check that the representations were in terms of unitary operators. This is because we were dealing with a finite group. In a finite group, every element g has an "order" m, i.e., a least positive integer for which $g^m = e$. If some operator L_g represents g, then (since the group representation preserves structure) $\left(L_g\right)^m = I$. Immediately, this means that any eigenvalue of L_g must satisfy $\lambda^m = 1$. Self-adjointness also follows – for the permutation group, since L_g is simply a re-ordering, for the dihedral group, because the L_g are rotations and reflections.

For infinite groups, we don't have $\left(L_g\right)^m = I$. But explicitly requiring that the that L_g are unitary, along with the Hilbert space structure, allows the main results from finite groups to generalize.

Example: the trivial representation

Finally, there is always the "trivial" representation, that takes every group element to the identity map on $Hom(V,V)$.

The character

The character χ_L of a representation L is a function from the group to the field. It is defined in terms of the trace: $\chi_L(g) = \mathrm{tr}(L_g)$. As noted above, the trace is the sum of the eigenvalues. In coordinates, this means that the trace is the sum of the diagonal elements.

A simple consequence of this is that the character is invariant with respect to inner automorphisms (mappings from the group to itself given by $g \to hgh^{-1}$). That is, $\chi_L(hgh^{-1}) = \chi_L(g)$. To see this, $\chi_L(hgh^{-1}) = \mathrm{tr}(L_{hgh^{-1}}) = \mathrm{tr}(L_h L_g L_{h^{-1}}) = \mathrm{tr}(L_h L_g (L_h)^{-1}) = \mathrm{tr}(L_g) = \chi_L(g)$, where we have used the fact that L preserves structure, and that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. Another way of putting this is that the character is constant on every "conjugate class" – the "conjugate class" of g is, by definition, the group elements of the form hgh^{-1}. Note that these observations become vacuous in a commutative group, since $g = hgh^{-1}$ so every conjugate class has just one element.

A couple of easy facts about characters: (i), The character of the trivial representation on a vector space V is equal to the dimension of V, since every group element is represented by the identity in V. (ii) The character of any representation at the identity element is the dimension of the representation, since the representation at the identity element is the identity matrix.

For the translation group on the line, the character of the nontrivial *irreducible* representations (i.e., the representations that cannot be broken down into smaller components –see below) will turn out to be the Fourier coefficients.

Combining representations: Direct sum

Two representations of the same group, U_1 in V_1 and U_2 in V_2, can be combined to make a composite representation in $V_1 \oplus V_2$. A group element g maps to $U_{1,g} \oplus U_{2,g}$, where $U_{1,g} \oplus U_{2,g}$ acts on $v_1 \oplus v_2$ in the obvious way: $\left(U_{1,g} \oplus U_{2,g}\right)(v_1 \oplus v_2) = U_{1,g}(v_1) \oplus U_{2,g}(v_2)$.

We have to check that $U_{1,g} \oplus U_{2,g}$ is unitary. First we need an inner product on the space $V_1 \oplus V_2$. It is natural to take $\langle v_1 \oplus v_2, w_1 \oplus w_2 \rangle = \langle v_1, w_1 \rangle_{V_1} + \langle v_2, w_2 \rangle_{V_2}$, where the two terms on the right-hand-side are inner products in V_1 and V_2 respectively. To show that $U_{1,g} \oplus U_{2,g}$ is unitary, we need $\left\langle \left(U_{1,g} \oplus U_{2,g}\right)(v_1 \oplus v_2), \left(U_{1,g} \oplus U_{2,g}\right)(w_1 \oplus w_2) \right\rangle = \left\langle (v_1 \oplus v_2), (w_1 \oplus w_2) \right\rangle$. This follows because

$$
\begin{aligned}
&\left\langle \left(U_{1,g} \oplus U_{2,g}\right)(v_1 \oplus v_2), \left(U_{1,g} \oplus U_{2,g}\right)(w_1 \oplus w_2) \right\rangle \\
&= \left\langle (U_{1,g} v_1 \oplus U_{2,g} v_2), (U_{1,g} w_1 \oplus U_{2,g} w_2) \right\rangle \\
&= \left\langle U_{1,g} v_1, U_{1,g} w_1 \right\rangle_{V_1} + \left\langle (U_{2,g} v_2 \otimes U_{2,g} w_2) \right\rangle_{V_2} \qquad , \\
&= \langle v_1, w_1 \rangle_{V_1} + \langle v_2, w_2 \rangle_{V_2} \\
&= \langle v_1 \oplus v_2, w_1 \oplus w_2 \rangle
\end{aligned}
$$

where the equalities follow from (i) the definition of $U_{1,g} \oplus U_{2,g}$, (ii) the definition of the inner product on $V_1 \oplus V_2$, (iii) the fact that $U_{1,g}$ and $U_{2,g}$ are unitary in V_1 and V_2, respectively, and (iv)again the definition of the inner product on $V_1 \oplus V_2$, but this time putting the pieces back together.

We can use general statements about how operators extend to direct sums to determine the characters of the composite representation L. Since $\chi_L(g) = \mathrm{tr}(L_g)$, we need to determine the sum of the eigenvalues of L_g. For a direct sum, the eigenvectors of $U_{1,g} \oplus U_{2,g}$ are $\nu_1 \oplus 0$, with eigenvalue λ_1, and $0 \oplus \nu_2$, with eigenvalue λ_2 (where v_j is an eigenvector of $U_{j,g}$ with eigenvalue λ_j, etc.). So each eigenvalue of $U_{1,g}$ and $U_{2,g}$ contributes once. So, $\chi_{U_1 \oplus U_2}(g) = \mathrm{tr}(U_{1,g}) + \mathrm{tr}(U_{2,g}) = \chi_{U_1}(g) + \chi_{U_2}(g)$, i.e., $\chi_{U_1 \oplus U_2} = \chi_{U_1} + \chi_{U_2}$.

Combining representations: Tensor product

We can also define a group representation on $V_1 \otimes V_2$ in the same way: $\left(U_{1,g} \otimes U_{2,g}\right)(v_1 \otimes v_2) = U_{1,g}(v_1) \otimes U_{2,g}(v_2)$. We also have to first check that this is unitary, which also requires defining an inner product on $V_1 \otimes V_2$. We take $\langle v_1 \otimes v_2, w_1 \otimes w_2 \rangle = \langle v_1, w_1 \rangle_{V_1} \langle v_2, w_2 \rangle_{V_2}$. Note that this definition respects the defining relationship of the tensor product space, namely, $a(v_1 \otimes v_2) = (av_1) \otimes v_2 = v_1 \otimes (av_2)$; the "direct sum" definition would not have done this.

The unitary nature of $U_{1,g} \otimes U_{2,g}$ follows from a calculation analogous to the one for $U_{1,g} \oplus U_{2,g}$:

$$\begin{aligned}
&\left\langle \left(U_{1,g} \otimes U_{2,g}\right)(v_1 \otimes v_2), \left(U_{1,g} \otimes U_{2,g}\right)(w_1 \otimes w_2) \right\rangle \\
&= \left\langle (U_{1,g} v_1 \otimes U_{2,g} v_2), (U_{1,g} w_1 \otimes U_{2,g} w_2) \right\rangle \\
&= \left\langle U_{1,g} v_1, U_{1,g} w_1 \right\rangle_{V_1} \left\langle (U_{2,g} v_2, U_{2,g} w_2) \right\rangle_{V_2} \qquad , \\
&= \langle v_1, w_1 \rangle_{V_1} \langle v_2, w_2 \rangle_{V_2} \\
&= \langle v_1 \otimes v_2, w_1 \otimes w_2 \rangle
\end{aligned}$$

where the equalities follow from (i) the definition of $U_{1,g} \otimes U_{2,g}$, (ii) the definition of the inner product on $V_1 \otimes V_2$, (iii) the fact that $U_{1,g}$ and $U_{2,g}$ are unitary in V_1 and V_2, respectively, and (iv)again the definition of the inner product on $V_1 \otimes V_2$, but this time putting the pieces back together.

For a tensor product, the eigenvectors of $U_{1,g} \otimes U_{2,g}$ are $\nu_1 \otimes \nu_2$, with eigenvalue $\lambda_1 \lambda_2$. So every product of eigenvalues, one from V_1 and one from V_2, contributes. So, $\chi_{U_1 \otimes U_2}(g) = \mathrm{tr}(U_{1,g})\,\mathrm{tr}(U_{2,g}) = \chi_{U_1}(g)\chi_{U_2}(g)$, i.e., $\chi_{U_1 \otimes U_2} = \chi_{U_1}\chi_{U_2}$.

The regular representation

The "regular representation" is a representation that we are guaranteed to have for any group, and it arises from considering how the group acts on functions on a set, when the set itself is the group. To build the regular representation: Let V be the vector space of functions $x(g)$ from G to $\mathbb{C}$. (This is the "free vector space" on G). We can make V into a Hilbert space by defining $\langle x, y \rangle = \sum_G x(g)\overline{y(g)}$.

Note that this makes sense for infinite groups – our Hilbert space then consists of functions on the group for which the inner product of a function with itself is finite. For infinite but discrete groups (such as the integers, under addition), the above expression works fine. For infinite but continuous groups (such as the reals, under addition), we instead use $\langle x,y\rangle = \int_G x(g)\overline{y(g)}dg$.

We define the regular representation *R* as follows:

For each element *p* of *G*, we need to define R_p, a member of $Hom(V,V)$. R_p takes *x* (a function on *G*) to the $R_p(x)$ (another function on *G*) whose value at *g* is given by

$$\left(R_p(x)\right)(g) = x(gp). \tag{7}$$

To see that R_p is unitary:
$\langle R_p(x), R_p(y)\rangle = \sum_{g\in G}\left(R_p(x)\right)(g)\overline{\left(R_p(y)\right)(g)} = \sum_{g\in G} x(gp)\overline{y(gp)} = \sum_{h\in G} x(h)\overline{y(h)}$. The reason for the final equality is that as *g* traverses *G*, then so does *gp* (but in a different order). Formally, change variables to $h = gp$; $g = hp^{-1}$ if hp^{-1} takes each value in *G* once, then so does *h*.

To see that R_p is a representation – i.e., that $R_pR_q = R_{pq}$: Here we are using the convention that R_pR_qx means, "apply R_p to the result of applying R_q to *x*". So we need to show that $R_p\left(R_q(x)\right) = R_{pq}(x)$ by evaluating the left and right hand side at every group element *g*.
On the left, say $y = R_q(x)$, so $y(g) = \left(R_q(x)\right)(g) = x(gq)$. Then $\left(R_p(y)\right)(g) = y(gp) = x(gpq)$. On the right, $\left(R_{pq}(x)\right)(g) = x(gpq)$.

Note that time translation as defined by (3) is an example of this: it is the regular representation of the additive group of the real numbers.

For a finite group, we can readily determine the character of the regular representation, as follows. We choose, as a basis for V, the functions on the group v_q, where $v_q(q) = 1$ and $v_q(g) = 0$ for $g \neq q$. R_p acts on V by permuting the v_q's: $(R_pv_q)(g) = v_q(gp)$, which is nonzero only at $gp = q$, i.e., $g = qp^{-1}$. So $R_pv_q = v_{qp^{-1}}$. If $p = e$, the identity, then every v_q is mapped to itself, i.e., $\mathrm{tr}(R_e) = \dim(V)$. But if $p \neq e$, every v_q is mapped to a different $v_{qp^{-1}}$. Viewed as a permutation matrix, R_p therefore must have its diagonal all 0's. So its trace is 0. Thus, for the regular representation, the character $\chi_R(s)$ is equal to 0 for all elements except the identity, and $\chi_R(e) = |G|$.

Irreducible Representations

An "irreducible representation" V is one that cannot be broken down into a direct sum of two representations. One-dimensional representations, such as in the example for $\mathbb{Z}_n$, are necessarily irreducible.

Less obviously, for a commutative group – such as the translations of the line -- every irreducible representation is one-dimensional. The reason is the following. Let's say you had some representation L of a commutative group that was of dimension 2 or more, and two group elements, say, g and h, for which the linear transformations L_g and L_h did *not* have a common set of eigenvectors. (If all L_x had a common set of eigenvectors, we could use this as a basis and decompose L into one-dimensional components.) If L_g and L_h did *not* have a common set of eigenvectors, then it would have to be that $L_g L_h \neq L_h L_g$, which would be a contradiction because $L_g L_h = L_{gh} = L_{hg} = L_h L_g$.

There is another characterization of an irreducible representation that will be useful below: if L is irreducible on V, then the only operators that commute with all L_g are 0 and the identity. For if some other linear operator commuted with all of the L_g, then, for each of its eigenvectors, the corresponding eigenspace would provide a way to reduce L.

Note also that a representation on a group always provides a representation on any subgroup. However, an *irreducible* representation may become reducible on a smaller subgroup – consider the trivial case in which the subgroup is the identity.

Breaking down a representation

As a first step in breaking down a representation into irreducible pieces, we can ask whether there is any part of it that is trivial. That is, does V have a subspace, say W, for which L_g acts like the identity element? It turns out that we can find W by creating a projection P_L from V onto W. We define this projection as follows:

$$P_L(v) = \frac{1}{|G|}\sum_g L_g(v). \tag{8}$$

That is, we let every L_g act on a vector v, and average the result. Intuitively, the average vector $P_L(v)$ cannot be altered by any further group action, e.g., by some L_h, and this makes P_L a projection. To show that P_L is a projection formally:

$P_L(L_h v) = \frac{1}{|G|}\sum_g L_g(L_h v) = \frac{1}{|G|}\sum_g L_{gh}(v) = \frac{1}{|G|}\sum_u L_u(v) = P_L(v)$, where in the next-to-the last step we've replaced $u = gh$, and observed that letting g run over all of G is the same as letting $u = gh$ run over all of G.

The trace of a projection P is the dimension of the space that it projects onto. That is because, when expressed in the basis of its eigenvalues, it looks like $\begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$, where the 1's correspond to basis vectors that are unchanged by P (and span its range), and the 0's correspond to the basis vectors that are set to 0.

So the dimension of W, the space on which L acts trivially, is given by $\mathrm{tr}(P_L)$. This yields the "trace formula":

$$\mathrm{tr}(P_L(v)) = \frac{1}{|G|}\sum_g \mathrm{tr}(L_g(v)) = \frac{1}{|G|}\sum_g \chi_L(g). \tag{9}$$

Finding parts of one representation inside another

With one more step, we will see that the regular representation contains all irreducible representations of G. The step is to use the above trace formula -- which tells how many copies of the identity representation are inside a given representation – in a way that counts the number of copies of irreducible parts of a representation L that match irreducible parts of a second representation M.

As motivation, here is an alternative characterization of an irreducible representation: L is an irreducible representation of G in V if the only elements of $Hom(V,V)$ that commute with all of the L_g are multiples of the identity. To see why this is true: if we could decompose $L = M_1 \oplus M_2$, where each M_i acted in a subspace V_i with $V = V_1 \oplus V_2$, then projections onto either V_i would commute with L_g. Conversely, say that $\varphi \in Hom(V,V)$ commuted with all the L_g, and that φ is not a multiple of the identity. Then take an eigenvalue λ of φ. $A = \varphi - I\lambda$ also commutes with all the L_g, and is not zero (since φ is not a multiple of the identity). So its null space (what it maps to zero) is less than all of V. This null space is preserved by all of the L_g, since $AL_g v = L_g A v$, which is zero whenever $Av = 0$. So the commuting operator φ has yielded a subspace of V that is invariant under all the L_g, and therefore demonstrates reducibility.

The more general setup: a representation L of G in V (i.e., for each group element g, a unitary transformation L_g in $Hom(V,V)$, and another representation M of G in W (i.e., for each group element g, a unitary transformation M_g in $Hom(W,W)$. The statement that there is a part of L

that corresponds to a part of *M* can be formalized by saying that there is a linear map φ in $Hom(V,W)$ for which $\varphi L_g = M_g \varphi$. That is, letting *L* act on a vector *v* in *V*, and then finding the image of $L_g(v)$ in *W*, is the same as finding the image $\varphi(v)$ of *v* in *W*, and letting M_g act on it.

So we want to know, whether any such φ's exist. To answer this, we will first construct a group representation Φ in $Hom(V,W)$ that builds on *L* and *M*. We will then show that if Φ acts like the identity representation on some homomorphism φ in $Hom(V,W)$ (i.e., if $\Phi_g(\varphi) = \varphi$ for all group elements *g*), then φ finds subspaces of *V* and *W* in which *L* and *M* act identically.

The above setup allows us to define the needed group representation Φ in $Hom(V,W)$. Φ is specified by the transformations Φ_g, each of which maps elements ϕ of $Hom(V,W)$ into other element $\Phi_g(\phi)$. And to specify $\Phi_g(\phi)$, we need to specify its action on any *v*. We choose:

$\Phi_g(\phi)(v) = M_g \phi L_g^{-1}(v)$. There are a few things to check – the most important of which is that it is a representation. That is, does $\Phi_{hg} = \Phi_h \Phi_g$? Making use of the fact that both *M* and *L* are group representations:

$$\Phi_{hg}(\phi)(v) = M_{hg}\phi L_{hg}^{-1}(v) = M_{hg}\phi L_{(hg)^{-1}}(v) = M_{hg}\phi L_{g^{-1}h^{-1}}(v) = M_h M_g \phi L_{g^{-1}} L_{h^{-1}}(v)$$
$$= M_h \Phi_g(\phi) L_{h^{-1}}(v) = \Phi_h\left(\Phi_g(\phi)\right)(v)$$

Notice that if there is some ϕ in $Hom(V,W)$ that every Φ_g leaves invariant (i.e., Φ_g acts like the identity on ϕ for all *g*), then $M_g \phi L_g^{-1} = \Phi_g(\phi) = \phi$, and therefore that $M_g \phi = \phi L_g$. Put another way, each operator in $Hom(V,W)$ for which Φ acts like the identity corresponds to a way of matching a component of *L* to a component of *M*. That is, the dimension of this space, which we will call $d(L,M)$, is the number of ways we can match components of *L* to components of *M* that preserve the action of ϕ. Since this dimension is the size of the space in which Φ acts trivially, we can find it by applying the trace formula, eq. (9), to Φ:

$$d(L,M) = \frac{1}{|G|}\sum_g \chi_\Phi(g).$$

So now we need to calculate χ_Φ, where Φ is built from *L* and *M* as described above. Just like we could evaluate $\chi_{L\otimes M}$ from $\mathrm{tr}(L_g \otimes M_g) = \mathrm{tr}(L_g)\,\mathrm{tr}(M_g)$ to get $\chi_{L\otimes M} = \chi_L \chi_M$, we can do the same for the representation Φ constructed in $Hom(V,W)$. The observation we need to do this is that there is a correspondence between $Hom(V,W)$ and $V^* \otimes W$.

The germ of the idea is as follows (see Q2 of Homework 4, 2016-2017 Linear Transformations and Group Representations, or Q2 of Homework 3, 2024-2035 Groups, Fields, and Vector Spaces). It is easier to show the correspondence in the reverse direction, i.e., for each element of $V^* \otimes W$, to find an element of $Hom(V,W)$. So say we have $\psi \otimes w$ in $V^* \otimes W$, where $\psi \in V^*$

and $w \in W$. We can then define the corresponding element $B(\psi)$ in $Hom(V,W)$ as the homomorphism that takes $v \in V$ to $\psi(v)w$. So we have a correspondence between elementary tensor products, and homomorphisms whose range is one-dimensional. We then need to check that this correspondence respects the way that tensors add to each other, and the way that homomorphisms add to each other, and we need to show that the correspondence can be inverted. We skip these details here.

Since Φ is a representation in $Hom(V,W)$, the correspondence between $Hom(V,W)$ and $V^* \otimes W$ gives us a representation $L^* \otimes M$ in $V^* \otimes W$. We then can calculate $\chi_\Phi(g) = \chi_{L^* \otimes M}(g) = \overline{\chi_L(g)}\chi_M(g)$.

Averaging $\chi_\Phi(g)$ over the group gives the number of times that Φ contains the identity, which – as we saw above – is the number of ways that parts of *L* can be found inside of *M*. This yields our main result:

$$d(L,M) = \frac{1}{|G|}\sum_g \overline{\chi_L(g)}\chi_M(g). \tag{10}$$

As a special case for $L = M$:

$$d(L,L) = \frac{1}{|G|}\sum_g |\chi_L(g)|^2. \tag{11}$$

The group representation theorem

While we have carried this analysis out for finite groups, everything we've done leading up to eq. (10) also works for infinite groups, provided that we can set up a Hilbert space in the functions on them (which amounts to being able to define integrals, so that there is a dot-product).

By applying eq. (10) to a few special cases, we obtain all the main properties of group representations, which are summarized in the "group representation theorem"– and which formalizes the non-accidental nature of the Fourier transform.

Recall that an "irreducible representation" is a representation cannot be written as a direct sum of group representations.

Here are the facts:

- The characters of irreducible representations are orthogonal. This follows from eq. (10) directly, since (according to the definition of irreducible representations), if *L* and *M* are two different irreducible representations, $d(L,M) = 0$.

- The character of an irreducible representation is an orthonormal function on the group. This follows from eq. (11), since in this case, $d(L,L) = 1$. It is standard to write the collect the characters into a "character table", in which the rows correspond to the individual irreducible representations and the columns correspond to the conjugate classes (since characters are constant on conjugate classes). In terms of the character table, this is "row orthonormality".

- *Every* irreducible representation L occurs in the regular representation, and the number of occurrences is equal to the dimension of L. This follows from eq. (10) by taking M to be the regular representation, R. The character of the regular representation is 0 for all group elements except the identity, and is $|G|$ at the identity. So the only term that contributes to the sum is the term for $g = e$. $\chi_L(e)$ is the dimension of L, since the representation of e is the identity matrix (and the trace just adds up the 1's on the diagonal).

- For a finite group, the sum of the squares of the dimensions of the distinct irreducible representations is the size of the group. This follows from eq. (10) with L and M both the regular representation: the right side of eq. (10) is the size of the group (the only term that contributes is the identity), and the left side contributes d^2 for every irreducible representation of dimension d.

- For a finite group, the number of distinct irreducible representations is the number of conjugacy classes. This follows from counting in the character table: the orthonormality of the rows (one row for each irreducible representation), and the orthogonality of the columns (one row for each conjugate class).

For finite, commutative groups, we can go further very easily by counting dimensions. Since every irreducible representation is one-dimensional, the number of different irreducible representations must be $|G|$. Thus, the characters of the irreducible representations form an orthonormal basis for functions on the group.

Algebraically, nothing changes when one goes from finite groups to infinite ones, but there are things to prove (about limits, integrals, etc.), which ultimately were the reasons we needed the Hilbert space structure.

Ignoring these "details", we apply the above to the additive group of the real numbers. Its regular representation is the time translation operators defined by eq. (3). All irreducible representations must be one-dimensional. Above we showed that each representation must be of the form $T \to e^{i\omega T}$. So this is the full set, and we have decomposed space of the regular representation (the space of functions of time) into one-dimensional subspaces, in which time translation by T acts like multiplication by $e^{i\omega T}$,

For finite but non-commutative groups, it is a bit more complex. There will always be some conjugate classes with more than one element, since there will be always some choice of g and h for which $hgh^{-1} \neq g$. So there have to be fewer conjugate classes than $|G|$ (since at least one of the conjugate classes has two or more elements). So there have to be fewer distinct irreducible representations than $|G|$, since their characters must be orthogonal (and hence, linearly independent) functions on the conjugate classes.

Orthonormal basis

With a bit more work, one can show that the matrix elements of the irreducible representations are orthogonal functions on the group, and, for an irreducible representation M, they can be made orthonormal by multiplying by a factor of $\sqrt{\dim M}$. The idea is to look at the tensor product of two irreducible representations $M \otimes \overline{N}$, and average it over the group. ($\overline{N}$ is the representation obtained by complex conjugation of all the elements of N; it is guaranteed to be a representation because complex conjugation is an automorphism of the base field.)

Note that matrix elements of $M \otimes \overline{N}$, summed over the group, correspond to inner products of matrix elements of M and N. Say v_i is an orthonormal basis for the vector space in which M acts, and w_k is an orthonormal basis for the vector space in which $\overline{N}$ acts, and that $m_{i,j}$ is the coefficient of v_j in Mv_i, and that $\overline{n_{k,l}}$ is the coefficient of w_l in $\overline{N}w_k$. Then, we can write the coefficient of $v_j \otimes w_l$ in $v_i \otimes w_k$ as $c_{ik,jl} = m_{i,j}\overline{n_{k.l}}$, where $m_{i,j}$, $\overline{n_{k,l}}$, and $c_{ik,jl}$ depend on the group element. But if M and N are distinct irreducible representations, then (by the argument used for the group representation theorem) then $\frac{1}{|G|}\sum_{g\in G}\overline{\chi_M(g)}\chi_N(g) = 0$. But $\frac{1}{|G|}\sum_{g\in G}\overline{\chi_M(g)}\chi_N(g) = \frac{1}{|G|}\sum_{g\in G}\overline{\chi_{M\otimes\overline{N}}(g)}$, which is the dimension of the space in which $M \otimes \overline{N}$ acts as the identity. Averaging $M \otimes \overline{N}$ over the group yields a projection onto this subspace, so the average of $M \otimes \overline{N}$ over the group must also be zero. And its coefficients are $\frac{1}{|G|}\sum_{g\in G}c_{ik,jl}(g) = \frac{1}{|G|}\sum_{g\in G}m_{i,j}(g)\overline{n_{k.l}}(g)$, which are all the inner products of the matrix elements of M and N.

But if $M = N$, then $M \otimes \overline{M}$, then, since the characters are orthonormal functions, $\frac{1}{|G|}\sum_{g\in G}\overline{\chi_M(g)}\chi_M(g) = 1$. So there is a one-dimensional subspace in which $M \otimes \overline{M}$ acts as the identity. Since each $M(g)$ is unitary, we can immediately find it: $\sum_{i=1}^{\dim V} v_i \otimes v_i$. In coordinates (using $minv_{k,i}(g)$ to denote the elements of the matrix inverse of M_g)

$$\left(M_g \otimes \overline{M}_g\right)\left(\sum_{i=1}^{\dim V} v_i \otimes v_i\right) = \sum_{j,k}^{\dim V}\left(\sum_{i=1}^{\dim V} v_j \otimes v_k\right)m_{i,j}(g)\overline{m}_{i,k}(g)$$

$$= \sum_{j,k}^{\dim V}\left(\sum_{i=1}^{\dim V} v_j \otimes v_k\right)m_{i,j}(g)minv_{k,i}(g) = \sum_{j,k}^{\dim V}(v_j \otimes v_k)\delta_{j,k} = \sum_j v_j \otimes v_j \,.$$

As in the previous case of $M \neq N$, the inner products of the elements of the matrix M with itself are the average of $M \otimes \overline{M}$ over the group, but it is no longer zero, but rather, the projection onto the subspace $\sum_{i=1}^{\dim V} v_i \otimes v_i$. This projection takes $v_k \otimes v_k$ to $\frac{1}{\dim V}\sum_{i=1}^{\dim V} v_i \otimes v_i$, and $v_j \otimes v_k$ ($j \neq k$) to zero, i.e., it is a projection P whose elements $p_{ki,ki} = \frac{1}{\dim V}$ and all other elements are zero. (Verify that P projects onto that space, and that it is idempotent – and recall that there is a unique correspondence between subspaces and projections.) So, the inner products of a matrix element of M with itself is $\frac{1}{\dim V} = \frac{1}{\dim M}$.

There's an intuition underlying this. Tensor products of vectors are generalizations of products of scalar variables. The unitary property is that inner products, and hence distances, are preserved. So, observing that $M \otimes \overline{M}$ preserves $\sum_{i=1}^{\dim V} v_i \otimes v_i$ corresponds to observing that a (squared) distance $\sum x_i^2$ is preserved.

Column orthogonality

Another important consequence is "column orthogonality" of the characters: for every pair of group elements g and h that are in different conjugacy classes, then, $\sum_L \chi_L(g)\chi_L(h) = 0$, where the sum is over all irreducible representations. This is sometimes referred to as "column orthogonality" of the character table.

The first step to see this is the following: for an irreducible representation L, consider $A = \frac{1}{|G|}\sum_{s \in G} L_{s^{-1}hs}$. This commutes with every L_g:

$AL_g = \frac{1}{|G|}\sum_{s \in G} L_{s^{-1}hs} L_g = \frac{1}{|G|}\sum_{s \in G} L_{s^{-1}hsg} = \frac{1}{|G|}\sum_{t \in G} L_{gt^{-1}ht} = \frac{1}{|G|}\sum_{t \in G} L_g L_{t^{-1}ht} = L_g A$, where we used the substitution $s = tg^{-1}$, so $s^{-1} = gt^{-1}$, and the usual observation that as s runs over the group, so does $t = sg$. So $A = \frac{1}{|G|}\sum_{s \in G} L_{s^{-1}hs}$ is a multiple of the identity. Now observe that as s ranges over the group, $tr(L_{s^{-1}hs}) = \chi_L(h)$, independent of s (because $L_{s^{-1}hs} = L_{s^{-1}} L_h L_s = \left(L_s\right)^{-1} L_h L_s$. So it follows that $\frac{1}{|G|}\sum_{s \in G} L_{s^{-1}hs} = \frac{1}{\dim L}\chi_L(h) I$.

Now we can show column orthogonality. For any two group elements g and h, let $C(g,h)$ be the number of elements s of G for which $s^{-1}gs = h$. $C(g,h) = 0$ if g and h are in different conjugate classes, and a positive integer if they are in the same conjugate class. If $s^{-1}gs = h$, then $s^{-1}gsh^{-1}$ is the identity, and $R_{s^{-1}gsh^{-1}}$, the regular representation at $sgs^{-1}h^{-1}$, is a $|G|\times|G|$ identity matrix, whose trace is $|G|$. But if $s^{-1}gsh^{-1}$ is not the identity then $R_{s^{-1}gsh^{-1}}$ is a permutation matrix that leaves no element unchanged, so its diagonal is all zeros. Consequently,

$$C(g,h)=\frac{1}{|G|}\sum_{s\in G}tr(R_{s^{-1}gsh^{-1}}).$$

Since the trace is invariant under a change of basis, we can choose the basis in which the regular representation is block-diagonalized into its component irreducible representations. The group representation theorem tells us that each irreducible representation L occurs in the regular representation with a multiplicity equal to $\dim(L)$.

$C(g,h)=\frac{1}{|G|}\sum_{s\in G}tr(R_{s^{-1}gsh^{-1}})=\frac{1}{|G|}\sum_{s\in G}\sum_{L}(\dim L)tr(L_{s^{-1}gsh^{-1}})$, where the inner sum is over all irreducible representations of G.

$$C(g,h)=\frac{1}{|G|}\sum_{s\in G}\sum_{L}(\dim L)tr(L_{s^{-1}gsh^{-1}})=\frac{1}{|G|}\sum_{s\in G}\sum_{L}(\dim L)tr(L_{s^{-1}gs}L_{h^{-1}})$$

$$=\frac{1}{|G|}\sum_{L}(\dim L)\sum_{s\in G}tr(L_{s^{-1}gs}L_{h^{-1}})=\frac{1}{|G|}\sum_{L}(\dim L)tr(\sum_{s\in G}L_{s^{-1}gs}L_{h^{-1}})$$

We now calculate this sum using the above result that $\frac{1}{|G|}\sum_{s\in G}L_{s^{-1}hs}=\frac{1}{\dim L}\chi_L(h)I$:

$$C(g,h)=\frac{1}{|G|}\sum_{L}(\dim L)tr(\sum_{s\in G}L_{s^{-1}gs}L_{h^{-1}})=\sum_{L}tr(\chi_L(g)IL_{h^{-1}})=\sum_{L}\chi_L(g)tr(L_{h^{-1}})$$

$$=\sum_{L}\chi_L(g)\overline{\chi_L(h)}$$

, i.e., the list of characters of non-conjugate elements are orthogonal functions.

Projecting onto the space of an irreducible representation

Column orthogonality also leads to a generalization of eq. (8) to an expression that projects onto the space in which a (possibly reducible) representation L acts as a specific irreducible representation M (taking M to be the trivial representation recovers eq. (8)).

$$P_{M,L}(v)=\frac{\dim M}{|G|}\sum_{g}L_g(v)\overline{\chi_M(g)}. \qquad (12)$$

This is Fourier inversion in a general context: it gives the projection onto the subspace corresponding to a given irreducible representation.

Several observations: (i) $P_{M,L}$ is a projection, (ii) the sum of these projections is the identity, i.e., it is a complete decomposition, (iii) $P_{M,L}$ commutes with each of the $L_s(v)$, i.e., the range of $P_{M,L}$ is invariant under L, (iv) the ranges of $P_{M,L}$ and $P_{N,L}$ $M \neq N$) are orthogonal. We sketch these.

(i) and (iv):

$$P_{M,L}\left(P_{N,L}(v)\right)=\frac{\dim M \dim N}{|G|^2}\sum_{g,h} L_g\left(L_h(v)\right)\overline{\chi_M(g)}\overline{\chi_N(h)}$$

$$=\frac{\dim M \dim N}{|G|^2}\sum_{g,h} L_{gh}(v)\overline{\chi_M(g)}\overline{\chi_N(h)}$$

$$=\frac{\dim M \dim N}{|G|^2}\sum_{h,r} L_r(v)\overline{\chi_M(rh^{-1})}\overline{\chi_N(h)}=\frac{\dim M \dim N}{|G|^2}\sum_{g,r} L_r(v)tr\overline{M_{rh^{-1}}}\,tr\overline{N_h},$$

$$=\frac{\dim M \dim N}{|G|^2}\sum_{h,r} L_r(v)tr\overline{M_r}\,\overline{M_{h^{-1}}}\otimes\overline{N_h}$$

$$=\frac{\dim M \dim N}{|G|^2}\sum_{h,r} L_r(v)tr\overline{M_r}M_h\otimes\overline{N_h}$$

where we have used the structure-preserving nature of representations (step 1), substituted $r=gh$ (step 2), relations between trace and character (step 3), relation between product of trace and trace of tensor product (step 4), and unitariness of M (step 5). Now sum over h. The argument (see "orthonormal basis" section, or material leading up to eq. (10)) means that the average over the group of $M_h\otimes\overline{N_h}$ is zero, unless $M=N$. But if $M=N$, (see "Orthonormal basis" section), the group-average is $\frac{1}{\dim M}$ (i.e., the sum over the group is $\frac{|G|}{\dim M}$), so

$$P_{M,L}\left(P_{M,L}(v)\right)=\frac{\left(\dim M\right)^2}{|G|^2}\sum_{r} L_r(v)tr\overline{M_r}\left(\frac{|G|}{\dim M}\right)=\frac{\dim M}{|G|}\sum_{r} L_r(v)tr\overline{M_r}$$

$$=\frac{\dim M}{|G|}\sum_{r} L_r(v)\overline{\chi_M(r)}=P_{M,L}(v)$$

And (ii):

$$\sum_M P_{M,L}(v)=\sum_M\frac{\dim M}{|G|}\sum_g L_g(v)\overline{\chi_M(g)}=\frac{1}{|G|}\sum_g L_g(v)\sum_M\left(\dim M\right)\overline{\chi_M(g)}.$$

But $\dim M=\chi_M(e)$. By column orthogonality,

$$\sum_M\left(\dim M\right)\overline{\chi_M(g)}=\sum_M\chi_M(e)\overline{\chi_M(g)}=\begin{cases}0, & g\neq e\\ \sum_M\left(\dim M\right)^2=|G|, & g=e\end{cases}.$$ So

$$\sum_M P_{M,L}(v)=\frac{1}{|G|}L_e(v)|G|=L_e(v)=v.$$

And (iii):

$$P_{M,L}\left(L_s(v)\right)=\frac{\dim M}{|G|}\sum_g L_g\left(L_s(v)\right)\overline{\chi_M(g)}=\frac{\dim M}{|G|}\sum_g L_{gs}(v)\overline{\chi_M(g)},$$

while (with $r=sgs^{-1},\ g=s^{-1}rs$)

$$L_s\left(P_{M,L}(v)\right)=\frac{\dim M}{|G|}\sum_g L_s\left(L_g(v)\right)\overline{\chi_M(g)}=\frac{\dim M}{|G|}\sum_r L_{rs}(v)\overline{\chi_M(s^{-1}rs)}.$$

The two expressions are equal since the character is independent of conjugate class.

Example: the representations of the cyclic group

To get an idea of what happens in the commutative case, here we consider a generic cyclic group $\mathbb{Z}_n$. We can regard this as the group generated by a single element *a*, of order *n*.
Since it is commutative, then all irreducible representations are one-dimensional. A unitary 1×1 matrix is simply a complex number of magnitude 1. Say a maps to the complex number *z*. Since $a^n=e$, it follows that $z^n=1$, i.e., that $z=\exp(\frac{2\pi i}{n}m)$ for some *m*. Each choice of m in $\{0,1,\ldots,n-1\}$ yields a different group representation, as it yields a distinct *z*. Since there are *n* such choices, we have found all the irreducible representations.

Summing up: the m th representation L_m is: $L_m(a)=\exp\left(\frac{2\pi i}{n}m\right)$ (and $L_m(a^j)=\exp\left(\frac{2\pi i}{n}mj\right)$), and its character is $\chi_{L_m}(a^k)=\exp\left(\frac{2\pi i}{n}mk\right)$. Writing a function on the group elements as a sum of the characters is the discrete Fourier transform.

The row orthonormality guaranteed by the group representation theory is that $d(L_m,L_p)=0$ for $m\neq p$ and $d(L_m,L_m)=1$, where

$d(L_m,L_p)=\frac{1}{n}\sum_{j=0}^{n-1}\overline{\chi_m(a^j)}\chi_p(a^j)$. This can be seen directly:
$d(L_m,L_p)=\frac{1}{n}\sum_{j=0}^{n-1}\exp\left(-\frac{2\pi i}{n}mj\right)\exp\left(\frac{2\pi i}{n}pj\right)=\frac{1}{n}\sum_{j=0}^{n-1}\exp\left(\frac{2\pi i}{n}(p-m)j\right)$. If $p=m$, all terms on the right hand side are 1. If $p\neq m$, the right side is a symmetric sum over distinct roots of unity.

Note that there is a symmetry between group elements and characters – in $\chi_{L_m}(a^k) = \exp\left(\frac{2\pi i}{n} mk\right)$, m and k play identical roles. (This is not generic!). So here, just as "row orthogonality" corresponds to Fourier decomposition, "column orthogonality" corresponds to Fourier inversion.

Example: the representations of the group of the cube

To get an idea of what happens in the non-commutative case, here we consider the group of the rotations of a standard 3-d cube. Since we can move any of its six faces into a standard position, and then rotate in any of 4 steps, this group has 24 elements. (Abstractly, this is also the same as the group of permutations of 4 elements – but we won't use that fact explicitly – you can see this by thinking about how rotations of the cube act on its four diagonals.).

Here we work out its "character table" – i.e., a table of the characters of all of its representations. It illustrates many of the properties of characters and representations.

The first step is determining the conjugate classes – these are the sets containing group elements that are identical up to inner automorphism. I.e., if two group elements are the same except for a relabeling due to rotation of the cube, they are in the same conjugate class.

We label three cube faces A, B, C, and their opposites A', B', and C'.

(1) The identity – always one element in this class.
(2) 90-deg rotations around a face – 6 faces, so 6 elements. This is the same as considering 90-deg clockwise or counterclockwise rotations around the axes AA', BB', or CC'.
(3) 180-deg rotations around an axis – 3 elements
(4) 120-deg rotations around a vertex – 8 elements
(5) 180-deg rotations around the midpoint of opposite edges – e.g., exchange A with B, A' with B', and C with C'. – 6 elements (there are 12 edges, so 6 pairs of edges to do this with)

As a check, we now have all 24 elements (1+6+3+8+6), in 5 conjugate classes.

Building the character table

We begin to write the "character table" by setting up a header row with the conjugate classes, and subsequent rows to contain the characters. The numbers in square brackets indicate the number of elements in the conjugate class. The first row is the trivial representation; it is one-dimensional and, since it maps each group element into 1, its character is 1.

	id[1]	*face*90[6]	*face*180[3]	*vertex*120[8]	*edge*180[6]
E	1	1	1	1	1

.

To find some other representations:

Every group element permutes the face-pairs – AA', BB', or CC'. They can thus be represented as permutation matrices on the three items, AA', BB', or CC'. Let's call this representation F (for faces). $\chi_F(g)$, which is the number of elements on the diagonal of F_g, is the number of face-pairs that are unchanged by the group element. For the identity, all are unchanged, so the character is 3. For a 90-deg rotation, one face-pair is unchanged and the other two are swapped, so the character is 1. For a 180-degree rotation, they're all preserved, so the character is 3. For the 120-deg rotation, they are cycled, so the character is 0 (none are preserved). For the edge-flip (e.g., around an edge between an *A*-face and a *B*-face), two pairs are interchanged, and the third pair is preserved, so the character is 1.

	id[1]	*face*90[6]	*face*180[3]	*vertex*120[8]	*edge*180[6]
F	3	1	3	0	1

.

We now need to check whether *F* is irreducible. According to the group representation theorem, it is irreducible if $d(F,F)=1$. So we calculate (using eq. (10)), and using the numbers in the square brackets to keep track of the number of elements in each conjugate class:

$d(F,F)=\frac{1}{24}\left(1\cdot 3^2+6\cdot 1^2+3\cdot 3^2+8\cdot 0^2+6\cdot 1^2\right)=\frac{48}{24}=2$, i.e., *F* is not irreducible.

F likely contains a copy of *E,* since all of its characters are non-negative, and so the dot-proeuct of the characters with that of *E* cannot be zero. If so, we can remove that copy (by $I-P_F$, where P_F is given by eq. (8)), and find a new irreducible representation. To see if F contains a copy of *E*, we calculate (using eq. (10)):

$$d(E,F)=\frac{1}{24}\left(1\cdot(1\cdot 3)+6\cdot(1\cdot 1)+3\cdot(1\cdot 3)+8\cdot(1\cdot 0)+6\cdot(1\cdot 1)\right)=\frac{24}{24}=1.$$

This means that *F* contains one copy of *E*. (Basically confirmatory: since the character of *F* was non-negative, then $d(E,F)$ had to be > 0; if $d(E,F)$ were 2, then *F* would just be two copies of the identity.) To find the other part of *F*, we could work out $I - P_F$ (using eq. (8)), to project onto a subspace that contains no copies of *E* – and hence, which contains some other representation, say F_0, with $F = E \oplus F_0$. But it is easier just to compute the character of F_0: $\chi_F = \chi_E + \chi_{F_0}$, so $\chi_{F_0} = \chi_F - \chi_E$. Entering this into the table:

	$id[1]$	$face90[6]$	$face180[3]$	$vertex120[8]$	$edge180[6]$
E	1	1	1	1	1
F_0	2	0	2	-1	0

.

Another representation is simply regarding these group elements as 3-d rotations, and writing them as 3×3 matrices. Let's call this *M*. To determine the character, we only need to write the matrix out for one example of each conjugate class, since the character is constant on conjugate classes. The identity group element of course yields the identity 3×3 matrix, and a character of 3. A 90-deg face rotation that rotates in the *XY*-plane has a matrix $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and a character of 1. A 180-deg rotation, which is the square of this matrix, has matrix $\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and a character of -1. A 120-deg rotation around a vertex permutes the axes, and so has matrix $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, and character 0. An edge-flip could exchange could *X* for *Y*, and *Y* for *X*, and invert *Z*, and thus, have matrix $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, and character -1. Using eq. (11), we find that *M* is irreducible:

$d(M,M) = \frac{1}{24}\left(1\cdot 3^2 + 6\cdot 1^2 + 3\cdot(-1)^2 + 8\cdot 0^2 + 6\cdot(-1)^2\right) = \frac{24}{24} = 1$. Adding this to the table:

	$id[1]$	$face90[6]$	$face180[3]$	$vertex120[8]$	$edge180[6]$
E	1	1	1	1	1
F_0	2	0	2	-1	0
M	3	1	-1	0	-1

.

As noted above, every group element acts on the three sets of axes, and permutes them. Some group elements cause an odd permutation of the axes (i.e., swap one pair), while others lead to an even permutation (i.e., don't swap any axes, or, cycle through all three of them). So there is a group representation Q that maps each group element to -1 or 1, depending on whether the permutation of the axes is odd or even. Since this is a one-dimensional representation, it must be irreducible. Adding it to the table:

	$\underline{id[1]}$	$\underline{face90[6]}$	$\underline{face180[3]}$	$\underline{vertex120[8]}$	$\underline{edge180[6]}$
E	1	1	1	1	1
F_0	2	0	2	-1	0
M	3	1	-1	0	-1
Q	1	-1	1	1	-1

.

Now let's use tensor products to create a representation. $M \otimes Q$ is a good choice: Q is all 1's and -1's, so if M is irreducible, then so will $M \otimes Q$. (See eq. (11): $\chi_{M\otimes Q} = \chi_M \chi_Q$, so $\left|\chi_{M\otimes Q}\right|^2 = \left|\chi_M\right|^2$, so $d(M \otimes Q, M \otimes Q) = d(M, M) = 1$.) Adding this to the table:

	$\underline{id[1]}$	$\underline{face90[6]}$	$\underline{face180[3]}$	$\underline{vertex120[8]}$	$\underline{edge180[6]}$
E	1	1	1	1	1
F_0	2	0	2	-1	0
M	3	1	-1	0	-1
Q	1	-1	1	1	-1
$M \otimes Q$	3	-1	-1	0	1

.

The table is now finished. We can verify that we've fully decomposed the regular representation – it should have each irreducible representation, repeated a number of times equal to the dimension of the representation, and indeed, $24 = 1^2 + 2^2 + 3^2 + 1^2 + 3^2$. One can also verify that the rows are orthogonal (and columns too!).

If you try to make new representations by tensoring these, you don't get anything new. For example (verify using the characters), $F_0 \otimes F_0 = E \oplus Q \oplus F_0$.

Note that while this character table only contained integers, this is not in general the case -- $\mathbb{Z}_n$, for example, contains non-integer entries.

Application: diffusion on a cube

We'll use the group representation table to analyze a kind of diffusion on the cube, as an example of how this machinery can use symmetry to facilitate the analysis of dynamics.

For definiteness, say the coordinates of the cube are $(\pm 1, \pm 1, \pm 1)$, with the x-axis to the right, the y-axis up, and the z-axis as towards the viewer. Say a particle is placed at the $(+1,+1,+1)$ corner, and, at each time step Δt, the cube is rotated by a quarter turn, but around a random axis. We'd like to know the probability distribution of the position of the particle at later times.

Here, we are thinking of a rotation process in which the probability of a quarter-turn rotation around the x-axis, rotating positive *y* to positive *z*, is $p_x \Delta t$, and in the opposite direction, $q_x \Delta t$. Similarly, rotations around the *y*-axis, rotating positive *z* to positive *x*, have probability $p_y \Delta t$, and in the opposite direction, probability $q_y \Delta t$; and rotations around the *z*-axis, rotating positive *x* to positive *y*, have probability $p_z \Delta t$, and probability $q_z \Delta t$ in the opposite direction. These motions are considered to be fixed in space, while the particle moves as if along a wire-frame cube (somewhat like "Rubik's cube" motions, but here, the front and back faces of the cube cannot move separately). After sufficient time, the particle will have no memory of its starting point. That is, the probability distribution will relax to a constant. Our goal is to determine the time constant(s) of that relaxation. In typical situations, $p_u = q_u$ or even all probabilities are equal; here we keep them unequal so that the structure of the analysis is clearer.

This is a standard way of approaching a diffusion problem, considering first time to be discrete, and then allowing the discretization interval to shrink. At any given time *t*, we have a probability distribution $\vec{s}_t$ (here, on the eight vertices), and we'd like to know how it evolves in time. We can think of $\vec{s}_t$ as a column vector of probabilities, and, according to the dynamics described above, $\vec{s}_{t+\Delta t} = A_{\Delta t}\vec{s}_t$, where $A_{\Delta t}$ is a matrix of transition probabilities. The evolution over a macroscopic period of time T are described by $T/\Delta t$ repeated applications of $A_{\Delta t}$, and thus yields a formal solution $\vec{s}_T = B_T \vec{s}_0$, where $B_T = \lim_{\Delta t \to 0}\left(A_{\Delta t}\right)^{T/\Delta t}$. This limit exists, because $A_{\Delta t} = I + D\Delta t$, so $B_T = \lim_{\Delta t \to 0}\left(I + D\Delta t\right)^{T/\Delta t} = e^{DT}$. To see what these formal solutions mean, we'd like to find a basis in which A is diagonal. Exponentials of *D* will then be easy to compute: in that basis, D will be diagonal as well, and will have the same eigenvectors as A, since $D = \dfrac{A_{\Delta t} - I}{\Delta t}$. So, to compute exponentials, $D = RMR^{-1}$,

$$e^{DT} = I + DT + \frac{D^2T^2}{2!} + \ldots = I + RMR^{-1} + R\frac{M^2T^2}{2!}R^{-1} + \ldots = R\,e^{MT}R^{-1}.$$

This analysis also shows that the eigenvalues of B -- the rates at which the distribution relaxes to a constant -- are the exponentials of the eigenvalues of D. For if $Dv = \lambda v$, then

$$\lim_{\Delta t \to 0}\left(A_{\Delta t}\right)^{T/\Delta t} v = \lim_{\Delta t \to 0}\left(I + D\Delta t\right)^{T/\Delta t} v = \lim_{\Delta t \to 0}\left(1 + \lambda\Delta t\right)^{T/\Delta t} v = e^{\lambda T} v$$

A, here, is 8×8. The group of the rotations of a cube acts on the probability distributions. The key observation is that the diffusion dynamics commutes with this action: the way that one of the elementary rotations (i.e., the moves that occur with probability p_x, q_x, p_y, q_y, p_z, q_z) act on a distribution is the same linear transformation as the way that they act on a distribution that has been acted on by the group. So the 8×8 space in which the group acts can be broken down into smaller subspaces, according to how this action decomposes into irreducible representations.

The group acts as a permutation on the 8 vertices of the cube. To decompose this representation, say, *L*: The identity preserves all of the 8 vertices; face rotation by 90 or 180 deg preserves none of them; rotation around a vertex preserves the vertex and its opposite, and rotation around an edge preserves none of them. It's character is added to the table below:

	$\underline{id[1]}$	$\underline{face90[6]}$	$\underline{face180[3]}$	$\underline{vertex120[8]}$	$\underline{edge180[6]}$
E	1	1	1	1	1
F_0	2	0	2	-1	0
M	3	1	-1	0	-1
Q	1	-1	1	1	-1
$M\otimes Q$	3	-1	-1	0	1
L	8	0	0	2	0

So $L = E\oplus M\oplus Q\oplus(M\otimes Q)$, and we have now found subspaces that break the 8-dimensional space in which *A* and *D* operate into two one-dimensional spaces and two three-dimensional spaces.

To find these subspaces, one can use the formula (12) to project the one-hots (or any other basis for the probability distributions) into each subspace. We order the vertices "lexicographically": $\begin{bmatrix} +1 & +1 & +1 \\ -1 & +1 & +1 \\ +1 & -1 & +1 \\ -1 & -1 & +1 \\ +1 & +1 & -1 \\ -1 & +1 & -1 \\ +1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix}$, and denote a vector in the space of probability distributions by the 8-vector of probabilities assigned to each of these vertices.

For the trivial representation E, projecting any one-hot (e..g, $v_{+++} = (1,0,0,0,0,0,0,0)^T$ yields $\frac{1}{8}(1,1,1,1,1,1,1,1)^T$. In this subspace, $A_{\Delta t}$ acts trivially: if the particle is equally likely to be at all positions prior to a rotation, it is equally likely to be at all positions after a rotation. So $P_E v_{+++} = \frac{1}{8}\left(v_{+++} + v_{-++} + v_{+-+} + v_{--+} + v_{++-} + v_{-+-} + v_{+--} + v_{---}\right)$. Any one-hot projects to the same vector. Once we are in this space, rotation of the cube does not change the distribution (it is still uniform). So $A_{\Delta t}P_E = I$, $DP_E = 0$, and the one eigenvector of D in this space has an eigenvalue of 0.

For the parity representation Q, note that any group element g for which $Q_g = -1$ maps v_{+++} into a one-hot with an odd number of -1's, and any group element for which $Q_g = +1$ maps v_{+++} into a one-hot with an even number of -1's. So, by the projection formula (12), the projection of v_{+++} into the space in which $A_{\Delta t}$ acts like Q is given by $P_Q v_{+++} = \frac{1}{8}\left(v_{+++} - v_{-++} - v_{+-+} + v_{--+} - v_{++-} + v_{-+-} + v_{+--} - v_{---}\right)$. Any other one-hot is either projected to this vector, or its negative. What is $A_{\Delta t}P_Q$? I.e., how does $A_{\Delta t}$ act on $P_Q v_{+++}$? Any rotation of the cube takes a one-hot with an even number of positive coordinates to a one-hot with an odd number of positive coordinates, i.e., it maps $P_Q v_{+++}$ to $-P_Q v_{+++}$. This happens with probability $p_{tot}\Delta t$, where $p_{tot} = p_x + q_x + p_y + q_y + p_z + q_z$. If no rotation occurs, then $P_Q v_{+++}$ is mapped to itself. So $A_{\Delta t}P_Q v_{+++} = (1 - p_{tot}\Delta t)P_Q v_{+++} - (p_{tot}\Delta t)P_Q v_{+++} = (1 - 2p_{tot}\Delta t)P_Q v_{+++}$, and $DP_Q = -2p_{tot}$, the rate at which $P_Q v_{+++}$ relaxes to the uniform distribution.

For the representations M and $M \otimes Q$, one could also use the formula (12). For M, this would identify four vectors that are the projections of the one-hots:

$P_M v_{+++} = \frac{1}{8}(+3,+1,+1,-1,+1,-1,-1,-3)^T$,

$P_M v_{-++} = \frac{1}{8}(+1,+3,-1,+1,-1,+1,-3,-1)^T$,

$P_M v_{+-+} = \frac{1}{8}(+1,-1,+3,+1,-1,-3,+1,-1)^T$,

$P_M v_{--+} = \frac{1}{8}(-1,+1,+1,+3,-3,-1,-1,+1)^T$.

The other one-hots project to the negatives of these. We now have four vectors that span the range of P_M, but we were expecting a 3-dimensional space. The resolution is that there is a linear dependence: $P_M(v_{+++} - v_{-++} - v_{-+-} + v_{--+}) = 0$. A similar phenomenon happens with $M \otimes Q$:

$P_{M\otimes Q} v_{+++} = \frac{1}{8}(+3,-1,-1,-1,-1,-1,-1,+3)^T$,

$P_{M\otimes Q} v_{-++} = \frac{1}{8}(-1,+3,-1,-1,-1,-1,+3,-1)^T$,

$P_{M\otimes Q} v_{+-+} = \frac{1}{8}(-1,-1,+3,-1,-1,+3,-1,-1)^T$,

$P_{M\otimes Q} v_{--+} = \frac{1}{8}(-1,-1,-1,+3,+3,-1,-1,-1)^T$, with the linear dependence

$P_{M\otimes Q}(v_{+++} + v_{-++} + v_{-+-} + v_{--+}) = 0$.

So maybe there is a nicer basis for the ranges of P_M and $P_{M\otimes Q}$.

Consider the vector that assigns ± 1 to the eight entries, depending on the *x*-coordinate: $v_x = (+1,-1,+1,-1,+1,-1,+1,-1)^T$, and similarly v_y and v_z. Each action of the group maps these vectors among themselves: since these vectors are linear functions of the coordinates, and the group actions never transform an axis into an oblique direction. One can now verify that the action of the group on $\{v_x, v_y, v_z\}$ has the same character as M.
Similarly, consider the vector that assigns ± 1 to the eight entries, depending on the product of the *y*- and *z*- coordinates: $v_{yz} = (+1,-1,-1,+1,-1,+1,+1,-1)^T$, and similarly v_{zx} and v_{xy}. Each action of the group also maps these vectors among themselves: since these vectors are the homogeneous quadratic functions of the coordinates, and again, the group actions never transform an axis into an oblique direction. The action of the group on $\{v_{yz}, v_{zx}, v_{xy}\}$ has the same character as $M\otimes Q$.

Note that we could have dealt with Q the same way: the projection we found via the formula (12) corresponds to a projection onto v_{xyz}, a vector whose entries are the products of the three coordinates. In fact, we've now decomposed the space of all functions of the 8 coordinates into a general cubic: $a + b_x v_x + b_y v_y + b_z v_z + c_{yz} v_{yz} + c_{zx} v_{zx} + c_{xy} v_{xy} + d_{xyz} v_{xyz}$. Any function of 8 coordinates can be expressed in this fashion, and we've backed into the observation that, under the group action, the coefficients of each order are transformed in a way that the zeroth-order (*a*), the first-order (*b*), the second-order (*c*), and the third-order (*d*) coefficients do not mix. This will generalize to analyzing dynamics on a sphere.

It remains to calculate the action of $A_{\Delta t}$ in P_M and $P_{M\otimes Q}$, using the above bases.

First, P_M, and the basis $\{v_x, v_y, v_z\}$. Tracking v_x over a timestep will show the general picture. With probability $1 - p_{tot}\Delta t$, there is no rotation, and v_x is of course unchanged. But v_x is also unchanged after a rotation in either direction around the *x*-axis, which occur with probability $(p_x + q_x)\Delta t$. For the other rotations, v_x is transformed into another basis vector. With probability $p_z\Delta t$, v_x is rotated into v_y, and with probability $q_z\Delta t$, it is rotated into $-v_y$. And with probability $p_y\Delta t$, v_x is rotated into $-v_z$ (since it rotates the positive *z*-axis into the positive *x*-axis), and with probability $q_y\Delta t$, it is rotated into v_y. Cyclically permuting (xyz) yields the other columns of the matrix.
We can therefore write out the transition matrix in the range of P_M:

$$A_{\Delta t}P_M = \begin{pmatrix} 1+(-p_{tot}+p_x+q_x)\Delta t & (-p_z+q_z)\Delta t & (p_y-q_y)\Delta t \\ (p_z-q_z)\Delta t & 1+(-p_{tot}+p_y+q_y)\Delta t & (-p_x+q_x)\Delta t \\ (-p_y+q_y)\Delta t & (p_x-q_x)\Delta t & 1+(-p_{tot}+p_z+q_z)\Delta t \end{pmatrix}.$$

So

$$DP_M = \begin{pmatrix} -p_{tot} + p_x + q_x & -p_z + q_z & p_y - q_y \\ p_z - q_z & -p_{tot} + p_y + q_y & -p_x + q_x \\ -p_y + q_y & p_x - q_x & -p_{tot} + p_z + q_z \end{pmatrix}.$$ If $p_u = q_u$, DP_M is immediately diagonalized, with diagonal elements $-2(p_y + p_z)$, $-2(p_z + p_x)$, and $-2(p_x + p_y)$. If all of the p's re equal, each of these rate constants is $-\frac{2}{3} p_{tot}$. In the worst case, with all of the p's and q's unequal, the eigenvalues are the roots of a cubic.

For $M \otimes Q$ and the basis $\{v_{yz}, v_{zx}, v_{xy}\}$: With probability $1 - p_{tot}\Delta t$, there is no rotation, and v_{yz} is unchanged. v_{yz} is inverted after a rotation in either direction around the *x*-axis in either direction. It's transformed to $-v_{zx}$ by a positive rotation around the z-axis (probability $p_z\Delta t$), since this takes the positive *y*-axis into the negative *x*-axis, and it's transformed to v_{zx} by a negative rotation around the z-axis (probability $q_z\Delta t$), which rotates the positive *y*-axis into the positive *x*-axis. Etc.

$$A_{\Delta t} P_{M \otimes Q} = \begin{pmatrix} 1 + (-p_{tot} - p_x - q_x)\Delta t & (+p_z - q_z)\Delta t & (-p_y + q_y)\Delta t \\ (-p_z + q_z)\Delta t & 1 + (-p_{tot} - p_y - q_y)\Delta t & (+p_x - q_x)\Delta t \\ (+p_y - q_y)\Delta t & (-p_x + q_x)\Delta t & 1 + (-p_{tot} - p_z - q_z)\Delta t \end{pmatrix},$$

$$DP_{M \otimes Q} = \begin{pmatrix} -p_{tot} - p_x - q_x & +p_z - q_z & -p_y + q_y \\ -p_z + q_z & -p_{tot} - p_y - q_y & +p_x - q_x \\ +p_y - q_y & -p_x + q_x & -p_{tot} - p_z - q_z \end{pmatrix}.$$

If $p_u = q_u$, $DP_{M \otimes Q}$ is immediately diagonalized, with diagonal elements $-4p_x - 2p_y - 2p_z$, $-2p_x - 4p_y - 2p_z$, and $-2p_x - 2p_y - 4p_z$. If all of the p's re equal, each of these rate constants is $-\frac{4}{3} p_{tot}$.

So in sum, the decomposition of the 8-dimensional representation into its irreducible components allows explicit computation of the rate constants that describe the evolution of the probability distribution in time. In the typical case of $p_u = q_u$, these group into a slowest triplet, corresponding to probability distributions that are linear combinations of $\{v_x, v_y, v_z\}$ around the mean (the component in which rotations act as M), a faster triplet, corresponding to probability distributions that are linear combinations of $\{v_{yz}, v_{zx}, v_{xy}\}$ around the mean (the component in which rotations act as $M \otimes Q$), and a fastest mode, probability distributions that are the mean plus some multiple of v_{xyz}. The evolution of an initial probability distribution can be calculated by projecting it into each of these subspaces; in each subspace, the relaxation to a uniform distribution is governed by one or three exponentials.

Representations of a continuous but non-commutative group

Here we sketch how this machinery extends to a scenario in which the group is non-commutative, and, rather than having discrete elements, is continuous. That is, the group structure is supplemented by a topology – a notion of nearness – with the condition that the topology is preserved by the group action. Intuitively, if two elements of the group (say g_1 and g_2) are "close", then, for any group element h, hg_1 and hg_2 are close, as are g_1h and g_2h. More formally, open sets are mapped to open sets under the group operation. Such groups are known as Lie groups.

The commutative and continuous Lie groups and their representations should be familiar: for the group of rotations of a circle, it is Fourier series. For the group of translations on a line, it is the Fourier transform. Since these groups are commutative, all of the irreducible representations are one-dimensional.

The archetypal example of a non-commutative Lie group is the group of rotations of a sphere in ordinary 3-dimensonal space, denoted $SO(3)$. This group is the substrate for analyzing rotational diffusion in which rotational movements are continuous, rather than restricted to the motions of a cube. Further details and a slightly different approach can be found at a previous homework (2021 Homework 2, Q2) and its extension (2021 Homework 2, Q2 extension). Many of the results that we were able to obtain by counting arguments still hold, but the machinery to prove them is necessarily different – and we omit that here.

There two related differences compared to the discrete-group case.

Conjugate classes and number of irreducible representations

Conjugate classes are no longer discrete, but rather, are continuously parameterized, and the number of irreducible representations is infinite. In the case of $SO(3)$, the conjugate classes are parametrized by a single scalar, say θ, with a range of $[0,\pi]$: every group element is a rotation, by some angle θ, about some axis, and the angle and axis determine the rotation uniquely. Rotations by the same angle, but about different axes (say, g_1 about axis $\vec{a}_1$, and g_2 about axis $\vec{a}_2$) are in the same conjugate class: to see this, find a rotation h that maps $\vec{a}_2$ into $\vec{a}_1$. Then $g_2 = h^{-1}g_1h$. Conversely, rotations by different angles are not conjugate. So the character of a representation, which is a function of conjugate classes, is a function of this parameter θ, rather than a function of a discrete conjugate class label. And to compute the character, we can choose the most convenient member of the conjugate class – typically, a rotation around the "*z*" axis.

For example, consider the "obvious" representation L_1 of $SO(3)$ by 3×3 matrices. A rotation by θ is $R_\theta=\begin{pmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}$. This has trace $1+2\cos\theta$, so we can write $\chi_{L_1}(R_\theta)=1+2\cos\theta$. But note also that we could have changed bases to $x+iy$, $x-iy$, and z, and in the new basis, $R_\theta=\begin{pmatrix}e^{i\theta} & 0 & 0\\ 0 & e^{-i\theta} & 0\\ 0 & 0 & 1\end{pmatrix}$, which of course has the same trace, naturally written as $\chi_{L_1}(R_\theta)=\sum_{k=-1}^{1}e^{ik\theta}$

The trivial representation, which maps every rotation to 1, will be denoted L_0, and $\chi_{L_0}(R_\theta)=1$.

Just as R_θ acts on the homogeneous linear forms in *x*, *y*, and *z* (by transforming them via rotation) can also be considered to act on all the homogeneous quadratic forms x^2, y^2, z^2, xy, xz, and yz (just like $M\otimes Q$ in the case of the cube). This will lead to a 6-dimensional representation, which is identical to $sym(L_1\otimes L_1)$ -- since the quadratics can also be viewed as tensor products. Moreover, $sym(L_1\otimes L_1)$ will not be irreducible, since any rotation preserves $x^2+y^2+z^2$, a one-dimensional subspace. It's not hard to show that after removal of that subspace, what's left is irreducible, and we'll call it L_2. We can compute its character in several ways. First, using the general formula for the trace of a symmetric tensor product: $tr(sym(A\otimes A))=\frac{1}{2}\left((tr(A))^2+tr(A^2)\right)$, the character of $sym(L_1\otimes L_1)$ is given by $\frac{1}{2}\left((\chi_{L_1}(R_\theta))^2+\chi_{L_1}(R_\theta{}^2)\right)$. Since $R_\theta^2=R_{2\theta}$, this works out to $2+2\cos\theta+2\cos2\theta$. But since $sym(L_1\otimes L_1)=L_2\oplus L_0$, we have $\chi_{L_2}(R_\theta)=1+2\cos\theta+2\cos2\theta$. But we could also have used a different basis set: x^2-y^2, ixy, $(x+iy)z$, $(x-iy)z$, and z^2, as this yields $R_\theta=\begin{pmatrix}e^{2i\theta} & 0 & 0 & 0 & 0\\ 0 & e^{-2i\theta} & 0 & 0 & 0\\ 0 & 0 & e^{i\theta} & 0 & 0\\ 0 & 0 & 0 & e^{-i\theta} & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}$. This suggests writing $\chi_{L_2}(R_\theta)=\sum_{k=-2}^{2}e^{ik\theta}$, which is the same as $\chi_{L_2}(R_\theta)=1+2\cos\theta+2\cos2\theta$. And there's a longhand approach in the links referred to above.

At this point, a pattern looks like it is emerging: we can consider $SO(3)$ to act on homogeneous polynomials of order *m*. This corresponds to a symmetric *m*-fold tensor product of L_1. We don't expect it to be irreducible, because of the invariance of $x^2+y^2+z^2\equiv r^2$. So, for example,

$sym(L_1 \otimes L_1 \otimes L_1)$ will have a subspace of dimension 3, spanned by xr^2, yr^2, and zr^2. Removing this space should lead to another irreducible representation. In the case of cubic forms, there are 10 dimensions in $sym(L_1 \otimes L_1 \otimes L_1)$: 3 pure cubics x^3, y^3, z^3; 6 terms like $x^2 y$, and then xyz. Removing the subspace of dimension 3 should then yield a 7-dimensional representation L_3, in which

$$R_\theta = \begin{pmatrix} e^{3i\theta} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-3i\theta} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{2i\theta} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-2i\theta} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{i\theta} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-i\theta} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \text{ and } \chi_{L_3}(R_\theta) = \sum_{k=-3}^{3} e^{ik\theta}.$$

It turns out that this procedure yields all the irreducible representations of $SO(3)$.

Averaging over the group

The other main difference is that averaging over the group by summing is now carried out by integration. Averaging over the group was critical -- for example, to project onto the subspaces corresponding to irreducible representations, eq. (12) – and the formula still holds, provided that we replace the summation by an appropriate integration.

One can readily parameterize the elements of $SO(3)$ -- for example, by the amount of angular rotation (here, θ), and by the direction of the axis (say, a 2-dimensional variable ψ). But the naïve approach of integrating $d\theta d\psi$ has a flaw: if the amount of angular rotation is small, then the axis of rotation does not matter very much, so, giving all θ's equal weight does not make much sense. Put another way, the "volume" of the group's elements in a given conjugate class depends on θ. So we need a weighting $w(\theta)$ that expresses this.

It turns out that for any continuous group (that is "locally compact", which is guaranteed if it is finite-dimensional), there is a quantity, the "Haar measure", that will provide this weighting. This is not trivial. The Haar measure indicates how much "volume" in the group is encompassed by an infinitesimal change in the variables (here, θ and ψ, used to parameterize the group. Integrating the Haar measure over each conjugate class yields the total volume of the conjugate class, as a fraction of the group volume. For $SO(3)$, it is $w(\theta) = \frac{1}{\pi}(1-\cos\theta)$ Note that $w(\theta)$ increases from 0 at $\theta = 0$ to its maximal value at $\theta = \pi$, corresponding to the intuition that the conjugate classes of larger rotations occupy larger volumes. Also, note that

$$\int_0^\pi w(\theta) d\theta = \frac{1}{\pi}\int_0^\pi (1-\cos\theta) d\theta = 1.$$

At this point, one can verify (see homework) that the characters of the irreducible representations are orthogonal, where orthogonality means integrating over the group with respect to the Haar measure:

$$\int_0^{\pi} \chi_{L_m}(R_\theta)\overline{\chi_{L_n}(R_\theta)}w(\theta)d\theta = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases}.$$

Similarly, the matrix elements of the irreducible representations are orthogonal – and these are the spherical harmonics.