Groups, Fields, and Vector Spaces

Overview

The goal is to understand the foundations of the mathematical methods for analyzing (neurophysiologic) data, and models of neurophysiologic processes. The specific targets are Fourier analysis and principal components analysis. The main challenge is that data are intrinsically multivariate – a time series, or an image, or an image that changes in time. We have lots of choices as to how to represent these quantities mathematically – even if at first, this might not seem to be the case. Some of these options, though at first un-natural, lead to important simplifications. We want to understand why this is the case, so that these simplifications no longer seem accidental, and we can see how to both generalize and specialize these basic mathematical approaches.

We will (temporarily) replace familiar objects with mathematical abstractions: signals, images, and movies will be considered elements in a "vector space." Ordinary numbers will be replaced by "fields." The fundamental structure, however, is the "group", an (abstract) set of elements with a single operation and a few key properties. These properties allow a group to be an abstraction of *a set of transformations of an object that preserves some specified aspects of that object's structure*. Studying groups will also serve us in another way, as the basic relationships between groups – homomorphisms and isomorphisms – carry over to the other kinds of mathematical structures as well.

We will find that groups and vector spaces interact in several natural ways. We will then apply this machinery to the group of time-translations and the vector space of functions of time. This will yield Fourier analysis, and the properties – and utility – of Fourier analysis will be natural consequences of the general machinery. Similar considerations will apply to images (functions of space) and the group of translations in space. Even though the symmetries (i.e., the "structure-preserving" aspects) of group operations are only approximate in the real world, the approach is still extremely useful.

Things become mathematically interesting because of the properties of groups, and because elements can play multiple roles: field operations form groups, and field elements operate on vector space elements in a way that forms a group. These distinctions are not so obvious with the group of time-translations (or space-translations), so it's worthwhile to consider more general groups to see this.

Three kinds of mathematical structures

In order of increasing number of kinds of components:

- Groups: one kind of element, one operation
- Fields: one kind of element, two operations ("addition" and "multiplication")

- Vector spaces: two kinds of elements (vectors and scalars); scalars form a field, and operations that apply to (vector, vector) pairs and to (vector, scalar) pairs

A particularly interesting kind of vector space is the set of mappings from elements of a group to a field.

Structure-preserving transformations and natural coordinates

These are the key to identifying natural "coordinates." Here, "coordinates" is used in a very general way, essentially as "labels".

Structure-preserving transformations can be sought for groups, fields, or vector spaces. Structure-preserving transformations always form a group, in their own right. This is a useful way to understand the generic nature of groups, rather than some of the simpler examples (real numbers under addition), since these "simpler examples" often have properties that are not generic to groups.

We will look at structure-preserving transformations of certain vector spaces, and use them to identify particularly natural basis sets for the vector spaces. We will apply this to vector spaces consisting of mappings from a group to a field. Fourier theory falls out from this.

Looking ahead: for a group $G = \mathbb{Z}_n$ (the integers, with addition as the group operation), we will get the discrete Fourier transform. For $G = \mathbb{R}$ (the real numbers, with addition as the group operation), we will get the Fourier transform. For G ={rotations of a circle}, we will get Fourier series.

Other groups lead to other useful constructs, though we won't pursue them here. For example, with G ={rotations of a sphere}, we get spherical harmonics. With G ={permutations of *n* objects}, translations in Euclidean *n*-space, or translations and rotations in Euclidean *n*-space, we get other useful things.

Groups

Group axioms

A group is a set of elements *a*, *b*, …, along with an operation $\circ$ that is a mapping from a pair of elements to a third element, i.e. $a \circ b = c$ (formally, $\circ : G \times G \to G$,which means that the set is closed under the group operation), for which the following hold:

G1: Associativity: $a \circ (b \circ c) = (a \circ b) \circ c$.
G2: Identity: There is a special element $e \in G$ for which, for every *a* in *G*, $a \circ e = a$ and $e \circ a = a$.

G3: Existence of inverses. For every *a* in *G*, there is a corresponding group element a^{-1} for which $a \circ a^{-1} = e$ and $a^{-1} \circ a = e$.

Other properties that many groups have, but are not required:

The group operation need not be commutative (i.e., satisfy $a \circ b = b \circ a$). A commutative group is also called an Abelian group.

A group may have a finite or an infinite number of elements.

An infinite group may, or may not, have a notion of "nearness" of elements. A group in which elements can be arbitrarily close to each other is called a continuous group; otherwise the group is discrete. Typically, the group operation preserves the notion of nearness, and in this case, a continuous group is called a Lie group. Saying that the notion of nearness is preserved by the group operation means that if *a* is near *b*, then $a \circ c$ must be near $b \circ c$ (and similarly $c \circ a$ must be near $c \circ b$).

A set that satisfies G1 but not G2 or G3 is a "semigroup". You can always make it satisfy G2 by adding an identity element, if it doesn't already have one.

Examples of groups

Some examples of groups in which the group operation is familiar addition or multiplication

- The (positive and negative) integers $\mathbb{Z}$, $\circ$ is ordinary addition
- The rational numbers $\mathbb{Q}$, $\circ$ is ordinary addition
- The real numbers $\mathbb{R}$, $\circ$ is ordinary addition
- The complex numbers $\mathbb{C}$, $\circ$ is ordinary addition
- $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$ with 0 omitted, $\circ$ is ordinary multiplication
- $m \times n$ matrices with entries drawn from $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, $\circ$ is matrix addition
- $m \times m$ invertible matrices with entries drawn from $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, $\circ$ is matrix multiplication

Some examples of groups in which the group operation is the composition of transformations:

- Rotations of a regular *k*-gon
- Rotations of a circle (limiting case of the above, "$k \to \infty$")
- Rotations and reflections of a regular *k*-gon
- Rotations and reflections of a circle
- Translations along a line
- Translations and rotations in Euclidean *n*-space
- Rotations of an *n*-sphere

- Permutations of a set of *n* objects

What kinds of structure do the above groups preserve?

Which of the above are commutative?
Are any of the above abstractly identical?
Which of the have an infinite number of elements? Of those, which have a notion of "nearness"?

It will be very helpful to identify properties of groups that apply both to finite groups and to infinite ones – especially if we are thinking of the group element as representing translation in time or in space.

Some basic group properties

We're doing this not just to provide "practice" with the group axioms, but also because of what they mean.

There is only one identity element. For if *e* and *f* were both identity elements, then
$e\circ f=e$ by G2, since *f* is an identity
$e\circ f=f$ by G2, since *e* is an identity
from which it follows that $e=f$.

An element can have only one inverse. For if $a\circ b=e$, then
$b=e\circ b$ by G2, since *e* is the identity
$b=(a^{-1}\circ a)\circ b$ by G3, since a^{-1} is an inverse of *a*
$b=a^{-1}\circ(a\circ b)$ by G1
$b=a^{-1}\circ e$ since we assumed that $a\circ b=e$
and hence,
$b=a^{-1}$ by G2, since *e* is the identity.

No element can have a "private" left or right identity. In other words, if an element *f* is an identity for some group element *a*, then it is the identity *e* for the whole group. For if $a\circ f=a$ (*f* is a "right identity"), then
$f=e\circ f$ by G2, since *e* is the identity
$f=(a^{-1}\circ a)\circ f$ by G3
$f=a^{-1}\circ(a\circ f)$ by G1
$f=a^{-1}\circ a$ since we assumed that *f* was a private identity for *a*, i.e., $a\circ f=a$,
and hence,
$f=e$, i.e., *f* is the group identity. (A similar argument works if we had assumed $f\circ a=a$, i.e., that *f* is a "left identity"). Another consequence of this (that we will use below) is that if $f\circ f=f$, then $f=e$. This is because $f\circ f=f$ means that *f* is a "private: identity for *f*.

The inverse of the product is the product of the inverses, in reverse order. To show this, we need to show that $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$, i.e., that $(a \circ b) \circ (b^{-1} \circ a^{-1}) = e$. --
$(a \circ b) \circ (b^{-1} \circ a^{-1}) = \left((a \circ b) \circ b^{-1}\right) \circ a^{-1} = \left(a \circ (b \circ b^{-1})\right) \circ a^{-1}$, each step by G1
$\left(a \circ (b \circ b^{-1})\right) \circ a^{-1} = (a \circ e) \circ a^{-1} = a \circ a^{-1} = e$, by G3, G2, and G3.

The group operation is one-to-one. That is, if $a \circ c = b \circ c$, then $a = b$. This, essentially, allows us to "cancel." Equivalently, if $x \circ z = y$, then $x = y \circ z^{-1}$
To show this: if $a \circ c = b \circ c$, then
$(a \circ c) \circ c^{-1} = (b \circ c) \circ c^{-1}$, then
$a \circ (c \circ c^{-1}) = b \circ (c \circ c^{-1})$ by G1,
$a \circ e = b \circ e$ by G3
$a = b$ by G2

Major consequence: every finite group G can be considered to be a permutation group. To do this, we need to assign, to each group element x, a permutation $\pi(x)$ on the group elements, and we need to show that this assignment is consistent with the group operation. The natural choice is to say that $\pi(x)$ maps the group element a to the group element $x \circ a$, i.e., $\pi(x)(a) = x \circ_G a$. Knowing that the group operation is one-to-one guarantees that this is a permutation. Equally important, it preserves the group structure. That is, permutations have a group operation, where $q \circ p$ means "first apply p then apply q." So, what we mean by saying that the mapping $x \rightarrow \pi(x)$ preserves structure is that $\pi(y) \circ_P \pi(x) = \pi(y \circ_G x)$ [*], where on the left, the operation is in the permutation group, and on the right, it is in G.

To show that this holds, we have to show that the permutations on both sides of the equation [*] act in the same way on every element $a \in G$. Start on the right, applying the permutation specified by $\pi(y \circ_G x)$ to an arbitrary *a*: $\pi(y \circ_G x)(a) = (y \circ_G x) \circ_G a$, by the definition of our mapping π from G into the permutation group, as applied to $y \circ_G x$. By the associative law in G, $(y \circ_G x) \circ_G a = y \circ_G (x \circ_G a)$. But
$y \circ_G (x \circ_G a) = y \circ_G \left(\pi(x)(a)\right)$, by applying the definition of $\pi(x)$. And
$y \circ_G \left(\pi(x)(a)\right) = \pi(y)\left(\pi(x)(a)\right)$, by applying the definition of $\pi(y)$. Finally,
$\pi(y)\left(\pi(x)(a)\right)$ means "first apply $\pi(x)$ to a, then apply $\pi(y)$ to the result" – which is exactly how we defined the permutation $\pi(y) \circ_P \pi(x)$. Putting this all together, we find that $\left(\pi(y) \circ_P \pi(x)\right)(a) = \pi(y \circ_G x)(a)$ for all a, and this establishes
$\pi(y) \circ_P \pi(x) = \pi(y \circ_G x)$.

Standard permutation notation: Every permutation can be broken down into disjoint cycles, by following around repeated application of the permutation to one element. The permutation that maps A to B, B to C, and C to A is written (ABC) (or, equivalently, (BCA) or (CAB)). The permutation that maps Q to R and R to Q is written (QR) or (RQ). The combination of the two is written, for example, $(ABC)(QR)$.

Typically, groups can also be considered to be permutation groups in other ways. See 2425 homework.

Intrinsic properties of group elements

The "order" of a group element *a* is the least (nonzero) integer *n* for which an *n*-fold product $a \circ a \circ \ldots \circ a$ is the identity, i.e., $a^n = e$. Note that associativity means that we don't have to specify how to put parentheses around $a \circ a \circ \ldots \circ a$; any way of doing it gives the same answer.

For finite groups, every element has a (finite) order. To see this, consider the series $a^0 = e$, a^1, a^2, a^3,…. Since the group is finite, eventually it must repeat. So say $a^m = a^n$. Then (assuming $m < n$),
$a^m = a^n$ implies
$$e = (a^m)^{-1} \circ a^m = (a^m)^{-1} \circ a^n = (a^m)^{-1} \circ (a^m \circ a^{n-m}) = \left((a^m)^{-1} \circ (a^m)\right) \circ a^{n-m} = a^{n-m}$$
so the order of *a* is at most $n - m$.

We can do better than this: for a finite group, the order of a group element is a factor of the size of the group. Here, size means number of elements, $\#(G)$.

We show this by showing something more general. First, define a subgroup: a subgroup of a group *G* is a subset of *H* of *G* that is, in its own right, a group. (Similarly: subfield, subspace, etc.) Note that the associativity law is automatic, so what must be shown is that *H* is closed under the group operation, and that it contains the identity (of *G*), and that it contains inverses of all of its elements.

Note also that if *a* is an element of *G,* and *n* is its order, then $H = \left\{e, a, a^2, \ldots a^{n-1}\right\}$ is a subgroup of *G*, and $\#(H) = n$ -- the notation means that the size of H is n. (Check: what is the group operation table for *H*? To compose elements in H, one adds the "exponents," remembering that a^p, a^{p+n}, a^{p+2n}, … are all the same element, since $a^n = e$. Are inverses always in *H*? Yes: the inverse of a^k is a^{n-k}). *H* is also known as the cyclic group generated by a.

So if we can show that the size of every subgroup is a factor of the size of the group, then we will have also shown that the order of every element is a factor of the size of the group – since the order of an element is the size of the cyclic group it generates.

We'll show this (that $\#(H)$ is a factor of $\#(G)$) by a counting argument: we will divide *G* up into pieces, each of which have the same size as *H*. The pieces are called "cosets. " The definition of a coset: for any element *b* in *G*, the coset *Hb* is the set of all of the elements *g* of *G* that can be written in the form $g = h \circ b$, for some element *h* in *H*.

Every element in *G* is in some coset: *g* is in the coset *Hg*, since $g = e \circ g$, and the identity, *e*, is in *H*.

How do we know that each coset has the same size as *H?* We show that right-multiplying by *b* is a 1-1 map between *H* and *Hb*. One element *h* in *H* maps to a single element hb in *Hb*. Two different elements in *H* cannot map to the same element in *Hb* because there is also an inverse map, from *Hb* to *H*. The inverse map is right-multiplying by b^{-1}: $\left(h \circ b\right) \circ b^{-1} = h \circ \left(b \circ b^{-1}\right) = h$. That is, the map consisting of right-multiplying by b^{-1} is the inverse of map consisting of right-multiplying by *b*, precisely because b^{-1} is the group-operation inverse of *b*.

So we now have to show that the cosets are non-overlapping. That is, either two cosets are disjoint, or they are identical. Say *Hb* and *Hc* are two cosets that are not disjoint. Then there is at least one element in common, i.e., for some h' and h'', $h' \circ b = h'' \circ c$. This means that $b = (h')^{-1} \circ h'' \circ c$. Now we can see that every element in *Hb* is contained in *Hc*: A typical element $g = h \circ b$ is also $g = h \circ \left((h')^{-1} \circ h'' \circ c\right) = \left(\left(h \circ (h')^{-1}\right) \circ h''\right) \circ c$ (after several applications of the associative law); the latter shows that *g* is also in *Hc*.

So *G* is a disjoint union of cosets of any subgroup *H*, and all the cosets have the same size. So the size of *H* must be a factor of the size of *G*.

Several notes, in order of increasing importance to us:

Here we used "right cosets". We also could have used "left cosets" *bH*. Note that a left coset *bH is* not necessarily the same as the right coset *Hb*. For non-commutative groups, a left coset and a right coset can overlap but the overlap can be only partial.

We can use facts about the order of group elements as an elementary way to establish some of the possibilities for the structure of groups of a given size. For prime numbers *p*, there is only one group (abstractly) that has size *p*, namely, the group generated by an element of order *p*. We can think of this as the rotations of a *p*-gon. For non-prime sizes, there are other possibilities; see homeworks 2014-2015b, Q2, for a few. This is the beginning of the broad problem of characterizing all possible groups.

We used a counting argument here to show that the size of a subgroup divides the size of the group, and counting arguments won't work for infinite groups. But the notion of "disjoint union" does work; even for infinite groups one can think of cosets as a way of decomposing a larger group *G* into "slices," each of which is based on the template of the smaller group *H*. This basic idea is a model for building larger structures out of smaller ones, Think of *G* as a space, *H* as a special plane in *G* that runs through the origin, and the cosets of *H* as planes that are parallel to *G*. Also, the notion that right-multiplying by a group element yields a 1-1 map between *H* and its cosets still holds.

This "coset decomposition" is the first instance of something that is easy to do with a finite group, and can be thought of as a toy example of a more general procedure that can be carried out for an infinite group. Another example is summing or averaging over the group. But infinite groups can be discrete or continuous, and if continuous, they can have a finite volume or an infinite volume (integers with ordinary addition is infinite but discrete; rotations of the circle is continuous and finite "volume"; reals with addition is continuous and infinite "volume"). Many aspects of the finite case are typically generic, but we need to keep in mind that some math – which we will skip – is necessary to prove this.

Two key definitions

A *normal* subgroup H is a subgroup for which the left cosets and the right cosets are the same. That is, for any group element b, the cosets Hb and bH are the same. That is, for any $h \in H$, there is also an $h' \in H$ for which $hb = bh'$, i.e, $b^{-1}hb = h'$. This can be written as $b^{-1}Hb \subset H$. Since this must hold for any b, we also have $bHb^{-1} \subset H$. These together imply that $b^{-1}Hb = H$, which is an equivalent definition of "normal." For commutative groups, all subgroups are normal.

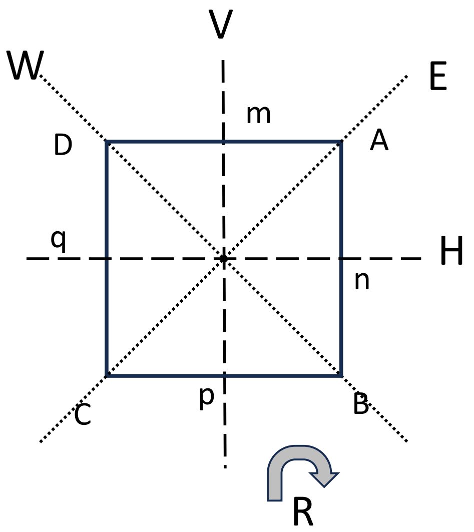
The *cyclic group* of order n, $\mathbb{Z}_n$, is, the group generated by a single element of order n. $\mathbb{Z}_n$ can be thought of in many ways, including:

- the rotations of a regular n-gon,
- cyclic permutations of n symbols,
- the integers $\{0,1,2,...,n-1\}$ under addition mod n -- that is, we compute $a+b$ in the standard way, but only keep track of the remainder after dividing by n.
- the complex numbers $e^{\frac{2\pi i}{n}k}$, for $k \in \{0,1,2,...,n-1\}$, under multiplication.

Abstractly, all of these are the same group, just with different conventions for labeling the elements and naming the operation.

Example: the group of the square

The above points, and some further points below, are illustrated by the group of symmetries of a square. (This is also known as D_4, a "dihedral group", and one can similarly



analyze D_n, the group of symmetries of a regular n-gon.) Other than the trivial transformation, this group contains a quarter-cycle right rotation R, its powers, and mirror reflections along the cardinal axes (H and V), and along the diagonal axes (E and W). The group operation is composition.

The group multiplication table is shown in the top half of the table below. Note that each row of the table contains each group element once. Note that the group operation is not commutative – for example, $R \circ H = W$ but $H \circ R = E$. Note also that the elements have orders 2 or 4, all factors of the size of the group, 8.

There are several subgroups. One is the subgroup of all rotations $\{I, R, R^2, R^3\}$, which is $\mathbb{Z}_4$. This is a normal subgroup: the nontrivial left coset consists of all the reflections, e.g., $\{H \circ I, H \circ R, H \circ R^2, H \circ R^3\} = \{H, E, V, W\}$, as does the nontrivial right coset (but in a different order): $\{I \circ H, R \circ H, R^2 \circ H, R^3 \circ H\} = \{H, W, V, E\}$

Another subgroup is $\{I, H\}$, since $H^2 = I$. This subgroup is not a normal subgroup – for example the left coset by R is $\{R, R \circ H\} = \{R, W\}$, but the right coset by R is $\{R, H \circ R\} = \{R, E\}$.

$x \circ a$		a							
		I	R	R^2	R^3	H	V	E	W
x	I	I	R	R^2	R^3	H	V	E	W
	R	R	R^2	R^3	I	W	E	H	V
	R^2	R^2	R^3	I	R	V	H	W	E
	R^3	R^3	I	R	R^2	E	W	V	H
	H	H	E	V	W	I	R^2	R	R^3
	V	V	W	H	E	R^2	I	R^3	R
	E	E	V	W	H	R^3	R	I	R^2
	W	W	H	E	V	R	R^3	R^2	I
permutations									
vertices		I	(ABCD)	(AC)(BD)	(ADCB)	(AB)(CD)	(AD)(BC)	(BD)	(AC)
edges		I	(mnpq)	(mp)(nq)	(mqpn)	(mp)	(nq)	(mn)(pq)	(mq)(np)
front/back		I	I	I	I	(fr,bk)	(fr,bk)	(fr,bk)	(fr,bk)
cardinal mirrors		I	(HV)	I	(HV)	I	I	(HV)	(HV)
oblique mirrors		I	(EW)	I	(EW)	(EW)	(EW)	I	I

Relationships among groups: homomorphisms

A (group) *homomorphism* is a **structure-preserving map** between two groups. Formally: if *G* and *H* are groups (*H* not necessarily a subgroup of *G*), then $\varphi : G \to H$ is a mapping from *G* to *H* for which

$\varphi(g_1 \circ g_2) = \varphi(g_1) \circ \varphi(g_2)$. Note that on the left side of the equation, $\circ$ is the group operation in *G*; on the right, $\circ$ is the group operation in *H.*

An *onto* homomorphism φ (a.k.a. "surjective" homomorphism) is a homomorphism from *G* and *H* for which all members of *H* are some $\varphi(g)$.

An *isomorphism* is an "onto" homomorphism φ from *G* to *H* if there is also an "onto" homomorphism $\varphi^{-1} : H \to G$, for which $\varphi^{-1}(\varphi(g)) = g$ (and also, $\varphi(\varphi^{-1}(h)) = h$).

An *automorphism* is an isomorphism from a group *G* to itself.

Each of these can also be defined in an analogous fashion for other algebraic structures, such as fields and vector spaces.

Examples of homomorphisms

The log is a homomorphism from $\mathbb{R} > 0$ (with $\circ$ as multiplication) to $\mathbb{R}$ (with $\circ$ as addition).

$\varphi(n) = 2n$ is a homomorphism from $\mathbb{Z}$ (with $\circ$ as addition) to $\mathbb{Z}$ (with $\circ$ as addition). "Remainder mod k" is a homomorphism from $\mathbb{Z}$ (with $\circ$ as addition) to $\mathbb{Z}_k$ (with $\circ$ as addition).

$\varphi(n) = -n$ is a homomorphism from $\mathbb{Z}$ (with $\circ$ as addition) to $\mathbb{Z}$ (with $\circ$ as addition).

$\varphi(z) = e^z$ is a homomorphism from $\mathbb{C}$ (with $\circ$ as addition) to nonzero elements of $\mathbb{C}$ (with $\circ$ as multiplication)

The correspondence between a finite group and its representation as a permutation group.

Which of the above examples are onto? Which are isomorphisms? Which are automorphisms?

A nontrivial homomorphism: parity

The parity of a permutation is a homomorphism from any permutation group (with $\circ$ as composition) to $G = \{+1, -1\}$ (with $\circ$ as multiplication). This will be crucial for constructing the determinant. The "parity" of a permutation is defined as follows. Any permutation can be built from a sequence of pairwise swaps. If the number of pairwise swaps is even, the parity is +1. If the number of pairwise swaps is odd, the parity is -1.

But we need to show that this is well-defined: that no matter how you build up a permutation from pairwise swaps, the parity will be the same. In other words, every permutation can be constructed either from an even number of pairwise swaps, or an odd number, but not both.

To show that the parity of a permutation τ (denoted $parity(\tau)$) is well-defined, we use a classic trick. Define a polynomial $P(X_1, X_2, \ldots, X_h) = (X_2 - X_1)(X_3 - X_1)(X_3 - X_2) \cdot \ldots \cdot (X_h - X_{h-1})$. This has one term $X_c - X_a$
for each pair of indices (a, c) with $a < c$. We will show that $P(X_{\tau(1)}, X_{\tau(2)}, \ldots, X_{\tau(h)}) = parity(\tau) \cdot P(X_1, X_2, \ldots, X_h)$. First, observe that when we apply τ to the subscripts, we simply scramble the order of the terms, and we also may change some terms into their negatives (if $a < c$ but $\tau(c) < \tau(a)$). So $P(X_{\tau(1)}, X_{\tau(2)}, \ldots, X_{\tau(h)}) = \pm P(X_1, X_2, \ldots, X_h)$. To show that the $\pm$ factor is $parity(\tau)$, we observe that if τ is a single pair-swap (say, of *a* and *c*, with $a < c$), then the sign of *P* is inverted. This is because we can catalog the effects of τ on *P*: the sign of $X_c - X_a$ is inverted (leading to a factor of -1), and, for all *b* between *a* and *c*, pairs of terms $(X_c - X_b)(X_b - X_a)$ become $(X_a - X_b)(X_b - X_c)$, which contributes no net sign change.

Computing the parity from standard permutation notation

We can now easily compute the parity of a permutation from its standard notation. The key observation is that a k-cycle can be created with $k-1$ pair-swaps. For example ($k = 3$), (ABC) is equivalent to (AB) followed by (AC), and ($k = 4$) $(ABCD)$ is equivalent to (AB) followed by (AC) followed by (AD). So the parity of a k-cycle is $+1$ if k is odd, and -1 if k is even. (The above analysis shows us that we only need to create a permutation from pair-swaps in one way, and that all other ways must have the same parity.) So to determine the parity of a permutation, we write it in standard notation and count up the number of cycles with an *even* number of elements. If the number of such cycles *odd*, then the permutation has an odd parity. Otherwise, it has an even parity.

While parity is an intrinsic property of a permutation – and groups can always be represented as permutations – parity is not an intrinsic property of a group element. Refer to the group of the square, unless one specifies how the group is represented. Considered as a permutation on the vertices, the cardinal-axis reflections $\{H, V\}$ are even permutations, since there are two pair-swaps, and the diagonal-axis reflections $\{E, W\}$ are odd, since there is only one pair-swap. Considered as a permutation on the edges, the cardinal-axis reflections are odd permutations, and the diagonal-axis reflections are even permutations. Considered as a permutation on the group elements themselves (i.e., the rows of the table), all group elements are even permutations: for example, left multiplication by R is the permutation $(IRR^2R^3)(HWVE)$, and left multiplication by H is the permutation $(IH)(RE)(R^2V)(R^3W)$.

The kernel

The kernel of a homomorphism $\varphi : G \to H$ is the set of elements of *G* for which $\varphi(g) = e_H$. Here, e_H is the identity for *H*. (Unfortunately, there is no obvious relationship to other uses of the term "kernel".)

The kernel of a homomorphism is always a subgroup. It's obviously a subset, so we need to show that G2 and G3 hold.

To show G2 (that there is an identity), we need to show that e is in the kernel. That is, we need to show that $\varphi(e)$ is the identity for *H*. $\varphi(e) = \varphi(e \circ e) = \varphi(e) \circ \varphi(e)$. So $\varphi(e) = e_H$, since it is the "private" identity for $\varphi(e)$.

To show G3 (that if *g* is in the kernel, then so is g^{-1}), we need to show that $\varphi(g) = e$ implies that $\varphi(g^{-1}) = e$. To do this:
$e_H \circ \varphi(g^{-1}) = \varphi(g) \circ \varphi(g^{-1}) = \varphi(g \circ g^{-1}) = \varphi(e) = e_H$. (Second equality uses the fact that a homomorphism is structure-preserving, last equality uses what we just showed, that $\varphi(e) = e_H$.

The argument for G2 also shows that $\left(\varphi(g)\right)^{-1} = \varphi(g^{-1})$. Note that the inverse on the left is found in *H*; the inverse on the right is found in *G*. This follows because $\varphi(g) \circ_H \varphi(g^{-1}) = \varphi(g \circ_G g^{-1}) = \varphi(e_G) = e_H$. Since inverses are unique, $\left(\varphi(g)\right)^{-1} = \varphi(g^{-1})$.

Objects playing several roles: automorphisms

We now show how the set of automorphisms of a group *G* can in turn be considered a group, which we will call $A(G)$. We need to define the group operation in $A(G)$, which must take a pair of automorphisms to a third. We'll use composition. (Here, we will use $\circ$ to denote the group operation in $A(G)$, and juxtaposition (e.g., *gh*) to denote the group operation in *G*.) Formally, to define $\varphi_1 \circ \varphi_2$, we need to define how it acts on an element of G, and to show that with this definition, $\varphi_1 \circ \varphi_2$ is itself an automorphism of G:
$\varphi_1 \circ \varphi_2(g) = \varphi_1\left(\varphi_2(g)\right)$.

To show that this is an automorphism:

$\varphi_1 \circ \varphi_2(gh) = \varphi_1\left(\varphi_2(gh)\right)$ (by the definition of the group operation in $A(G)$)
$= \varphi_1\left(\varphi_2(g)\varphi_2(h)\right)$ (since φ_2 is a homomorphism)
$= \varphi_1\left(\varphi_2(g)\right)\varphi_1\left(\varphi_2(h)\right)$ (since φ_1 is a homomorphism)
$= \left(\varphi_1 \circ \varphi_2(g)\right)\left(\varphi_1 \circ \varphi_2(h)\right)$ (by the definition of the group operation in $A(G)$, applied to each factor)

We next need to show that this operation leads to a group structure on $A(G)$. Associativity follows from the fact that the operation is a composition. The presence of an identity in $A(G)$ follows from the fact that the trivial map from *G* to itself is an automorphism (but not an interesting one). The presence of inverses in $A(G)$ follows from the fact that an automorphism has an inverse (since it is an isomorphism).

A special set of automorphisms: the "inner" automorphisms. For any element α in *G*, let's look at the map $\varphi_\alpha(g) = \alpha g \alpha^{-1}$. It's easy to see that φ_α is an automorphism of *G*: It preserves structure:
$\varphi_\alpha(gh) = \alpha(gh)\alpha^{-1} = \alpha(g\alpha^{-1}\alpha h)\alpha^{-1} = (\alpha g \alpha^{-1})(\alpha h \alpha^{-1}) = \varphi_\alpha(g)\varphi_\alpha(h)$.

To see that the "inner" automorphism group contains identities and inverses (as automorphisms), we need to see how inner automorphisms compose:
$\left(\varphi_\alpha \circ \varphi_\beta\right)(g) = \varphi_\alpha\left(\varphi_\beta(g)\right) = \varphi_\alpha(\beta g \beta^{-1}) = \alpha\beta g \beta^{-1}\alpha^{-1} = \alpha\beta g(\alpha\beta)^{-1} = \varphi_{\alpha\beta}(g)$
so
$\varphi_\alpha \circ \varphi_\beta = \varphi_{\alpha\beta}$ (where the subscript on the right is the group operation in *G*).
As a consequence, $\left(\varphi_\alpha\right)^{-1} = \varphi_{\alpha^{-1}}$, i.e., φ_α is invertible and its inverse is also an inner automorphism.

We can think of the "inner" automorphisms as a model for change of coordinates.

Summing up: For any group *G*, we have a group of automorphisms $A(G)$, and a homomorphism from *G* into a subgroup of $A(G)$, the "inner" automorphisms: This mapping, the adjoint map, $Adj : G \rightarrow A(G)$, takes a group element α into the inner automorphism φ_α. The action of φ_α on G is defined by $\varphi_\alpha(g) = \alpha g \alpha^{-1}$.

What is the kernel of *Adj*? Say γ is in the kernel of *Adj*. This means that φ_γ is the identity transformation on *G*. That is, $\varphi_\gamma(g) = g$ for all *g* in *G*. That is, $\gamma g \gamma^{-1} = g$ for all *g* in *G*.Or, $\gamma g = g\gamma$. In other words, the kernel of *Adj* is the set of elements γ in *G* that commute with all elements in *G*. (This is known as the "center" of *G*).

If *G* is commutative (i.e., everything commutes), the center of *G* is *G* itself, and *Adj* is trivial – in other words, all inner automorphisms are the identity. But there may still be some nontrivial members of $A(G)$.

Examples of automorphisms, inner automorphisms, etc.

$\mathbb{Z}$ (with $\circ$ as addition): It is commutative, so all inner automorphisms are trivial. But $\varphi(n) = -n$ is an automorphism (that is nontrivial, and not an inner automorphism).

Invertible $m \times m$ matrices: For generic matrices *M*, $\varphi_M(G) = MGM^{-1}$ is a nontrivial inner automorphism. The center of the group of invertible $m \times m$ matrices, i.e., the matrices that commute with all others, and therefore lead to the trivial inner automorphisms, are multiples of the identity matrix.

Referring again to the group of the square: there are four inner automorphisms: the trivial inner automorphism φ_I, and automorphisms φ_R, φ_H, and φ_E. (Note, $\varphi_{R^2} = \varphi_I$, $\varphi_{R^3} = \varphi_R$, $\varphi_V = \varphi_H$, and $\varphi_W = \varphi_E$.) There are also outer automorphisms, in which cardinal reflections are exchanged for the diagonal reflections. This corresponds to reversing the roles played by the vertices and the edges. Rotations remain rotations, but mirrors that preserve one pair of edges are exchanged for mirrors that preserve one pair of vertices.

Fields

Field axioms

A field is a set of elements $\alpha, \beta, ...$ along with two operations, + and $\cdot$.

For the operation +, the elements form a commutative group. The identity is denoted by 0. The inverse of α is denoted $-\alpha$.

For the operation $\cdot$ (typically denoted by juxtaposition), the elements other than 0 form a commutative group, and the identity is denoted by 1. The inverse of *a* is denoted $1/\alpha$ or α^{-1}.

The operations + and $\cdot$ are linked by the distributive law, $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

Fields can be finite or infinite.

Field examples

The real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ are the familiar ones, and the ones we typically use to represent scalar quantities

$\mathbb{C}$ has a very non-obvious property that $\mathbb{R}$ does not have: in $\mathbb{C}$, every polynomial equation has roots. ("$\mathbb{C}$ is an algebraically closed field.") Not the case in $\mathbb{R}$: for example, $x^2+1=0$ has no roots. On the other hand, $\mathbb{R}$ is an "ordered field" ("order" here meaning size rank, not the group-theoretic meaning of "order"): if $\alpha \neq \beta$, then either $\alpha < \beta$ or $\beta < \alpha$, but not both.

The integers $\mathbb{Z}$ do not form a field, since there are no inverses.

There are many other fields, including the rational numbers ($\mathbb{Q}$) and finite fields.

Relationships between fields

Familiar example: the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$
We write complex numbers as $z = x + yi$, where $i^2 = -1$ and *x* and *y* are reals.

If all we are told is that the α's and β's are drawn from a field *k*, and that θ is a symbol that can be added and multiplied by the α's and β's, in a manner that follows the distributive law, we would know how to add quantities like $\alpha_0 + \alpha_1\theta$. For example, $(\alpha_0 + \alpha_1\theta) + (\beta_0 + \beta_1\theta) = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1)\theta$. We could also try to multiply them, but we would find:

$$\begin{aligned}(\alpha_0 + \alpha_1\theta)\cdot(\beta_0 + \beta_1\theta) &= \alpha_0 \cdot (\beta_0 + \beta_1\theta) + \alpha_1\theta \cdot (\beta_0 + \beta_1\theta) \\ &= \alpha_0\beta_0 + (\alpha_0\beta_1 + \alpha_1\beta_0)\theta + \alpha_1\beta_1\theta^2\end{aligned} \quad [*]$$

which is a "problem", since there is a θ^2-term. So, to ensure that the result of the multiplication is of the same form ($\gamma_0 + \gamma_1\theta$), we need to have a way to write θ^2 in terms of θ and field elements, i.e., an equation of the form $\theta^2 + c_1\theta + c_0 = 0$. We want to choose c_1 and c_0 so that there is no solution of $x^2 + c_1x + c_0 = 0$ in *k*, since if there *was* a solution, i.e, $x^2 + c_1x + c_0 = 0$ for *x* in *k*, then we'd have two ways of writing the same thing (*x* and θ). So this process of "extension" only works if we choose a polynomial that does not have a root in the starting field. Extending from the real numbers $\mathbb{R}$ to the complex numbers $\mathbb{C}$ chooses the simplest such polynomial: $\theta^2 = -1$. Putting this in the above equation [*]yields the familiar rule for multiplying complex numbers:

$$(\alpha_0 + \alpha_1\theta)\cdot(\beta_0 + \beta_1\theta) = \alpha_0\beta_0 - \alpha_1\beta_1 + (\alpha_0\beta_1 + \alpha_1\beta_0)\theta$$

The extension process guarantees that the extension field has a nontrivial automorphism; in the familiar case of extending from $\mathbb{R}$ to $\mathbb{C}$, this automorphism is complex conjugation. To see this: Equations of the form $\theta^2 + c_1\theta + c_0 = 0$ are expected to have two roots. We don't need to "solve" the quadratic; we merely need to observe that

$x^2+c_1x+c_0=(-c_1-x)^2+c_1(-c_1-x)+c_0$, so if θ solves the quadratic, so does $\theta^*=-c_1-\theta$. This means that bothθ and $\theta^*=-c_1-\theta$ yield the same formula for reducing their squares. So, even though $x+\theta y$ and $x+\theta^* y$ are different elements of the extension field, the multiplication rules are identical. (And clearly, the addition rules are identical too). That is, that there is an automorphism of the extension field, *conj*, which maps $x+\theta y$ into $conj(x+\theta y)=(x+\theta^* y)$.

Note also that $(x+\theta y)(x+\theta^* y)$ is unchanged by *conj*. So, it must be in the base field. Therefore, when it is written in the form $a+b\theta$, it cannot involve θ (i.e., $b=0$). So the mapping from it is a mapping from $x+\theta y$ into $(x+\theta y)(x+\theta^* y)$ is a mapping from the extension field back to the base field that preserves multiplication.

The mapping *conj* lets us see that the extension field has multiplicative inverses: $\frac{1}{(x+\theta y)}=\frac{(x+\theta^* y)}{(x+\theta y)(x+\theta^* y)}=\frac{x}{D}+\frac{y}{D}\theta^*$, for $D=(x+\theta y)(x+\theta^* y)$ which we know is in the base field.

In the familiar case of $\mathbb{R}$ and $\mathbb{C}$, *i* solves $x^2=-1$ (so, $c_1=0$ and $i^*=-i$), so *conj* is the familiar $conj(x+iy)=(x-iy)$, complex conjugation, more typically written z^*. For $z=x+iy$, we write $|z|^2=zz^*=x^2+y^2$.

Choosing other polynomials $x^n+c_{n-1}x^{n-1}+\ldots c_1x^1+c_0=0$ (with *x* in $\mathbb{R}$) does not yield anything new. Attempting to extend $\mathbb{C}$ by this strategy also does not yield anything new. This is because $\mathbb{C}$ is “algebraically closed”—any polynomial with real or complex coefficients has a full set of roots in $\mathbb{C}$. This is a fundamental reason that $\mathbb{C}$ is so useful.

But if we choose polynomials such as $x^n+c_{n-1}x^{n-1}+\ldots c_1x^1+c_0=0$ with *x* in other fields, we would get new and interesting structures. For the field $\mathbb{Z}_p$, we would get the Galois fields of size p^n. For *p*=2, this is the algebraic structure underlying “m-sequences”, a way of producing pseudorandom cyclic binary sequences that have some very nice properties for experimental design. The main property of an m-sequence is that a cyclic shift of the sequence is very nearly orthogonal to the original sequence is a consequence of the field axioms. There is also a great deal of interesting structure choosing the field of rational numbers as the base field and $n=2$ (Pell’s equation and related) or $n=3$ (elliptic function cryptography).

Finite fields

For each size that is a power of a prime number, p^n, there is exactly *one* finite field, known as a Galois field. Finite fields are important for experimental design. Galois fields can be constructed by extending a field of size p by a polynomial in the field of degree

n that cannot be factored. (If it could be factored, the factors would construct two nonzero field elements whose product is zero. This cannot happen, since then neither would have a multiplicative inverse.)

The field of size p is denoted $\mathbb{Z}_p$ or $GF(p,1)$, and consists of the integers $\{0,1,\ldots,p-1\}$. Addition and multiplication are "mod p". That is, carry out addition and multiplication in the ordinary fashion, and then find the remainder after dividing by p. We'll look at $n>1$ later; that construction generalizes the relationship of the complex numbers to the real numbers.

To see that $\mathbb{Z}_p$ (the $n=1$-case) is in fact a field:

The additive group is the cyclic group, generated by 1. But are there multiplicative inverses? I.e., given an α in $\{0,1,\ldots,p-1\}$, are we guaranteed to find a β such that $\alpha\beta=1$ (mod p)?

Two very different ways to see that multiplicative inverses exist.

Method 1: use the fact that automorphisms form a group, and apply this to $\mathbb{Z}_p$. Viewed abstractly, the additive group of $\mathbb{Z}_p$ is a cyclic group. We are guaranteed that the order of every nonzero element of $\mathbb{Z}_p$ (under addition) is p, since the order of every element must be a factor of $\#(\mathbb{Z}_p)=p$, and p is prime. We also know that the map $\varphi_\alpha(x)=\alpha x$ is a homomorphism of the additive group of $\mathbb{Z}_p$. This follows from the distributive law: $\varphi_\alpha(x+y)=\alpha(x+y)=\alpha x+\alpha y=\varphi_\alpha(x)+\varphi_\alpha(y)$. Now choose x to be any element that is not the identity. If $\alpha<p$, then αx cannot be the identity, since the order of x is p. (Note that αx is NOT the group operation for the additive group; it means $x+x+\ldots+x$ a total of α times.) Since αx is not the identity, its order must be p (since the only possible orders are factors of p, and p is prime). Therefore, successive applications of + to αx will produce all of the members of the group $\mathbb{Z}_p$. Therefore $\varphi_\alpha(x)=\alpha x$ is an isomorphism, not just a homomorphism. Since isomorphisms form a group, φ_α must have an inverse. Call it ψ. $\psi(x)$ must be something in $\mathbb{Z}_p$, so let's call $\psi(x)=\beta x$. Since ψ preserves structure, $\psi(2x)=\psi(x+x)=\psi(x)+\psi(x)=\beta x+\beta x=\beta(2x)$, and similarly for 3$x$, … so $\psi(y)=\beta y$ for any y. Finally, since ψ and φ_α are inverses, $x=\psi(\varphi_\alpha(x))=\psi(\alpha x)=\beta\alpha x$, so $x=\beta\alpha x$. That is, x added to itself $\beta\alpha$ times is x, i.e., $\beta\alpha-1$ is a multiple of p. And β is a multiplicative inverse for α (mod p).

Method 2: use the Euclidean algorithm to construct an inverse.

$\alpha\beta=1$ (mod p) is equivalent to $\alpha\beta+pq=1$, for some integer q.

In general: for a given *A* and *P*, $AB+PQ=1$ has a solution in integers when, and only when, *A* and *P* are relatively prime. (This is the classic Euclidean algorithm). If *p* is prime, α and *p* are guaranteed to be relatively prime, and consequently, $\alpha\beta+pq=1$ has a solution. The solution is obtained by "descent", using the Euclidean algorithm: The Euclidean algorithm is easiest to explain by example. Say we want to find the multiplicative inverse of 18 in $\mathbb{Z}_{79}$. That is, we want to find integers β and *q* that solve $\alpha\beta+pq=1$ for $p=79$ and $\alpha=18$. To solve $18\beta+79q=1$ in integers:

Step 1: Note that $79=4\cdot 18+7$. So,
$18\beta+79q=1$ is equivalent to $18\beta+(4\cdot 18+7)q=1$, or, $18(\beta+4q)+7q=1$.
So if $18\beta'+7q'=1$, we can solve $18\beta+79q=1$ with $\beta=\beta'-4q$ and $q=q'$.

Step 2: Note that $18=2\cdot 7+4$. So,
$18\beta'+7q'=1$ is equivalent to $(2\cdot 7+4)\beta'+7q'=1$, or, $4\beta'+7(2\beta'+q')=1$.
So if $4\beta''+7q''=1$, we can solve $18\beta'+7q'=1$ with $\beta'=\beta''$ and $q'=q''-2\beta'$.

Step 3: Note that $7=1\cdot 4+3$. So,
$4\beta''+7q''=1$ is equivalent to $4\beta''+(1\cdot 4+3)q''=1$, or, $4(\beta''+q'')+3q''=1$.
So if $4\beta'''+3q'''=1$, we can solve $4(\beta''+q'')+3q''=1$ with $\beta''=\beta'''-q''$ and $q''=q'''$.

We are guaranteed to have smaller and smaller coefficients, since at each stage we reduce one coefficient to its remainder when divided by the other.
An integer solution of $4\beta'''+3q'''=1$ is "obvious" – if not, we could go one more stage.
$\beta'''=1$, $q'''=-1$; working backwards yields
$\beta''=2$, $q''=-1$; then
$\beta'=2$, $q'=-5$; then
$\beta=22$, $q=-5$. So, $18\cdot 22-79\cdot 5=1$, i.e., $18\cdot 22=1$ (mod 79), i.e., 22 is the inverse of 18 in $\mathbb{Z}_{79}$.

Vector Spaces

Vector space axioms

Vector spaces have two kinds of elements: vectors (*v*, *w*, …) drawn from a set *V* and scalars (α, β, …) drawn from a set *k*.

The scalars form a field *k*, operations are scalar addition, +, and multiplication, $\cdot$.

The vectors form a commutative group under addition, operation is vector addition, +. The additive inverse of v is $-v$.

There is an operation “scalar multiplication” that maps a scalar α and a vector *v* into a vector αv. It satisfies two kinds of distributive laws,

$\alpha(v+w)=\alpha v+\alpha w$ and
$(\alpha+\beta)v=\alpha v+\beta v$. As a consequence of the latter, $0v$ must be the identity for vector addition, since $\alpha v=(\alpha+0)v=\alpha v+0v$ (and the identity is unique). And the additive inverse of αv, $-(\alpha v)$, is the same as $(-\alpha)v$, since inverses are unique, and
$0v=(\alpha+(-\alpha))v=\alpha v+(-\alpha)v$.

There is also a kind of associative law that relates scalar and field multiplication: $(\alpha\beta)v=\alpha(\beta v)$. (The multiplication $\alpha\beta$ is in the field.) As a consequence of this, $1v=v$. (Calculate $1(1v-v)=1(1v)-1v=(1\cdot 1)v-1v=1v-1v=0$.)

Unless stated otherwise, we work with the field $k=\mathbb{C}$.

Nothing is said about length, angles, or dimension.

Vector space examples

A field can always be considered as a vector space over itself, with the vector operations identical to the field operations.

Ordered *n*-tuples of field elements form a vector space. To define the operations: say $v=(\alpha_1,\alpha_2,\ldots,\alpha_n)$ and $w=(\beta_1,\beta_2,\ldots,\beta_n)$. Operations work component-by-component. Vector addition: $v+w=(\alpha_1+\beta_1,\alpha_2+\beta_2,\ldots,\alpha_n+\beta_n)$.
Scalar multiplication: $\lambda v=(\lambda\alpha_1,\lambda\alpha_2,\ldots,\lambda\alpha_n)$.

An extension field is a vector space over the base field. The vector operations are identical to the field operations. One can think of $z=a_0+a_1\theta+\ldots+a_{n-1}\theta^{n-1}$ as an ordered *n*-tuple $(a_0,a_1,...,a_{n-1})$, with the powers of θ merely tagging the places. So, the complex numbers are a vector space (of dimension 2, see below) over the reals. But extension fields have a lot more structure – for example, elements can be multiplied. It is important to remember that vector spaces typically don’t have this behavior.

The set of functions from any set S to the field *k* forms a vector space. This generalizes the ordered *n*-tuple example; the subscripts were merely placeholders, i.e., $S=\{1,2,...,n\}$. To be explicit: Say f and g are functions on S. To define $f+g$, we need to define how it acts on elements of S: $(f+g)(s)=f(s)+g(s)$, where the addition is in k. To define αf, we need to define its action: $(\alpha f)(s) = \alpha \cdot (f(s))$ where the "outer" multiplication on the right is in the field k. This construction is known as the "free vector space" on S.

Particularly important vector space are the set of functions on the real numbers, or on the complex numbers. We often restrict consideration to the functions that are continuous, or have other conditions – such as that their integrals are finite, or that they have derivatives. It is worth checking that these restrictions do result in vector spaces. The key thing to check is that they are closed under vector addition and scalar multiplication.

Linear independence, span, and basis sets

Definition of linear independence: A set of vectors $\{v_1,\ldots,v_h\}$ is linearly independent if $\sum_{k=1}^{h} \alpha_k v_k = 0$ implies that each $\alpha_k = 0$.

Definition of linear span (or just span): The span of a set of vectors $\{v_1,\ldots,v_h\}$ is the set of all vectors *v* that can be written as a linear combination $v = \sum_{k=1}^{r} \alpha_k v_k$ of members of the set. Note that the span of a set of vectors is always a vector space.

Definition of a basis: A set of vectors $\{v_1,\ldots,v_h\}$ is a basis for a vector space *V* if (i) it is linearly independent, and (ii) its span is the entire vector space, i.e, any vector *v* in *V* can be written as $v = \sum_{k=1}^{h} \alpha_k v_k$. The α_k are the "coordinates" for *v*, with respect to the basis set $\{v_1,\ldots,v_h\}$.

Once we have chosen the basis set, the coordinates are unique. That is, if

$v = \sum_{k=1}^{h} \alpha_k v_k$ and also $v = \sum_{k=1}^{h} \beta_k v_k$, then $\sum_{k=1}^{h} \alpha_k v_k - \sum_{k=1}^{h} \beta_k v_k = 0$, so $\sum_{k=1}^{h} (\alpha_k - \beta_k) v_k = 0$, and each $\alpha_k - \beta_k = 0$ (since the basis set is linearly independent).

As an example: For the free vector space on *S*, a basis set consists of the vectors δ_s, one for each element *s* of S, defined as follows: $\delta_s(s) = 1$, and $\delta_s(t) = 0$ if $t \neq s$. To see that it is a basis, note that $f = \sum_{s \in S} f(s)\delta_s$. This is sometimes called the "one-hot" basis.

Note that if a set of vectors $\{v_1,\ldots,v_r\}$ is linearly independent, then it is always a basis set for something, namely, its span.

Slightly less obviously: if a set of vectors $\{v_1,\ldots,v_r\}$ spans a vector space *V*, then it always has a subset that is a basis for *V*. To see this: The original set could fail to be a basis if its members are not linearly independent, i.e., that there is some set of field elements β_k for which $\sum_{k=1}^{r}\beta_k v_k = 0$, with at least one $\beta_j \neq 0$. This means that $v_j = -\sum_{k=1,k\neq j}^{r}\frac{\beta_k}{\beta_j}v_k$. This in turn allows us to eliminate v_j from the set, and still be able to represent any vector:

$$v=\sum_{k=1}^{r}\alpha_k v_k = \sum_{k=1,k\neq j}^{r}\alpha_k v_k + \alpha_j v_j = \sum_{k=1,k\neq j}^{r}\alpha_k v_k - \alpha_j\sum_{k=1,k\neq j}^{r}\frac{\beta_k}{\beta_j}v_k = \sum_{k=1,k\neq j}^{r}\left(\alpha_k - \alpha_j\frac{\beta_k}{\beta_j}\right)v_k$$

We then continue eliminating until we can no longer find a relationship of linear dependence. The resulting set is the required basis. Note that the field properties (existence of a multiplicative inverse) play an important role.

Dimension

If the size of a basis set is finite, then this size is an intrinsic characteristic of the vector space, namely, its dimension.

We need to see that any two basis sets for a vector space have the same size (if the set size is finite). Suppose to the contrary, and that we've found the smallest such example. To be specific, say that $S_v = \{v_1,\ldots,v_h\}$ is a basis for *V*, and so is $S_w = \{w_1,\ldots,w_r\}$, with $h > r$. We need to show that this situation does not allow the elements of S_w to be linearly independent (but we don't want to resort to coordinates, counting degrees of freedom, etc.)

The proof is surprisingly tricky, and the reason is that the finiteness of the set size is critical. It relies on successively replacing each element of the smaller set (S_w) by one of the larger set, without changing its span. Once all elements have been replaced, then S_w has the same span as a proper subset of S_v, and this shows that the elements of S_v could not have been linearly independent.

Details: We begin by adjoining a vector from S_v to S_w. Since $S_w = \{w_1,\ldots,w_r\}$ is a basis, we can write v_1 as a linear combination of S_w , $v_1 = \sum_{k=1}^{r}\beta_k w_k$ At least one of the coefficients, say β_j, must be nonzero (since otherwise we would have found a linear relationship among the elements of S_w). So we can use this to eliminate w_j from the adjoined set. This results in a new basis set, containing v_1 and all the w_k except w_j. By our construction, the new set still spans *V*, and it is linearly independent. The latter follows because if there were a linear dependence, we could eliminate yet another vector, resulting in a smaller example. (We assumed we were dealing with the smallest example.)

We can now continue the swapping, each time bringing in another element of S_v. We can always eliminate a "w" from the augmented S_w, since, if there were a linear dependence among just the v's, then they could not have been linearly independent. After r steps, we've replaced all of the w's in S_w by a v , but there are still $h-r$ v's in S_v. So now there is a contradiction: at each stage, we showed that S_w is a basis, so now there must be a way to write one of these remaining v's as a linear combination of the ones we swapped into S_w.

The definitions of linear independence and of a basis set make sense for infinite sets $\{v_1,\ldots,v_h,\ldots\}$, i.e., for infinite-dimensional vector spaces. But one cannot claim that basis sets have a definite "size".

Combining vector spaces

General set-up here: V and W are vector spaces over the same field k.
$\{v_1,\ldots,v_m\}$ is a basis for V, and $\{w_1,\ldots,w_n\}$ is a basis for W.

Purposes: (a) a review of linear algebra, (b) setting up the material we need to describe how groups transform vectors (data), (c) a coordinate-free definition of the determinant.

Direct sum

The direct sum of V and W, $V\oplus W$, is a vector space consisting of ordered pairs of elements from V and W, i.e., (v,w).

Vector-space operations are defined as component by component: vector addition, $(v,w)+(v',w')=(v+v',w+w')$, and scalar multiplication, $\lambda(v,w)=(\lambda v,\lambda w)$.

Each of the properties required needed for $V\oplus W$ to be a vector space follow from the vector-space properties for V and W. For example, to show that $(\lambda\mu)(v,w)=\lambda\big(\mu(v,w)\big)$: $(\lambda\mu)(v,w)=((\lambda\mu)v,(\lambda\mu)w)=(\lambda(\mu v),\lambda(\mu w))=\lambda(\mu v,\mu w)=\lambda\big(\mu(v,w)\big)$. The first, third, and fourth equalities follow from the definition of scalar multiplication in the direct-sumspace, the second equality follows from the properties of scalar multiplication within V and W.

A basis for $V\oplus W$ is the $(m+n)$-element set, $\{(v_1,0),\ldots,(v_m,0),(0,w_1),\ldots,(0,w_n)\}$.

Homomorphism

$Hom(V,W)$ indicates the set of homomorphisms (structure-preserving maps) from *V* to *W*.

In this context, “preserves structure” means that the vector-space operations are preserved. That is, if $\varphi \in Hom(V,W)$, then $\varphi(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1\varphi(v_1) + \alpha_2\varphi(v_2)$. In other words, a homomorphism of vector spaces is a linear transformation. The above equation is a compact way of combining two aspects of linearity: scaling, $\varphi(\alpha v) = \alpha\varphi(v)$, and superposition. $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2)$.

It is worth noting that homomorphisms include not only the familiar linear transformations of a finite-dimensional vector space – such as rotations and projections -- but also, mappings from one infinite-dimensional vector space to another, or from one infinite-dimensional vector space to a finite-dimensional one. For example, say *V* is the space of infinitely-differentiable or analytic real-valued functions on the line (these are two ways of formalizing “smooth and otherwise nicely-behaved). Then the derivative, which maps a function $f(x)$ to the function $\frac{df}{dx}$, is a member of $Hom(V,V)$. And evaluation at a point *p*, which maps a function $f(x)$ to the value $f(p)$, is a member of $Hom(V,\mathbb{R})$.

The set of homomorphisms between two vector spaces is, itself, a vector space over *k*. To see this, we need to define the vector space operations among the homomorphisms. That is, if φ and ψ are both homomorphisms from *V* to *W* and λ is a scalar, we need to define vector addition, $\varphi+\psi$, and scalar multiplication, $\lambda\varphi$. Since both of these need to be in $Hom(V,W)$, we define them by their actions on V.

Addition: $\varphi+\psi$ is defined by $(\varphi+\psi)(v) = \varphi(v) + \psi(v)$.
(Right hand side is addition in *W*).

Scalar multiplication: $\lambda\varphi$ is defined by $(\lambda\varphi)(v) = \lambda(\varphi(v))$. (Right hand side is scalar multiplication in *W*).

The definitions are almost automatic, but it is worthwhile seeing how it is guaranteed that (a) they preserve the structure of *V* (and therefore are members of $Hom(V,W)$), and (b) that the vector space axioms are obeyed by the above definitions. For example, to show that $(\varphi+\psi)$ preserves the additive structure, we need to show that $(\varphi+\psi)(v_1+v_2) = (\varphi+\psi)(v_1) + (\varphi+\psi)(v_2)$. So we first apply the definition of $(\varphi+\psi)$, and then use the fact that both φ and ψ are homomorphisms from *V* to *W*:

$$\begin{aligned}&(\varphi+\psi)(v_1+v_2) = \varphi(v_1+v_2) + \psi(v_1+v_2)\\ &= \varphi(v_1) + \varphi(v_2) + \psi(v_1) + \psi(v_2)\\ &= \varphi(v_1) + \psi(v_1) + \varphi(v_2) + \psi(v_2)\\ &= (\varphi+\psi)(v_1) + (\varphi+\psi)(v_1)\end{aligned} .$$

Given a basis $\{v_1,\ldots,v_m\}$ for *V*, and $\{w_1,\ldots,w_n\}$ for *W*, we can build a basis for $Hom(V,W)$.

Consider the mapping φ_{ij} from *V* to *W* defined by $\varphi_{ij}(\sum_{k=1}^{m}\alpha_k v_k)=\alpha_i w_j$. In other words, let φ_{ij} (a) map the *i*th element of $\{v_1,\ldots,v_m\}$ to the *j*th element of $\{w_1,\ldots,w_n\}$, (b) map every other basis vector to zero, and (c) extend to the rest of *V* as required by linearity.

The set $S_{VW}=\{\varphi_{11},\ldots,\varphi_{1n},\varphi_{21},\ldots,\varphi_{2n},\ldots,\varphi_{m1},\ldots,\varphi_{mn}\}$ form a basis set for $Hom(V,W)$. (Also makes sense if either *V* or *W* is infinite-dimensional.)

To express an arbitrary φ in terms of S_{VW}, we note that we only have to express how φ acts on each basis element of *V* (since the fact that φ is linear allows us to extend the action of φ from any basis of IV to the whole space). So we simply express the action of φ on a basis element *vᵢ*. in terms of the basis $\{w_1,\ldots,w_n\}$ of *W*: $\varphi(v_i)=\sum_{j=1}^{n}\gamma_{ji}w_j$. Then, it follows from the definition of φ_{ij} that $\varphi=\sum_{i=1}^{m}\sum_{j=1}^{n}\gamma_{ji}\varphi_{ij}$.

To see that S_{VW} has no linear dependencies, we suppose that we had some linear combination of its elements that is zero: $\psi=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ji}\varphi_{ij}=0$. We need to show that this forces each of the c_{ji} to be 0. (The order of the subscripts of c_{ji} is reversed, because of matrix conventions, see below.) If $\psi=0$, then its action on every element of V must yield 0. Specifically, its action on any basis element v_k of *V* must be 0.
$0=\psi(v_k)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ji}\varphi_{ij}(v_k)=\sum_{j=1}^{n}c_{jk}\varphi_{kj}(v_k)=\sum_{j=1}^{n}c_{jk}w_j$, where the last two equalities follow from the definition of φ_{ij}. Since $\{w_1,\ldots,w_n\}$ is a basis set for *W*, $\sum_{j=1}^{n}c_{jk}w_j=0$ can only happen if all $c_{jk}=0$.

Coordinates

This should begin to look a lot like matrices. The connection is explicit if we choose specific basis sets $\{v_1,\ldots,v_m\}$ and $\{w_1,\ldots,w_n\}$.

We think of *V* as a set of *m* numbers in a column, and choose as a basis for *V* the following:

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, v_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$ So a vector $v = \sum_{k=1}^{m} \alpha_k v_k$ corresponds to $v = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}$

Similarly, think of *W* as a set of *n* numbers in a column, and choose

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, w_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, w_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$ So a vector $w = \sum_{j=1}^{n} \beta_j w_j$ corresponds to $w = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$.

Transformations from *V* to *W* can now be thought of as arrays of *n* rows, *m* columns:

$$\varphi = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nm} \end{pmatrix}.$$ A basis element φ_{ij}, which maps v_i to w_j, corresponds to a matrix in which the element in the *i*th column and *j*th row (γ_{ji}) is equal to 1, and all other elements are 0. With these correspondences, $\varphi v = w$ is equivalent to

$$\begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{n1} & \cdots & \gamma_{nm} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix},$$ with the usual rules of matrix multiplication: $\beta_j = \sum_{i=1}^{m} \gamma_{ji} \alpha_i$.

Note that the above would hold no matter what basis we choose – and there is no reason at this point to choose this particular basis. In another basis, the coordinates α_i, β_j, and γ_{ij} that represent particular vectors and transformations would be different.

If we fix a particular vector *v*, its representation in coordinates is *completely arbitrary* – i.e., by changing the basis, we could have used *any* set of numbers to represent it (other than all zeros). Put another way, if we are given the coordinates of a vector, we know nothing about its intrinsic properties, since the same set of numbers could represent any vector.

To what extent is the numerical representation of a member of $Hom(V,W)$ arbitrary? It can't be completely arbitrary, since the dimension of the range of φ is an intrinsic property, so it must somehow be embedded in the numerical representation. And there are other constraints that will arise when we consider $Hom(V,V)$.

Coordinate change, and a preview of the determinant

But first let's work out how a coordinate change affects the numerical representation of a an element φ in $Hom(V,W)$. As above, vectors v in V are represented by columns of length m, and vectors w in W are represented by columns of length n. The homomorphism φ is represented by a matrix of size $n \times m$, which we will call L, and $w = \varphi v$ corresponds to the ordinary linear algebra equation $w = Lv$.

Let's assume that the numerical representations of vectors in the new coordinate system and the original coordinate system for V are related by $v = Av'$. (We write the coordinate change this way, as opposed to $v' = Zv$, as it will avoid some inverses later on.) Similarly, for W, the change of coordinates corresponds to $w = Bw'$. We interpret these equation as in ordinary linear algebra: vectors are columns of field elements, A is an invertible $m \times m$ matrix, B is an invertible $n \times n$ matrix. To work out the representation of φ in the new coordinate system, we observe that $w' = B^{-1}w = B^{-1}Lv = B^{-1}LAv'$. So if φ is represented in the old coordinate system by the matrix L, in the new coordinate system, φ is represented by a matrix L' given by $L' = B^{-1}LA$. It is usually not confusing and simpler to just write $\varphi' = B^{-1}\varphi A$.

Similarly, an isomorphism φ in $Hom(V,V)$ is represented by an invertible $m \times m$ matrix L. With a change of coordinates in V given by $v = Av'$ (and $w = Aw'$), $w = Lv$ is equivalent to $w' = A^{-1}LAv'$, and $\varphi' = A^{-1}\varphi A$.

Thus, a change in coordinates $v = Av'$ in the vector space V induces a change in coordinates of $Hom(V,V)$. But we can also view $v = Av'$ (abstractly) as an isomorphism of V, and we see that it induces an isomorphism of the vector space $Hom(V,V)$, in which each L is mapped to $\Psi_A(L) = A^{-1}LA$. To check that Ψ_A is an isomorphism in $Hom(V,V)$, we need to verify that $\Psi_A(\alpha L) = \alpha\Psi_A(L)$ for any scalar α, that $\Psi_A(L+M) = \Psi_A(L) + \Psi_A(M)$ for any L and M in $Hom(V,V)$, and that Ψ_A has an inverse (see homework 2020-2021c, Q2).

There's one more piece of structure, and it's important. The isomorphisms of V have a natural group structure, since they are invertible transformations of V. The group structure corresponds to composition: if $v' = A^{-1}v$ and $v'' = v'B^{-1}$, then $v'' = v'B^{-1} = vB^{-1}A^{-1} = v(AB)^{-1}$. The correspondence from isomorphisms A of V to the isomorphisms Ψ_A within $Hom(V,V)$ preserves this structure: $\Psi_A \circ \Psi_B = \Psi_{AB}$, where the group operation $\circ$ means "followed by" That is,

$$\begin{aligned}&\left(\Psi_A \circ \Psi_B\right)(L) = \Psi_B\left(\Psi_A(L)\right) = \Psi_B\left(A^{-1}LA\right) = B^{-1}(A^{-1}LA)B = (B^{-1}A^{-1})L(AB)\\ &= (AB)^{-1}L(AB) = \Psi_{AB}(L)\end{aligned}.$$

Although we may have used coordinates to inspire this construction, the above equation can be interpreted in a coordinate-free way: *A* and *B* are just linear transformations on *V*.

This will carry through for other vector spaces that we build from *V*. The composition structure of the transformations on *V* will carry through to the composition structure on the vector spaces built from *V*. So if we can build a one-dimensional vector space from *V* – in which the only linear transformations are scalar multiplication and without resorting to coordinates -- we will have found a way to map transformations into scalar multiplication that preserves composition.

The dual space

An important special case of $Hom(V,W)$ is that of $W = k$, the base field. $Hom(V,k)$ is thus the set of linear mappings from V to the scalars, and is also known as the dual space of *V*, V^*.

If *V* has some finite dimension *m*, then the dimension of V^* is also *m* (since the dimension of *k* is 1).

Importantly, and perhaps non-obviously, there is no natural relationship between *V* and its dual. This seems surprising because in some sense, V^* and *V* have the same intrinsic structure – they are abstractly the same as a free vector space on a set of size *m*. The problem appears when we try to set up a correspondence between *V* and V^*. The obvious way to proceed is to take a vector in *V*, determine its coordinates with respect to some basis set $\{v_1, \ldots, v_m\}$, and then find the element in of V^* that has the same coordinates. The problem with this construction is that when we change coordinates, the vectors in *V* and in V^* change in different ways. Let's say we happened to have an element in *v* in *V* and an element φ in V^* that had an *intrinsic* relationship, for example, $\varphi(v) = 1$. For $v = \sum_{k=1}^{m} \alpha_k v_k$ and $\varphi = \sum_{j=1}^{m} \gamma_{1j}\varphi_{j1}$,

$\varphi(v) = \sum_{j=1}^{m} \gamma_{1j}\varphi_{j1}\left(\sum_{k=1}^{m} \alpha_k v_k\right) = \sum_{j=1}^{m} \gamma_{1j} \sum_{k=1}^{m} \alpha_k \varphi_{j1}(v_k) = \sum_{j=1}^{m} \gamma_{1j}\alpha_j$. If you transform the α-and γ-coordinates in the same way (e.g., double them, because you halved the lengths of the basis vectors), you would change the value of $\varphi(v)$, and this shouldn't happen – you would want $\varphi(v) = 1$ in all coordinate systems.

Working out the change of coordinates in the standard linear-algebra in a slightly more compact form: as above, vectors in V are represented by columns of length m, but now φ is represented by a row of size $m \times 1$, say, R. We want the value of applying the dual element φ to v to be independent of the change of coordinates, i.e., we want $\varphi v = \varphi' v'$. $Rv = RAv'$. So in the new coordinates, we need to represent φ' by RA. We multiplied v by A on the left, but we multiplied R by A on the right. That is, a coordinate change is $v = Av'$ in *V* corresponds to a coordinate change $\varphi' = \varphi A$ in V^*.

Note that in the above "matrix" model, if elements of *V* are represented by column vectors of length *m*, elements of V^* are represented by *row* vectors of length *m*. This what allows (actually, requires) the coordinates in *V* and V^* to change in different ways when you change basis sets.

This lack of a natural correspondence of *V* and V^* is fixed by adding a little more structure to *V*, namely, an inner product (or dot product). The dot-product implicitly defines distances, perpendicularity, projection, angles, etc. We can always (for finite-dimensional *V*) impose a dot-product once we have chosen coordinates, but it is important to recognize when this is arbitrary.

When dealing with a vector space of signals or stimuli, the dot-product is typically arbitrary. This means that the distinction between *V* (data) and V^* (mappings from data to the field) are different kinds of objects.

An important example is imaging data. Considering an image as a set of values at pixels, $x^{[i]} = \sum_{k=1}^{m} \alpha_k{}^{[i]} x_k$, elements of *V*. ($\alpha_k$ is the value at pixel *k*, x_k is an image consisting of a unit intensity at pixel *k*, and 0 elsewhere): One general task is to describe a set of images $\left\{x^{[i]}\right\}$; here, the simplest kind of solution would consist of an average image, $\frac{1}{N}\sum_{i=1}^{N} x^{[i]}$ image, which is an element of *V*. If you transformed the images (e.g., scaled them up or blurred them – both linear transformations), the average image would be transformed in the same way.

A second task is to distinguish one set of images from another, e.g., to distinguish $\left\{x^{[i]}\right\}$ from some other set $\left\{y^{[i]}\right\}$. Here, a useful strategy is to identify a "decision function" φ, for which the values of $\varphi(x^{[i]})$ are different from the values of $\varphi(y^{[i]})$. Simple (linear) decision functions are part of V^*, not *V*, even though you can describe them by their weights at each pixel. Decision functions and images are different kinds of objects – an image has dimensions of intensity units, while decision functions tell you how much you multiply a pixel value by to get a contribution to the decision function, so it has units of 1/intensity. If you scaled up the image, you'd want to scale down the weights to get the same value of the decision function applied to the image. If you blurred the image, you'd need to try to sharpen the decision function to keep its value unchanged. These show that images and decision functions transform in different ways.

Another example is the distinction between lights (described by an intensity at each wavelength) and neural mechanisms for color (described by mappings from lights into responses). Here, the vector space of lights is infinite-dimensional, $I(\lambda)$. Vector-space operations in the space of lights can be defined (e.g., superposition), but there is no first-principles way to make a correspondence between lights and mechanisms.

Tensor products

One more way to combine vector spaces. Strange at first, but this is the foundation for (a) finding the intrinsic properties of $Hom(V,V)$, and (b) making a bridge between linear procedures and nonlinear ones.

Same set-up: *V* and *W* are vector spaces over the same field *k*.

The "tensor product" of *V* and W, $V \otimes W$, is the set of elements $v \otimes w$ and all of their formal linear sums, e.g., $\lambda(v \otimes w) + \lambda'(v' \otimes w')$, along with the following rules for reduction:

$(v \otimes w) + (v \otimes w') = v \otimes (w + w')$,
$(v \otimes w) + (v' \otimes w) = (v + v') \otimes w$,
$\lambda(v \otimes w) = (\lambda v \otimes w) = (v \otimes \lambda w)$.

$v \otimes w$ is known as an "elementary tensor product."

Intuitively, the tensor product space is the substrate for functions that act bi-linearly on *V* and *W*. Put another way, if a function $f(v,w)$ acts linearly on each argument *v* and *w* separately, then it can always be extended to a function that acts linearly on $V \otimes W$. The first of the above rules ensure linearity when components are added in *V*, the second ensures linearity when components are added in *W*, and the third, ensures linearity for scalar multiplication.

When *V* and *W* are both finite-dimensional (with $\{v_1, \ldots, v_m\}$ is a basis for *V*, and $\{w_1, \ldots, w_n\}$ is a basis for *W*), then $V \otimes W$ is of dimension *mn*, and it has a basis consisting of the elementary tensor products $v_m \otimes w_n$. Writing *q* in $V \otimes W$ as $q = \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij} (v_i \otimes w_j)$, we can see that elements in $V \otimes W$ can be thought of as rectangular arrays, and they are added coordinate-by-coordinate.

The three laws, together, enable us to rewrite any elementary tensor product $v \otimes w$ in this basis. For if $v = \sum_{i=1}^{m} a_i v_i$ and $w = \sum_{j=1}^{n} b_j w_j$, then $v \otimes w = \sum_{i=1}^{m} \sum_{j=1}^{n} q_{ij} (v_i \otimes w_j)$ for $q_{ij} = a_i b_j$. Note, though, that generic elements *q* of $V \otimes W$ are not elementary tensor products, i.e., cannot be written as just one term $v \otimes w$, since this requires that $q = \sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij}(v_i \otimes w_j)$ where q_{ij} is "separable", i.e., $q_{ij} = a_i b_j$.

$V \otimes W$ has the same dimension as $Hom(V,W)$, but (just like for V and V^*), there is no intrinsic relationship between them. But see the homework (2012-2013c, Q1C): there is a coordinate-free relationship between $(V \otimes W)^*$ and $Hom(V,W^*)$.

A useful way to see that $V \otimes W$ and $Hom(V,W)$ are intrinsically different is to see how their representations change when we change coordinates in standard linear-algebra notation. We represent the coordinate change in V by an invertible $m \times m$ matrix A, and $v = Av'$; we represent the coordinate change in W by an invertible $n \times n$ matrix B, and $w = Bw'$. For $\varphi \in Hom(V,W)$ and $w = \varphi v$, $w' = B^{-1}w = B^{-1}\varphi v = B^{-1}\varphi A v'$, so $\varphi' = B^{-1}\varphi A$ (interpreted as arrays of field elements and ordinary matrix operations). That is, $\varphi'_{k,l} = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(B^{-1}\right)_{k,i}\varphi_{i,j}A_{j,l}$.

For $q \in V \otimes W$, standard linear algebra notation fails us. We can write q as a sum of elementary tensor products $q = \sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}(v_i \otimes w_j)$. Then, since $v_i = \sum_{k=1}^{m} A_{i,k}v'_k$ and $w_j = \sum_{l=1}^{n} B_{j,l}w'_l$, $v_i \otimes w_j = \sum_{k=1}^{m} A_{i,k}v'_k \otimes \sum_{l=1}^{n} B_{j,l}w'_l = \sum_{k=1}^{n}\sum_{l=1}^{m} A_{i,k}v'_k \otimes B_{j,l}w'_l$. Our goal is to write q in the new coordinates, i.e., $q = \sum_{k=1}^{m}\sum_{m=1}^{n} {q_{k,l}}'(v'_k \otimes w'_l)$ and express the $q'_{k,l}$ in terms of the $q_{i,j}$. To do this, we first apply the linearity rules for tensor products: $v_i \otimes w_j = \sum_{l=1}^{n}\sum_{k=1}^{m} A_{i,k}v'_k \otimes B_{j,l}w'_l = \sum_{l=1}^{n}\sum_{k=1}^{m} A_{i,k}B_{j,l}\left(v'_k \otimes w'_l\right)$. So now we can express q in terms of the elementary tensor products in the new coordinate system, $v'_k \otimes w'$:

$$\begin{aligned}
q &= \sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}(v_i \otimes w_j) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}\left(\sum_{l=1}^{n}\sum_{k=1}^{m} A_{i,k}B_{j,l}\left(v'_k \otimes w'_l\right)\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{m} q_{i,j}A_{i,k}B_{j,l}\left(v'_k \otimes w'_l\right) \\
&= \sum_{l=1}^{n}\sum_{k=1}^{m}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} q_{i,j}A_{i,k}B_{j,l}\left(v'_k \otimes w'_l\right)\right)
\end{aligned}.$$

So the coefficient $q'_{k,l}$ of $v'_k \otimes w'_l$ is $q'_{k,l} = \sum_{i=1}^{m} \sum_{j=1}^{n} q_{i,j} A_{i,k} B_{j,l}$.

We can extend the tensor product construction to $\left(V \otimes W\right) \otimes X$, etc. We can also verify that $\left(V \otimes W\right) \otimes X$ and $V \otimes \left(W \otimes X\right)$ are abstractly identical, and that there *is* a coordinate-free correspondence, namely, $(v \otimes w) \otimes x \leftrightarrow v \otimes (w \otimes x)$. (The way to show this is that both $\left(V \otimes W\right) \otimes X$ and $V \otimes \left(W \otimes X\right)$ are the same – they are the substrates for the tri-linear functions on *V*, *W*, and *X*.) This extends to *h*-fold tensor products. Since the associative law holds, we don't need to pay attention to the parentheses when we write out multiple tensor products.

Relationship to familiar (physical) tensors

What is the relationship to more familiar "tensors", such as the diffusion tensor and the conductivity tensor? The short, informal answer is that these objects are elements of a tensor product space in which *V* is 3-dimensional, and their entries transform like the above tensors, under coordinate transformation.

The diffusion tensor is based on a model that particles diffuse via Brownian motion in a medium that may be anisotropic. For ordinary Brownian motion in a 1-dimensional medium, the expected mean-squared distance moved by a particle in time *t* is proportional to time, and the proportionality is the diffusion constant, namely, $< x^2 >= Dt$. In a 3-d medium, if diffusion along each coordinate axis is independent, this generalizes to $< x^2 >= D_{xx}t$, $< y^2 >= D_{yy}t$, and $< z^2 >= D_{zz}t$. But one could imagine that the fastest axis of diffusion is along some oblique axis, i.e., that a cohort of particles released at the origin would tend to form a cloud that is elongated along an oblique axis. So the position of a typical particle along each axis need not be independent; i.e., $< xy > \neq 0$. Working out the physics leads to $< xy >= D_{xy}t$.

So for the "standard" *x*, *y*, *z* coordinate system for *V*, we can characterize the variances and covariances of the particle position by an array $D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}$. This array must be symmetric, since $D_{xy}t =< xy >=< yx >= D_{yx}t$.

To see that *D* conforms to our more abstract notion of a tensor, we have to verify that *D* transforms in the proper way when we change coordinates. It does. For example, if we choose units in V that leads to b times the numerical value for a vector that has the same physical length, then we multiply the numerical values of the variances by b^2, and hence, the entries of D by b^2. We also need to check that D transforms properly when we choose oblique axes, but this works out too.

In other physical situations, the tensor need not be symmetric. For example, a "conductivity tensor" M is (abstractly) a quantity in $(V^* \otimes V)^*$ such that $M(E \otimes v)$ gives the current in the direction v induced by an electric field E in V^*. An electric field is considered to be a member of the dual space since it is a way of assigning a value (the potential) to an exploring vector. The clue is that electric fields have units that include a *reciprocal* length (volts/cm), while vectors have units of length (cm).

The determinant

Here we use these ingredients to define the determinant and derive its properties.

Permutations act on tensor products of a vector space with itself

$V \otimes W$ has additional structure if $V = W$: This arises because of some homomorphisms on $V \otimes V$. We build these homomorphisms by permuting the copies of V.

In this simplest case, there is only one nontrivial permutation: a permutation τ that takes 1 to 2, and 2 to 1.This provides a homomorphism on $V \otimes V$ that swaps the first and second copies. Say $q = v^{[1]} \otimes v^{[2]}$, an elementary tensor product. Define $\sigma_\tau(q) = v^{[2]} \otimes v^{[1]}$ for elementary tensor products, and use linearity to extend σ_τ to all of $V \otimes V$. σ_τ is a homomorphism on $V \otimes V$.

For some tensors x, $\sigma_\tau(x) = x$ -- these are the "symmetric" tensors (such as the diffusion tensor). The symmetric tensors form a subspace of $V \otimes V$ -- one way to see this is that the symmetric tensors are the kernel of $\sigma_\tau - I$, as $(\sigma_\tau - I)(x) = 0$ is equivalent to $\sigma_\tau(x) = Ix = x$. So any transformation A that acts on V can be thought of as acting linearly in $V \otimes V$, and hence, in the symmetric subspace of $V \otimes V$. We want to generalize this idea, to find a one-dimensional subspace, derived from V, in which A acts linearly. In this example, if V has dimension n, then $V \otimes V$ has dimension n^2 (it has basis elements $v_i \otimes v_j$). The symmetric part of $V \otimes V$ has dimension $n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$, namely, the elements $v_i \otimes v_i$ and $\frac{1}{2}\left(v_i \otimes v_j + v_j \otimes v_i\right)$ (for $i \neq j$).

The first step in the generalization is that the above construction extends to h-fold tensor products $V^{\otimes h} = V \otimes V \otimes \cdots \otimes V$, and to any permutation τ on the set $\{1, 2, \ldots, h\}$. For

example, if $\tau(1)=4$, $\tau(2)=2$, $\tau(3)=1$, and $\tau(4)=3$ and $q=v^{[1]}\otimes v^{[2]}\otimes v^{[3]}\otimes v^{[4]}$, then $\sigma_\tau(q)=v^{[\tau(1)]}\otimes v^{[\tau(2)]}\otimes v^{[\tau(3)]}\otimes v^{[\tau(4)]}=v^{[4]}\otimes v^{[2]}\otimes v^{[1]}\otimes v^{[3]}$.
Thus, for each permutation τ, we have a homomorphism (actually, an isomorphism) on $V^{\otimes h}$.

The plan is to show that when we choose *h* to be the dimension *m* of *V*, that there is a unique one-dimensional subspace of $V^{\otimes h}$ that we can identify *without resorting to coordinates*. (This is the *m*-fold antisymmetrized tensor product space $anti(V^{\otimes m})$, whose elements are $anti(v^{\otimes m})$, which we will define below.)

Now, assume that we had some linear transformation *A* in $Hom(V,V)$. We could have carried out the above procedure on vectors *v* in *V*, or, on the transformed space vectors Av. Since $anti(V^{\otimes m})$ is one-dimensional, i.e., the field *k*, the two results $anti(v^{\otimes m})$ and $anti((Av)^{\otimes m})$, would have to differ by some scalar factor, i.e., by some multiplier in *k*. This number (the ratio) is the determinant of *A*, $\det(A)$: $\det(A)=\dfrac{anti((Av)^{\otimes m})}{anti(v^{\otimes m})}$, which is independent of $v^{\otimes m}$ as long as $v^{\otimes m}\neq 0$. Put another way, $\det(A)$ indicates how much $anti(V^{\otimes m})$ expands, when *V* is transformed by *A*.

Since we defined the determinant without resorting to coordinates for *V*, we are guaranteed that the determinant is independent of coordinates. The other key properties of the determinant follow immediately.

Property 1: The determinant of a product is the product of the determinants: $\det(BA)=\det(B)\det(A)$. Since applying *BA* to *v* is the same as applying *B* to *Av*, which is in turn the same as applying *A* and then *B*, we can calculate the expansion of $anti(V^{\otimes m})$ induced by *BA* two ways:

(1) apply *BA* to *v*; this yields $\det(BA)$.
(2) in stages: apply *A* to *v*, yielding a factor of $\det(A)$; apply *B* to Av, yielding a factor of $\det(B)$. Since the two results must be identical, $\det(BA)=\det(B)\det(A)$.
Symbolically,

$$\det(BA)=\frac{anti((BAv)^{\otimes m})}{anti(v^{\otimes m})}=\frac{anti((BAv)^{\otimes m})}{anti((Av)^{\otimes m})}\frac{anti((Av)^{\otimes m})}{anti(v^{\otimes m})}=\det(B)\det(A).$$

Property 2: The determinant of a mapping to a lower-dimensional space is zero. This will follow because we will show that for a vector space of dimension $\leq m-1$, the dimension of $anti(V^{\otimes m})$ is 0. The converse also holds: if the determinant is not zero, then the mapping covers the entire space (i.e., a basis is mapped to a basis). We will see that vectors must span an *m*-dimensional space to have a nonzero antisymmetrized tensor product of *m* vectors.

There is also a very nice, simple geometric view of the determinant as the quotient of $anti((Av)^{\otimes m})$ and $anti(v^{\otimes m})$: it is the amount that the volume of a parallelepiped spanned by v expands, when *v* is transformed to *Av*. The fact that this expansion factor is independent of the choice of *v* (i.e., the choice of parallelepiped) also has a simple geometric interpretation (from Bruce Knight): one could always space-fill one parallelepiped with smaller copies of another one of a different shape, and (just by counting) see that the volume expansion ratio has to be independent of shape.

Construction of the antisymmetrized tensor product

We want to generalize the following from 2 copies of *V* to multiple copies:
For any $q = v^{[1]} \otimes v^{[2]}$ in $V \otimes V$, we have a homomorphism $\sigma_\tau(q) = v^{[2]} \otimes v^{[1]}$ based on the permutation τ that takes 1 to 2, and 2 to 1. Similarly, we can write $\sigma_e(q) = q$, where *e* is the trivial permutation (that takes 1 to 1, and 2 to 2). Now define

$sym = \frac{1}{2}(\sigma_e + \sigma_\tau)$ and $anti = \frac{1}{2}(\sigma_e - \sigma_\tau)$. So
$sym(v^{[1]} \otimes v^{[2]}) = \frac{1}{2}\left(v^{[1]} \otimes v^{[2]} + v^{[2]} \otimes v^{[1]}\right)$, and
$anti(v^{[1]} \otimes v^{[2]}) = \frac{1}{2}\left(v^{[1]} \otimes v^{[2]} - v^{[2]} \otimes v^{[1]}\right)$. The homomorphisms *sym* and *anti* can be thought of as symmetrizing, and antisymmetrizing, the tensors $q = v^{[1]} \otimes v^{[2]}$. That is, $sym(q)$ is unchanged by swapping the components of *q*, and $anti(q)$ is negated by swapping the components of *q*.
(Note also that $sym(sym(q)) = sym(q)$, $anti(anti(q)) = anti(q)$, and $sym(anti(q)) = anti(sym(q)) = 0$).

For three copies, *sym* and *anti* are:

$$sym(v^{[1]} \otimes v^{[2]} \otimes v^{[3]}) = \frac{1}{6}(v^{[1]} \otimes v^{[2]} \otimes v^{[3]} + v^{[2]} \otimes v^{[1]} \otimes v^{[3]} + \\ v^{[3]} \otimes v^{[1]} \otimes v^{[2]} + v^{[1]} \otimes v^{[3]} \otimes v^{[2]} + v^{[2]} \otimes v^{[3]} \otimes v^{[1]} + v^{[3]} \otimes v^{[2]} \otimes v^{[1]})$$

and

$$anti(v^{[1]} \otimes v^{[2]} \otimes v^{[3]}) = \frac{1}{6}(v^{[1]} \otimes v^{[2]} \otimes v^{[3]} - v^{[2]} \otimes v^{[1]} \otimes v^{[3]} + \\ v^{[3]} \otimes v^{[1]} \otimes v^{[2]} - v^{[1]} \otimes v^{[3]} \otimes v^{[2]} + v^{[2]} \otimes v^{[3]} \otimes v^{[1]} - v^{[3]} \otimes v^{[2]} \otimes v^{[1]}).$$

The general form is

$$sym(v^{[1]} \otimes v^{[2]} \otimes \cdots \otimes v^{[h]}) = \frac{1}{h!}\sum_\tau v^{[\tau(1)]} \otimes v^{[\tau(2)]} \otimes \cdots \otimes v^{[\tau(h)]}$$

and

$$anti(v^{[1]}\otimes v^{[2]}\otimes\cdots\otimes v^{[h]})=\frac{1}{h!}\sum_{\tau}parity(\tau)\left(v^{[\tau(1)]}\otimes v^{[\tau(2)]}\otimes\cdots\otimes v^{[\tau(h)]}\right)$$

where the summation is over all permutations τ of $\{1,\ldots,h\}$, and $parity(\tau)$ is +1 or -1, depending on whether the number of pairwise swaps required to make τ is even or odd.

More compact form, writing $z=v^{[1]}\otimes v^{[2]}\otimes\cdots\otimes v^{[h]}$:

$sym(z)=\frac{1}{h!}\sum_{\tau}\sigma_{\tau}(z)$ and $anti(z)=\frac{1}{h!}\sum_{\tau}parity(\tau)\sigma_{\tau}(z)$.

The above makes it explicit that *sym* and *anti* are averages over a group (here, the permutation group).

The fact that *sym* and *anti* are averages over a group leads to two properties: if ρ is a permutation that swaps a single pair of indices, then $sym(\sigma_{\rho}z)=sym(z)$ and $anti(\sigma_{\rho}z)=-anti(z)$.

The *sym* property is straightforward:

$$sym(\sigma_{\rho}z)=\frac{1}{h!}\sum_{\tau}\sigma_{\tau}(\sigma_{\rho}z)=\frac{1}{h!}\sum_{\tau}\sigma_{\tau\circ\rho}(z)=\frac{1}{h!}\sum_{\tau}\sigma_{\tau}(z)=sym(z).$$

The first equality follows from the definition of *sym*, the second from the definition of the group operation (composition) for permutations. The third equality, which is the critical one, from the fact that, since composition by ρ is invertible (because of the group properties), a sum over all permutations $\tau\circ\sigma$ is the same as a sum over all permutations τ.

For the *anti* property, we also need to notice that $parity(\tau_1\circ\tau_2)=parity(\tau_1)parity(\tau_2)$. (That is, *parity* is a homomorphism from the permutation group to $\{+1,-1\}$ under multiplication – as we showed above.) This is because *parity* counts the number of pair-swaps, and $\tau_1\circ\tau_2$ can always be constructed by first applying the pairs needed to make τ_2, and then the pairs needed to make τ_1. So if ρ is a pairwise swap, $parity(\rho)=-1$, and $parity(\tau)=-parity(\tau\circ\rho)$. Consequently,

$$anti(\sigma_{\rho}z)=\frac{1}{h!}\sum_{\tau}parity(\tau)\sigma_{\tau}(\sigma_{\rho}z)=\frac{1}{h!}\sum_{\tau}parity(\tau)\sigma_{\tau\circ\rho}(z)=-\frac{1}{h!}\sum_{\tau}parity(\tau\circ\rho)\sigma_{\tau\circ\rho}(z)$$
$$=-anti(z)$$
.

To complete the construction of the determinant, we need to count the dimensions of $anti(V^{\otimes h})$ -- and to show that it is one-dimensional.

Dimension count

To count the dimensions of $anti(V^{\otimes h})$, we count the size of a basis. We start with a basis for $V^{\otimes h}$, and let *anti* act on it. A basis for $V^{\otimes h}$ can be built from the basis $\{v_1,\ldots,v_m\}$ for *V*: select one element of $\{v_1,\ldots,v_m\}$ for each of the *h* copies in the tensor product space. So a typical basis element is $v_{i_1}\otimes v_{i_2}\otimes\ldots\otimes v_{i_h}$, where each of the subscripts $i_1,\ldots,i_h$ is drawn from $\{1,\ldots,m\}$. We can write this compactly as $z_{\vec{i}}=v_{i_1}\otimes v_{i_2}\otimes\ldots\otimes v_{i_h}$.

What happens when *anti* acts on $z_{\vec{i}}$? If any of the subscripts $i_1,\ldots,i_h$ match, then we have to get 0. This is for the following reason. Say τ is a permutation that swaps two of the identical subscripts. On the one hand, $anti(z_{\vec{i}})=anti(\sigma_\tau(z_{\vec{i}}))$, since $z_{\vec{i}}$ and $\sigma_\tau(z_{\vec{i}})$ are identical. But since τ is a pair-swap, we also have $anti(z_{\vec{i}})=-anti(\sigma_\tau(z_{\vec{i}}))$. So both quantities must be 0.

So, *anti* maps a basis element $z_{\vec{i}}$ of $V^{\otimes h}$ to 0 if any of its subscripts match. If none of the subscripts match, *anti* maps $z_{\vec{i}}$ to a linear combination of distinct basis elements of $V^{\otimes h}$, which therefore cannot be 0. (For the same reason, $z_{\vec{i}}$'s with distinct subscripts must be linearly independent.) The dimension of $anti(V^{\otimes h})$, which is the count of the number of basis elements that do not map to 0, is the number of ways of choosing *h* distinct elements out of *m* – which is $\begin{pmatrix} m \\ h \end{pmatrix}=\frac{m!}{(m-h)!h!}$.

For *sym* – which we don't need for the determinant, but is useful for other purposes -- the dimension count is the number of ways of choosing *h* elements out of *m* that need not be distinct. This is $\begin{pmatrix} m-1+h \\ m-1 \end{pmatrix}=\frac{(m-1+h)!}{(m-1)!h!}$. (This follows from a standard counting argument, sketched here: To choose a list of *h* items out of the numbers $\{1,\ldots m\}$: imagine you start a counter at 1. At each instant, you take one of two options: either record ("R") the value on the counter or increment it ("I"). After you have made *h* "record" moves and $m-1$ "increment" moves, you will have chosen *h* numbers, possibly with repetition, and the counter will now read *m*, so the process terminates. Every unique set of choices corresponds to a unique sequence of *h* "R" moves and *m*-1 "I" moves. The number of ways of labeling a sequence of $m-1+h$ steps as either R's or I's is the above binomial coefficient.)

Note that for $h=2$, the dimension of the antisymmetric space is $\frac{m(m-1)}{2}$, and the dimension of the symmetric space is $\frac{m(m+1)}{2}$, which adds up to m^2, the dimension of $V\otimes V$. So for $h=2$, we have completely decomposed $V^{\otimes 2}=V\otimes V$ into two parts: $sym(V^{\otimes 2})$ and $anti(V^{\otimes 2})$, and there is nothing left over. For $h\geq 3$, there is a similar decomposition of $V^{\otimes h}$; it involves these two parts and additional parts with more complex symmetries.