# Handout #4: Vector spaces and the statistical algorithm

This handout is intended to serve as a user-friendly compendium of certain basic facts about vectors spaces, and about the way in which vector spaces are used in quantum mechanics to represent both physical states of systems and "observables" pertaining to those systems. I will try to be somewhat more rigorous than Albert and somewhat less rigorous than Hughes. Let's start with a review of elementary features of vector algebra.

## I. Vector spaces

### A. Examples

All of the following are examples of vector spaces:

(i) in a plane, the set of all arrows radiating from the origin;

(ii) the set of all pairs (x,y) of real numbers;

(iii) the set of all triples (x,y,z) of real numbers;

(iv) the set of all pairs (a,b) of *complex* numbers;

(v) more generally, the set of all n-tuples $(a_1,a_2,\ldots,a_n)$ of numbers—real or complex—for any n;

(vi) the set of all real-valued functions f(x) such that $\int(f(x))^2dx$ [evaluated from x = $-\infty$ to x = $+\infty$] is well-defined;

vii) the set of all complex-valued functions f(x) (of a real variable x) such that $\int f(x)f^*(x)dx$ [evaluated from x = $-\infty$ to x = $+\infty$] is well-defined ($f^*(x)$ is the complex conjugate of f(x)).

What makes each of these collections of mathematical objects *vector spaces* is simply that, in each case, there are well-defined operations of adding vectors and scaling vectors. (There's a bit more: see Hughes for a rigorous definition of what makes some set of mathematical objects a vector space.) Let's discuss these operations in turn.

B. Adding vectors

One central feature of a vector space is that any two vectors in it can be "added" to yield another vector. Thus, in the case of arrows radiating from the origin in a plane, we "add" them in the usual head-to-tail fashion. In the case of the vectors consisting of all pairs of real numbers, we add them as follows: (x,y) + (w,z) = (x+y,w+z). Similarly for vectors consisting of pairs of complex numbers, and indeed for vectors consisting of n-tuples of numbers, real or complex. In the case of the so-called "square-integrable" functions, we add them in the obvious way: e.g., (f+g)(x) = f(x) + g(x).

C. Scaling vectors

A second central feature of a vector space is that any vector in it can be "scaled" by multiplying it by a number. Thus, multiplying an arrow by a number is understood to either lengthen or shrink the arrow—and, if the multiplier is negative, to reverse the arrow's direction. Multiplying a pair (x,y) by a number c yields c(x,y) = (cx,cy); more generally, multiplying an n-tuple $(a_1,a_2,\ldots,a_n)$ by a number c yields $c(a_1,a_2,\ldots,a_n) = (ca_1,ca_2,\ldots,ca_n)$. Multiplying a function f(x) by a number c yields, not surprisingly, cf(x).

D. Linear combinations of vectors

Suppose that $\phi$ and $\psi$ are two vectors in some vector space. Then, for any two numbers a and b, $a\phi + b\psi$ is likewise a vector: for multiplying $\phi$ by a yields a vector, and multiplying $\psi$ by b yields a vector, and adding these two vectors yields a vector. More generally, suppose that $\phi_1$, $\phi_2$, …, $\phi_n$ are all vectors, and $a_1$, $a_2$, …, $a_n$ are all numbers; then there is a vector $\gamma$ such that

$$\gamma = a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n.$$

When an equation like this holds, we call $\gamma$ a *linear combination* of the vectors $\phi_1$, $\phi_2$, …, $\phi_n$.

E. Subspaces

Suppose we have some vectors $\phi_1$, $\phi_2$, …, $\phi_n$ in some vector space. Then the set of all possible linear combinations of these vectors is called a *subspace* for the vector space. As an example, suppose our vector space is just the ordinary three-dimensional Cartesian vector space ($R^3$). Then the subspaces we can construct from vectors in this vector space include (i) every line through the origin; (ii) every plane that includes the origin; (iii) the entire vector space $R^3$ itself; (iv) the "null" subspace consisting just of the zero vector.

### F. Bases

Suppose that some vectors $\phi_1$, $\phi_2$, …, $\phi_n$ in some vector space have the following feature: *any* vector in the vector space is equal to some linear combination of $\phi_1$, $\phi_2$, …, $\phi_n$. Then we will say that the vectors $\phi_1$, $\phi_2$, …, $\phi_n$ form a *basis* for the vector space. As an example, the vectors (0,1) and (1,0) form a basis for the vector space consisting of all pairs of real numbers. So do the vectors (1,2) and (5,17); so do the vectors (1,1), (3,4), (7,-7); etc. But notice that the vectors (1,1) and (2,2) *don't* form a basis for this vector space; for example, there is no way to get the vector (3,4) to come out as a linear combination of them.

## II. Inner product

The next important notion to understand is that of the *inner product* between any two vectors. Abstractly, the inner product is just a function that takes pairs of vectors $\phi$ and $\psi$ as input and produces as output a number which we will write $<\phi \mid \psi>$. There are three defining features of this function:

(i) $<\phi \mid \psi> = <\psi \mid \phi>^*$ (where c is a complex number, c* is its complex conjugate);

(ii) $<\phi \mid a\psi_1 + b\psi_2> = a<\phi \mid \psi_1> + b<\phi \mid \psi_2>$.

(iii) $<\phi \mid \phi>$ is real and $\geq 0$; $<\phi \mid \phi> = 0$ iff $\phi$ is the zero vector.

### A. Examples

The most familiar example of an inner product is the so-called "dot product": $(x,y)\bullet(w,z) = xw + yz$. We will write this as follows: $<(x,y) \mid (w,z)> = xw + yz$. Less familiar—at least, when characterized as an operation on vectors—is the usual inner product for the vector space consisting of all square-integrable functions: $\langle f(x) \mid g(x)\rangle = \int f^*(x)g(x)dx$, evaluated from $x = -\infty$ to $x = +\infty$.

B. Orthogonality

One of the crucial things that the notion of inner product allows us to do is to define a notion of "orthogonal" (i.e., "at right angles"), as applied to vectors. This is simple: we say that vectors $\phi$ and $\psi$ are orthogonal iff their inner product is zero, i.e. $\langle\phi \mid \psi\rangle = 0$.

C. Length

The other crucial thing that the notion of inner product allows us to do is to define a notion of "length". Again, this is easy: the length of a vector $\phi$ — typically written $|\phi|$ —is simply taken to be the (positive) square root of its inner product with itself: $|\phi| = \sqrt{\langle\phi \mid \phi\rangle}$.

D. Orthonormal bases

Now we can combine various definitions. Suppose we have some vectors $\phi_1, \phi_2, \ldots, \phi_n$ in some vector space. Suppose first that every vector in the vector space is a linear combination of these vectors. Suppose next that each of these vectors has length 1. And suppose finally that any two of these vectors are orthogonal to each other. Then this set of vectors $\{\phi_1, \phi_2, \ldots, \phi_n\}$ is called an *orthonormal basis* for the vector space. We will see that orthonormal bases are very important.

## III. The statistical algorithm

Now we can finally state a much more general version of the statistical algorithm. Suppose we have some physical system S. At any moment of time, S will have some physical state. We will take its physical state at a given moment of time to be represented by a vector in some vector space. (Not just any vector space! Rather, the vector space suitable for representing the physical states of S will be determined by the exact physical make-up of S: how many particles it contains, and what kinds they are. We mostly won't be concerned with the details.) Next, there are various *experiments* (or "measurements", if you want to indulge in horribly sloppy talk) we can perform on this system, and each such experiment will have various possible *outcomes.* And it turns out that there are rigorous rules for associating with each such experiment an *orthonormal basis*. And, finally, if we are given the vector that represents the state of our system S (at, say, the time we begin performing the experiment), and we are given the orthonormal basis associated with the experiment we are performing, then there is a neat algorithm for calculating the probability of any given outcome.

Let's see how this works. Suppose our system is in the state represented by vector $\psi$. And suppose we are performing experiment E, which is associated with the orthonormal basis $\{\phi_1, \phi_2, \ldots, \phi_n\}$. And suppose we want to calculate the probability of some particular outcome $\alpha$. Then the rules that associate the orthonormal basis $\{\phi_1, \phi_2, \ldots, \phi n\}$ with E will tell us that certain of the basis vectors "correspond to" the outcome $\alpha$. Suppose, for definiteness, that these are $\phi_2$, $\phi_8$, and $\phi_{17}$. Then to calculate the probability of $\alpha$, we proceed as follows:

<u>First step</u>: Express the vector $\psi$ as a linear combination of the vectors in the set $\{\phi_1, \phi_2, \ldots, \phi_n\}$: $\psi = a_1\phi_1 + a_2\phi_2 + \ldots + a_n\phi_n$. It turns out that there is a *unique* way to do this.

<u>Second step</u>: Take the coefficients $a_2$, $a_8$, and $a_{17}$, and square them.

<u>Third step</u>: Add the result. That number — $|a_2|^2 + |a_8|^8 + |a_{17}|^{17}$ — is the probability that performing experiment E on a system S in state $\psi$ (ooops! I mean: in a state *represented* by the vector $\psi$) will yield the outcome $\alpha$.