January 30

Hilbert Space

In order to explain the foundations of quantum mechanics, I first need to give you a math lesson. In this lecture, I will discuss vector spaces, both finite and infinite dimensional. We will find that this mathematical discussion gives a natural home to some of the concepts that we saw arising in the theory of the Schrödinger equation.

You are familiar with the notion of vectors. To a physics student, a vector is an n-tuple of numbers

$$v = (v_1, v_2 \cdots, v_n)$$

From a set of m linearly independent vectors $\{v_1, \ldots, v_m\}$, we can build additional vectors by taking linear combinations

$$w = a_1 v_1 + \cdots + a_m v_m$$

The set of all such linear combinations is a *vector space*. The set of all n-tuples is a vector space. In this space, there are n linearly independent vectors; any vector in the space can be represented as a linear combination of these. In fact, any set of n independent vectors can be used to build up the whole space. We call such a set of independent vectors a *basis* for the vector space.

Mathematicians enshrine this concept by giving a set of axioms for an abstract vector space. They define a vector space V as a set of elements $\{v_i\}$ such that the following operations are well defined and yield another element of the space: addition

$$v_1, v_2 \in V \quad \text{mple} \quad (v_1 + v_2) \in V$$

inversion

$v \in V$ implies $(-v) \in V$ such that $v + (-v) = 0$

and multiplication by a real or complex number

$v \in V$ implies $av \in V$

Addition and multiplication must also satisfy the usual commutative, associative, and distributive laws.

A vector space is *finite-dimensional* if there is a finite basis of vectors $\{t_i\}$, $i = 1, \ldots m$ such that any vector can be constructed as a linear combination.

$$V = \sum_{i=1}^{m} a_m t_m$$

The minimum value of m is the *dimension* of the space

From here on, I will specialize to vector spaces over the complex numbers. This is the case relevant to quantum mechanics.

An *inner product* on the vector space V is an operator that takes two vectors v_1, v_2 in V and returns a complex number. I will denote the inner product as

$$\langle V_1 | V_2 \rangle$$

(in math books, this is usually written $\langle v_1, v_2 \rangle$). The inner product should be linear in each argument, in the following way:

$$\langle V | a_1 W_1 + a_2 W_2 \rangle = a_1 \langle V | W_1 \rangle + a_2 \langle V | W_2 \rangle$$

$$\langle a_1 V_1 + a_2 V_2 | W \rangle = a_1^* \langle V_1 | W \rangle + a_2^* \langle V_2 | W \rangle$$

It should, further, be symmetric,

$$\langle w|v\rangle = (\langle v|w\rangle)^*$$

The postulate of symmetry implies that $\langle v|v\rangle$ is a real number. Then, finally, we demand that the inner product be positive

$$\langle v|v\rangle > 0 \qquad \text{unless} \qquad v = 0$$

For the vector space of n-tuples of complex numbers, there is a natural inner product that satisfies these axioms

$$\langle v|w\rangle = v_1^* w_1 + v_2^* w_2 + \cdots v_n^* w_n$$

Then

$$\langle v|v\rangle = \sum_1^n |v_i|^2$$

is the square of the length of the vector.

Following this, we define the *norm* of v as

$$\|v\| = (\langle v|v\rangle)^{\frac{1}{2}}$$

This satisfies the relation

$$\| a.v \| = |a| \cdot \| v \|$$

and the triangle inequality

$$\| x+y \| \leq \| x \| + \| y \|$$

Finally,

$$\| v \| = 0 \quad \text{or} \quad \langle v | v \rangle = 0$$

implies $v = 0$. Similarly, if two vectors v and w satisfy

$$\langle v | w \rangle = 0$$

we say that they are *orthogonal.*

We are used to thinking about finite-dimensional vector spaces, but there is no reason that a vector space cannot be infinite-dimensional. There are many systems that are usefully thought of infinite-dimensional vector spaces. We will meet some in a moment. To deal with this, it is useful to define: A *Hilbert space* is a vector space, not necessarily finite-dimensional, with a positive inner product, that is *complete* with respect to the metric $|x - y| = \| x - y \|$.

A complete space is one in which a convergent sequence has a limit point that belongs to the space. This is a very useful property. It allows us to prove that we can minimize a function and find the minimum as a point in the space. Any finite-dimensional vector space is complete, but this postulate gives us control also over infinite-dimensional cases.

An important class of Hilbert spaces is that of square-integrable functions on an interval. For example, consider the space of complex-valued functions satisfying

$$f(x) = 0 \qquad \text{for} \quad x = 0 \,, \quad x = 1$$

with

$$\int_0^1 dx \, |f(x)|^2 \qquad \text{finite}$$

We can define an inner product satisfying the axioms by

$$\langle f | g \rangle = \int_0^1 dx \, f^*(x) \, g(x)$$

The *distance* between two functions $f(x)$ and $g(x)$ is

$$\| f - g \| = \left[\int_0^1 dx \, |f(x) - g(x)|^2 \right]^{1/2}$$

Let's now return to the space of n-tuples. A common operation on n-tuples is matrix multiplication

$$w = Mv \qquad \text{or} \qquad w_i = \sum_j M_{ij} v_j$$

This operation is linear

$$M(av + bw) = a \cdot Mv + b \cdot Mw$$

It is useful to study

$$\langle v | Mw \rangle = \sum_{ij} v_i^* M_{ij} w_j$$

We can rearrange this expression into

$$\sum_{ij} (M_{ij}^* v_i)^* w_j$$

or

$$\sum_{ij} (M_{ji}^\dagger v_i)^* w_j = \langle M^\dagger v | w \rangle$$

defining

$$(M^\dagger)_{ji} = (M_{ij})^*$$

Then the inner product can be viewed either as one with M acting on w or with $M^\dagger$ acting on v. The matrix $M^\dagger$ is called the *adjoint* or *Hermitian conjugate* of M. For example

$$M = \begin{pmatrix} 1 & i \\ 2 & 3i \end{pmatrix} \qquad M^\dagger = \begin{pmatrix} 1 & 2 \\ -i & -3i \end{pmatrix}$$

A special class of matrices satisfies

$$M^+ = M$$

These are called *self-adjoint* or *Hermitian*. They satisfy

$$\langle v | Mw \rangle = \langle Mv | w \rangle$$

We can define the *eigenvalue problem* associated with a matrix in the same way that we defined the eigenvalue problem for the Schrödinger equation. The problem is to find vectors v_a and associated complex numbers λ_a such that

$$M v_a = \lambda_a v_a$$

The vectors are called the *eigenvectors* and the associated numbers are the *eigenvalues*. For example, for

$$M = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

It is easy to check directly that the following are eigenvectors and eigenvalues:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \lambda_1 = 3 \qquad\qquad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \lambda_2 = -1$$

The eigenvectors and eigenvalues of self-adjoint matrices have special properties. I will now prove two of these using a method similar to one that we used in our discussion of the Schrödinger equation. First, the eigenvalues λ_a are real. Consider

$$\langle v_a | M v_a \rangle = \langle v_a | \lambda_a v_a \rangle = \lambda_a \langle v_a | v_a \rangle$$

$$= \langle M v_a | v_a \rangle = \langle \lambda_a v_a | v_a \rangle = \lambda_a^* \langle v_a | v_a \rangle$$

Since $\langle v_a | v_a \rangle$ cannot be zero unless $v_a = 0$, we can divide through by this factor and find

$$\lambda_a = \lambda_a^*$$

Second, eigenvectors with different eigenvalues are orthogonal.

$$\langle v_a | M v_b \rangle = \lambda_b \langle v_a | v_b \rangle$$

$$= \langle M v_a | v_b \rangle = \lambda_a^* \langle v_a | v_b \rangle = \lambda_a \langle v_a | v_b \rangle$$

If $\lambda_b \neq \lambda_a$, these relations are compatible only if

$$\langle v_a | v_b \rangle = 0$$

These are the properties of the eigenvectors and eigenvalues of the Schrödinger equation that we proved in an earlier lecture. Notice how the axioms and definitions given here reduce these proofs to simple two-line arguments.

In an n-dimensional vector space, there can be at most n eigenvectors with distinct eigenvalues. Actually, it is not difficult to show that a self-adjoint matrix has exactly n independent eigenvectors. I will give the proof in detail here; it contains some important concepts.

To begin, let M be a self-adjoint matrix on a vector space of n dimensions. Define the function on vectors v

$$m(v) = \frac{\langle v | Mv \rangle}{\langle v | v \rangle}$$

Since

$$\langle v | Mv \rangle = \langle Mv | v \rangle = (\langle v | Mv \rangle)^*$$

this function is real-valued. It is also not difficult to show that the function has a lower bound. For example, if M_* is the largest element of M in absolute value, then

$$m(v) > -n|M_*|$$

This means that $m(v)$ has a minimum, and, by completeness, we can find the vector v_1 that minimizes $m(v)$.

I claim that v_1 is an eigenvector of M, with eigenvalue

$$\Lambda_1 = m(v_1)$$

I will give a proof by contradiction. Assume that

$$Mv_1 = Av_1 + w$$

with $w \neq 0$. If w has a component parallel to v_1, we could have included this term in Av_1; thus, we can assume that

$$\langle v_1 | w \rangle = 0$$

Then

$$\Lambda_1 = m(v_1) = \frac{\langle v_1 | Av_1 + w \rangle}{\langle v_1 | v_1 \rangle} = A$$

Now consider

$$m(v_1 + bw) = \frac{\langle v_1 + bw | M(v_1 + bw) \rangle}{\langle v_1 + bw | v_1 + bw \rangle}$$

where b is an arbitrary complex number. The denominator evaluates to

$$\langle v_1 + bw | v_1 + bw \rangle = \langle v_1 | v_1 \rangle + |b|^2 \langle w | w \rangle$$

$$= \langle v_1 | v_1 \rangle + \mathcal{O}(b^2)$$

The numerator is

$$\langle v_1 + bw | M(v_1 + bw) \rangle = \langle v_1 | Mv_1 \rangle + b^* \langle w | Mv_1 \rangle + b \langle v_1 | Mw \rangle + \mathcal{O}(b^2)$$

Now, by the self-adjoint property

$$\langle w | Mv_1 \rangle = \langle Mw | v_1 \rangle = (\langle v_1 | Mw \rangle)^*$$

so we can rewrite this as

$$\Lambda_1 + 2\,\mathrm{Re}\left[\, b\langle v_1|Mw\rangle \,\right] + \mathcal{O}(b^2)$$

We find, finally

$$m(v_1+bw) = \Lambda_1 + 2\,\mathrm{Re}\left[\, b\langle v_1|Mw\rangle\right] + \mathcal{O}(b^2)$$

In this argument, b was an arbitrary complex number. We can always choose a b such that

$$b\langle v_1|Mw\rangle$$

is a negative real number that is as small as we like. If this number is small enough that terms of order b^2 can be neglected, then

$$m(v_1+bw) < \Lambda_1 = m(v_1)$$

But v_1 was chosen to be the minimum of $m(v)$, so this contradicts the original assumption.

As a byproduct of this proof, we obtain an important result: If Λ_1 is the lowest eigenvalue of M, then

$$\Lambda_1 \leq \frac{\langle v|Mv\rangle}{\langle v|v\rangle} \quad \text{for any } v$$

We can converge to the eigenvector by improving the upper bound. This gives a method for finding v_1, called the *variational principle*, that is often useful in practice.

We have now found the first eigenvector of M. The same method can be used to find the rest. Let $V^\perp$ be the vector space of vectors w orthogonal to v_1

$$V^\perp = \{ x \in V \text{ such that } \langle x | v_1 \rangle = 0 \}$$

The matrix M operates within this space, since if $x \in V^\perp$, $Mx \in V^\perp$,

$$\langle v_1 | Mx \rangle = \langle M v_1 | x \rangle = \Lambda_1 \langle v_1 | x \rangle = 0$$

M is self-adjoint in $V^\perp$. So we can minimize $m(x)$ over vectors in $V^\perp$ to find a second eigenvector v_2. Proceeding in this way, we can find n eigenvectors (some of which might have equal eigenvalues). These n vectors are independent, and, in fact, orthogonal. Any other vector in the space can be written as a linear combination of these, so there are no more independent eigenvectors left to find.

We have now proved the *Spectral Theorem* for a n-dimensional Hilbert space V: If M is a self-adjoint matrix acting on V, M has n orthogonal eigenvectors, and these form a basis for V.

Any $w \in V$ can then be expanded

$$w = \sum_j c_j v_j$$

In our study of the Schrödinger equation, we found it useful to normalize the eigenfunctions, rescaling them so that

$$\|v_j\|^2 = \langle v_j | v_j \rangle = 1$$

Then the coefficients c_n are given by

$$c_j = \langle v_j | w \rangle$$

just as we saw in the Schrödinger equation examples.

Finding a eigenvectors and eigenvalues of a matrix is a well-studied problem in algebra. For 2×2 matrices, there is a relatively simple general solution. Consider in particular a general self-adjoint matrix

$$M = \begin{pmatrix} a & b \\ b^* & c \end{pmatrix} \qquad a, c \text{ real}$$

Let

$$v = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

be an eigenvector with eigenvalue λ. Then multiplying out the eigenvalue condition gives

$$a\alpha + b\beta = \lambda\alpha$$

$$b^*\alpha + c\beta = \lambda\beta$$

Let $\gamma = \alpha/\beta$. Then

$$b = (\lambda - a)\gamma$$

$$b^*\gamma = (\lambda - c)$$

Eliminating γ,

$$(\lambda - a)(\lambda - c) = |b|^2$$

This is a quadratic equation whose solutions are

$$\lambda_\pm = \frac{a+c}{2} \pm \left[\left(\frac{a-c}{2}\right)^2 + |b|^2\right]^{\frac{1}{2}}$$

These solutions are both real, as required. The corresponding eigenvectors are

$$\begin{pmatrix} \alpha_\pm \\ \beta_\pm \end{pmatrix} = \mathcal{C}_\pm \begin{pmatrix} \lambda_\pm - c \\ b^* \end{pmatrix} = \mathcal{D}_\pm \begin{pmatrix} b \\ \lambda_\pm - a \end{pmatrix}$$

For the discussion to follow, I will normalize these vectors to

$$|\alpha_+|^2 + |\beta_+|^2 = 1 \qquad |\alpha_-|^2 + |\beta_-|^2 = 1$$

Orthonormality implies that the matrix

$$U = \begin{pmatrix} \alpha_+ & \alpha_- \\ \beta_+ & \beta_- \end{pmatrix}$$

satisfies the condition

$$\begin{pmatrix} \alpha_+^* & \beta_+^* \\ \alpha_-^* & \beta_-^* \end{pmatrix} \begin{pmatrix} \alpha_+ & \alpha_- \\ \beta_+ & \beta_- \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

or

$$U^\dagger U = \underline{\underline{1}}$$

This matrix can be used to simplify the form of M. Notice that

$$MU = \begin{pmatrix} \lambda_+ \alpha_+ & \lambda_- \alpha_- \\ \lambda_+ \beta_+ & \lambda_- \beta_- \end{pmatrix}$$

and so

$$U^\dagger M U = \begin{pmatrix} \alpha_+^* & \beta_+^* \\ \alpha_-^* & \beta_-^* \end{pmatrix} \begin{pmatrix} \lambda_+ \alpha_+ & \lambda_- \alpha_- \\ \lambda_+ \beta_+ & \lambda_- \beta_- \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix}$$

That is, action by the matrix U transforms M to diagonal form, with the eigenvalues as the diagonal elements. The matrix U implements a change of basis from the original coordinate system to the basis of eigenvectors of M.

A matrix satisfying the property of U above

$$U^\dagger = U^{-1}$$

is called a *unitary* matrix. If U is unitary, matrix multiplication by U preserves the inner product,

$$\langle Uv | Uw \rangle = \langle v | U^\dagger U w \rangle = \langle v | w \rangle$$

The converse is also true: Any linear transformation of a vector space V that preserves all inner products is a unitary transformation.

Finally, I would like to connect these ideas to our results on the solutions of the Schrödinger equation. In the earlier lectures of this course, we solved a number of eigenvalue problems for the Schrödinger equation with different potentials. In each case, we were actually working with a Hilbert space. For example, the set of square-integrable functions on $(-\infty, \infty)$ such that

$$|f(x)| < C e^{-a|x|} \quad \text{as} \quad |x| \to \infty$$

is a complete vector space with inner product

$$\langle f|g\rangle = \int_{-\infty}^{\infty} dx\, f^*(x)\, g(x)$$

On this space, we can study the properties of the linear operator on the right-hand side of the Schrödinger equation

$$S = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x) \right)$$

The transformation

$$f(x) \to Sf(x)$$

is a linear operation on the space of functions. We could call this a matrix operation, though more typically we call $\mathbf{S}$ a *linear operator*.

Actually, the operator S is self-adjoint. We can prove this by integration by parts

$$\langle f | Sg \rangle = \int_{-\infty}^{\infty} dx \; f^*(x) \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x) \right) g(x)$$

$$= \int_{-\infty}^{\infty} dx \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} f(x) + V(x) f(x) \right)^* g(x) = \langle Sf | g \rangle$$

The boundary terms vanish by the restriction that I have placed on the behavior of the functions as $|x| \to \infty$. If the potential $V(x)$ is bounded below

$$V(x) > V_0$$

then operator $\mathcal{S}$ is also bounded below. To see this, integrate once by parts

$$\langle f | Sf \rangle = \int_{-\infty}^{\infty} dx \; f^* \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} f + V(x) f \right)$$

and rearrange this expression to

$$= \int_{-\infty}^{\infty} dx \left\{ \frac{\hbar^2}{2m} \frac{df^*}{dx}\frac{df}{dx} + f^* V(x) f \right\}$$

$$\int_{-\infty}^{\infty} dx \left\{ \frac{\hbar^2}{2m} \left| \frac{df}{dx} \right|^2 + V(x) |f(x)|^2 \right\} \geq V_0 \cdot \langle f | f \rangle$$

Then

$$\frac{\langle f | Sf \rangle}{\langle f | f \rangle} > V_0$$

We can add a constant to $V(x)$ without affecting the physics. For the argument below, it is useful to add such a constant so that

$$V_0 > 0$$

You can now see that the method used above to find the eigenvectors of a self-adjoint matrix works in this general context for a self-adjoint linear operator. We construct the expression

$$m(f) = \frac{\langle f | Sf \rangle}{\langle f | f \rangle}$$

which is now a real-valued function (or *functional*) of functions $f(x)$. We find a function f_1 in the Hilbert space that minimizes $m(f)$. This will be an eigenfunction of S. We then consider the smaller Hilbert space of functions orthogonal to f_1, and minimize $m(f)$ in that space to find a second eigenfunction f_2. Continuing in this way, we will find an infinite sequence of eigenfunctions f_n, with eigenvalues satisfying

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$$

In many of the Schrödinger equation problems that we studied, the spectrum of eigenfunctions was discretely spaced and extended to infinity. This was true, for example, for the Schrödinger particle in an infinite square well or a harmonic oscillator potential. Let me assume, that the spectrum of eigenvalues of S is such that, for any value Ω, no matter how large, there is an integer I such that

$$\lambda_j > \Omega \quad \text{for all} \quad j > I$$

This assumption is correct for the examples just given but not for the Hydrogen atom, which is a mathematically tricky special case. The assumption is also not correct for a problem with continuous spectrum, such as a free particle or a particle in a potential that levels off at infinity. In these cases we can make the spectrum discrete by putting the system in a very large box of size L.

In any event, I will now make the assumption above and work out its consequences. First, the assumption implies that, for a function $g(x)$ orthogonal to $f_1(x), \ldots, f_I(x)$,

$$m(g) = \frac{\langle g|Sg\rangle}{\langle g|g\rangle} > \Omega$$

Now consider $h(x)$, an arbitrary function in the Hilbert space. Let

$$S = \langle h|Sh\rangle$$

Choose some $\Omega \gg S$. For this Ω, find I and expand

$$h(x) = \sum_{n=1}^{I} c_n f_n(x) + g(x)$$

where $g(x)$ is orthogonal to the $f_i(x)$. We may assume that the $f_i(x)$ are normalized. Then

$$\langle h|h\rangle = \sum_{n=1}^{I} |c_n|^2 + \langle g|g\rangle$$

and

$$\langle h|Sh\rangle = \sum_{n=1}^{I} \lambda_n |c_n|^2 + \langle g|Sg\rangle$$

Then

$$S = \langle h|Sh\rangle > \sum_{n=1}^{\infty} \lambda_n |c_n|^2 + \Omega\langle g|g\rangle$$
$$> \Omega\langle g|g\rangle$$

so

$$\langle g | g \rangle < \frac{S}{\Omega}$$

By increasing Ω, it is possible to make this expression as small as we wish. This implies that

$$\| h - \sum_{n=1}^{I} c_n f_n \| \to 0$$

as I increases. More explicitly,

$$\int_{-\infty}^{\infty} dx \left(h(x) - \sum_{n=1}^{I} c_n f_n(x) \right)^2 \to 0$$

This proves a more general case of the Spectral Theorem. Let V be a Hilbert space V. If $\mathcal{S}$ is a self-adjoint operator on V that is bounded below, with discrete spectrum extending to infinity, then the eigenfunctions of $\mathcal{S}$ provide a basis for V. Any element of V can be approximated arbitrarily well, by the criterion of the integral above, as a linear combination of eigenvectors.

Using more sophisticated mathematical techniques, it is possible to extend this theorem to more general classes of self-adjoint operators, and also to problems with continuous spectra. It is also possible to prove that, if the function $h(x)$ is smooth, it is well approximated not simply in the integral but at every point. Further details can be found, for example, in the textbooks on Functional Analysis by Walter Rudin and John Conway.

The Spectral Theorem brings many ideas from our study of the Schrödinger equation into clearer focus. Some special cases of this theorem are:

1. Any function $f(x)$ on an interval $[a, b]$ can be represented by a Fourier series.
2. Any function $f(x)$ on the interval $[-1, 1]$ can be represented as a linear combination of Legendre polynomials.

3. Any function $f(\theta, \phi)$ on the sphere can be represented as a linear combination of spherical harmonics.

More generally, the concepts of Hilbert space provide the proper general context for understanding the solutions of the Schrödinger equation. In fact, they allow us to move beyond the Schrödinger equation and to formulate a general framework for quantum mechanics. We will see that in the next lecture.