

Chapter 1

The Story

Following the work of Thompson (See "The Origins of Dalton's Chemical Atomic Theory"), physicists came to the conclusion that matter was composed of atoms, fused into molecules. In 1897 (see "Thompson's 1906 Nobel Lecture: Carriers of Negative Electricity"), physicist J.J. Thompson argued that all negative electrical charge was carried by particles carrying both mass and electricity. Thus the first elementary particle was discovered. Using cathode rays, he measured the ratio between the unit of electrical charge e and the mass m of an electron. In 1913, Millikan found an experiment to measure the electrical charge individually. The mass is on the order of 9×10^{-31} kilograms, and the electrical charge is on the order of 2×10^{-19} Coulombs. The resolution of current measurement technology still leads us to believe that the electron is a *point particle*, i.e. it has no extent, existing in a single point in space.

Accumulated matter, like solid materials or gasses, on averages carries little to no electrical charge, or objects at the human scale would constantly be repulsed from one another. Thus there must exist matter with positive charge. Thompson speculated that electrons were embedded in a uniform sphere of jelly-like positive charge – the *plum pudding model*. However, Rutherford proved in 1903 that the carriers of positive charge must also behave like points; by firing a positively charged helium nuclei (an alpha particle) through a thin metal foil. The metal foil is much heavier than the alpha particle, The even distribution of positive charge lead Rutherford to expect that the helium nuclei would pass through the foil. However, Rutherford instead observed that the helium atoms scattered in various directions – distributed in the same way as would be expected via the Kepler problem, where the trajectories are hyperbolas.

It follows from Rutherford's experiment that the positive charge in the metal foil must be concentrated in very small regions of space; Rutherford had discovered that electrons existed separately from positively charged nuclei. Each nucleus can have various different types of charge, itself being formed from several more elementary particles; positively charged protons, and chargeless neutrons. We often let z denote the number of protons in a nucleus. In nature, z lies between $\{1, \dots, 92\} - \{43, 61, 85\}$; the other values $\{43, 61, 85, 93, \dots, 109\}$ have all been produced in experiments, but decay very quickly. The values $\{84, \dots, 92\}$ are also unstable, decaying over time, and producing *radioactive energy*.

Though consisting of several elementary particles, a nucleus itself is not a point mass. Nonetheless, in practice the protons and neutrons are concentrated in a region of radius 10^{-15} meters, which is often much smaller than the scale of the problem under consideration, and so a nucleus can also be treated as a point mass. The mass of a proton is on the order of 2×10^{-27} kilograms, more than

1000 times the mass of an electron. The mass of a naturally occurring nucleus is observed to have mass between z protons, and $3z$ protons. Thus a nucleus is in practice much much heavier than an electron, and can be considered infinite for most points, i.e. unmovable relative to the motion of an electron, just like the sun is often regarded as having infinite mass relative to the motion of planets. The protons repel one another, but are held together by the *nuclear force* between themselves, and the neutrons in the nucleus, and is about 100 times stronger than the corresponding electrical force. Thus for most of our purposes it is natural to consider a nucleus as a point particle with infinite mass.

With the plum pudding model thrown out, it was reasonable to try and apply the theory of electrostatics to electrons and nuclei, existing apart from one another; electrostatics tells us that the potential energy between two particles with charges q_1 and q_2 existing at two points $x, y \in \mathbb{R}^3$ is given by

$$k \frac{q_1 q_2}{|x - y|},$$

where $k = 9 \times 10^9 \text{Nm}^2/\text{C}^2$ is Coulomb's constant.

Consider the simplest possible case; the analysis of the Hydrogen atom, with one electron, and one proton. Placing the proton at the origin, and suppose the electron is at a position p , with momentum q , the Hamiltonian of the system is given by

$$H(p, q) = \frac{p^2}{2m} - \frac{ke^2}{|x|}.$$

This Hamiltonian is, up to constants, identical to the celestial two body problem; Kepler's analysis thus shows that stable orbits of the electron around the Helium nucleus are ellipses, with the nucleus at one focus. As the energy of the atom decreases, the electron gets closer and closer to the nucleus.

This analysis, however, results in several paradoxes. Firstly, an orbiting electron about a proton is an accelerating, charged mass. Maxwell theory of electromagnetism tells us that such an object should emit electromagnetic energy away from the mass. As the energy decreases, the electron orbits closer and closer to the nucleus of the atom. But the electron can get arbitrarily close to the proton, since both are point masses, and as this happens the energy of the atom can become arbitrarily negative. But this means that hydrogen atoms would continuously emit an infinite amount of energy over time; matter in the universe would always tend to shrink indefinitely in size, emitting electromagnetic waves as they do so. But we don't observe this happening.

Quantum mechanics fixes this problem. As an electron is forced to get closer and closer to a proton, the *uncertainty principle* tells us the kinetic energy of the electron is forced to increase. Thus the total energy of the hydrogen atom actually increases to ∞ as the electron gets closer and closer to the atom.

Chapter 2

The Setup

In any physical theory, we must characterize mathematically the *state* of a system (all information describing the situation of a physical system at a particular time), and the *observables*, the functions of a state, which give ways in which the state of a system can be reduced to quantities that can be observed experimentally. For instance, in Hamiltonian classical mechanics, the state of a system is given by a point in a symplectic manifold M , and observables given by functions $f : M \rightarrow \mathbb{R}$, which should be continuous if we are to correctly measure these observables up to a small degree of error. The observables are then ‘second order’ as they are defined in terms of states, but we can also reverse the situation, describing the observables as the C^* algebra $A = C(M)$. The states (elements of M) are then identified by a *positive* linear functional $\phi : A \rightarrow \mathbb{R}$ with $\phi(1) = 1$ (the evaluation functional at the corresponding point in M).

It is natural in the later quantum mechanics to complexify the C^* algebra A . Then the observables become the *self-adjoint* elements of A , and the states the linear functionals $\phi : A \rightarrow \mathbb{C}$ with $\phi(1) = 1$ and with $\phi(X) \geq 0$ if $X \geq 0$. In our original setting of a symplectic manifold M , our states are now extended to include all non-negative Borel probability measures on M . The Riesz representation theorem allows us to identify an arbitrary positive linear functional $\phi : A \rightarrow \mathbb{R}$ such that $\phi(1) = 1$ with a Borel probability measure μ on M . We then think of an element $X \in A$ as a *random variable* over the probability space (M, μ) , because we then have

$$\mathbb{E}_\phi[X] = \int X d\mu = \phi(X).$$

Similarly,

$$\sigma_\phi(X)^2 = \mathbb{V}_\phi(X) = \phi(X^2) - \phi(X)^2.$$

The *pure*, deterministic states ϕ can then be identified from general *mixed states* as the *extremal points* of the convex set of probability measures.

However, this formulation fails to explain quantum mechanical phenomena. The most fundamental experimental observation in the theory is the *uncertainty principle*. The behaviour of atomic spectra lead Heisenberg to postulate that in any physical system, if $p : A \rightarrow \mathbb{C}$ and $q : A \rightarrow \mathbb{C}$ are the position and momentum observables, then for any state ϕ ,

$$\sigma_\phi(p)\sigma_\phi(q) \geq \hbar/2,$$

where \hbar is *Planck's constant*. But in the algebra $A = C(M)$, there are no two observables $\phi : A \rightarrow \mathbb{R}$ with this property for all classical states, because

$$\sigma_\phi(p) = \sigma_\phi(q) = 0$$

for any deterministic state. Thus it appears that the only physically possible states ϕ *must be uncertain* in a suitable sense; this is the *uncertainty principle*.

It turns out the correct approach is to model p and q as (unbounded) self-adjoint operators on a Hilbert space H . A *pure state* is then a functional ϕ corresponding to some $x_0 \in H$ with $\|x_0\| = 1$, such that for any bounded operator A ,

$$\phi(A) = \langle Ax_0, x_0 \rangle.$$

The advantage to this approach is that self-adjoint operators on a Hilbert space have a rich spectral calculus, which allows us to write

$$A = \int \lambda d\pi$$

for a projection valued measure $d\pi$. Thus we can define operators $A_E = \int_E \lambda d\pi$ for any bounded set E . Given a state ϕ , we thus have a probability distribution

$$\mathbb{P}_\phi(A \in E) = \phi(A_E)$$

Note, in particular, that the support of the distribution is $\sigma(X)$.

As it comes to the Heisenberg uncertainty principle, if X and Y are any observables with $\phi(X) = \phi(Y) = 0$, one has

$$\phi(X^2)\phi(Y^2) \geq \phi(i[X, Y])^2/4.$$

In general, this means that for any state ϕ , and any $X, Y \in A$,

$$\sigma_\phi(X)\sigma_\phi(Y) \geq \phi[i[X, Y]].$$

Thus if we pick X and Y where $[X, Y] = -i\hbar$, then in any state ϕ , the uncertainty principle holds. Such a choice is possible provided we take X and Y to be *unbounded operators*.

Remark. As an aside, to prove the inequality above, note that the matrix

$$M = \begin{pmatrix} \phi(X^2) & 0.5\phi(i[X, Y]) \\ 0.5\phi(i[X, Y]) & \phi(Y^2) \end{pmatrix}$$

is positive-semidefinite, since for any $v = (\alpha, \beta)^T \in \mathbb{R}$,

$$\begin{aligned} v^T M v &= \phi(X^2)\alpha^2 + \phi(i[X, Y])\alpha\beta + \phi(Y^2)\beta^2 \\ &= \phi((\alpha X - i\beta Y)(\alpha X + i\beta Y)) \\ &\geq 0. \end{aligned}$$

Thus $\det(M) = \phi(X^2)\phi(Y^2) - \phi(i[X, Y])^2/4$ is non-negative.

For position and momentum, a natural choice over $H = L^2(\mathbb{R})$ are the unbounded self-adjoint operators

$$Xf \mapsto xf \quad \text{and} \quad Mf = -if'.$$

This choice leads us to think of the state $\phi(A) = \langle Af_0, f_0 \rangle$, with $\|f_0\|_{L^2(\mathbb{R})} = 1$, as corresponding to a random state, whose position is given randomly with probability density function $|f_0|^2$, and whose momentum is given randomly with probability density function $|\hat{f}_0|^2$ (Thus f_0 and λf_0 have the same state for any complex number λ with $|\lambda| = 1$). Given this assumption, an operator X representing position would have to have

$$\langle Xf_0, f_0 \rangle = \int x|f_0(x)|^2 dx$$

and a momentum operator M would have to satisfy

$$\langle Mf_0, f_0 \rangle = \int \xi |\hat{f}_0(\xi)|^2 d\xi,$$

and these operators are precisely those given above.

The class of states on $B(H)$ is in general much more complicated than the pure states induced by elements of H . But if we limit ourselves to a smaller C^* subalgebra of $B(H)$, the complexity of the kinds of states we can consider also becomes smaller. In particular, the states on the C^* algebra $K(H)$ of compact operators on H correspond to functionals of the form

$$\phi(A) = \text{Tr}(DA),$$

where D is a positive semidefinite Hermitian operator of trace one. The pure states of this C^* algebra (the extremal states) are precisely those given by some vector in H . The main reason we deal with a larger family of states is that it allows us to discuss random quantum systems (a quantum system drawn at random from a classical probability distribution), and to discuss the phenomenon of entanglement more easily, e.g. by taking conditional expectations: Given two quantum systems over the Hilbert spaces H_1 and H_2 , the two systems considered together have state space $H_1 \otimes H_2$ (e.g. two \mathbb{R} valued systems should correspond to a \mathbb{R}^2 valued system, and $L^2(\mathbb{R}^2) = L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$). Given a pure state ϕ on $H_1 \otimes H_2$ given by $f \in H_1 \otimes H_2$, and given an observable A in H , the expected value of A on the state ϕ should be

$$\phi(A \otimes I) = \langle (A \otimes I)f, f \rangle.$$

If $f = f_1 \otimes f_2$, where $\|f_1\| = \|f_2\| = 1$, then $\phi(A \otimes I) = \langle Af, f \rangle$, and so we obtain a classical state on H_1 . But if f is not a simple tensor product, the induced state is not in general a pure state.

2.1 Classical and Quantum Probabilities

Often we deal with an intermixing of classical and quantum probability. For instance, we may consider a system given by drawing randomly from a family of quantum systems with respective states ϕ_1, \dots, ϕ_n . If ϕ_j is selected with probability p_j , where $\sum p_j = 1$, the expected value of an observable A is given by

$$\sum p_j \phi_j(A).$$

The operator $A \mapsto \sum p_j \phi_j(A)$

This gives rise to a new random state which cannot be written purely in the quantum form. In this setting, a state is a positive semidefinite operator $D \in B(H)$ with $\text{Tr}(D) = 1$; if H is finite dimensional, D is called a *density matrix*. Given an observable A , the expected value of the observable with respect to the state D is

Chapter 3

Quantum Information Theory

The simplest unit of information in classical physics is a *bit*, represented by an element of $\{0, 1\}$. We can generalize the state of a collection of n bits are represented by an element of $\{0, 1\}^n$. From the quantum perspective, a *quantum bit*, or *qubit*, is represented by an element $\psi = \psi_0\langle 0| + \psi_1\langle 1|$ of a two dimension Hermitian product space with orthonormal basis $\{\langle 0|, \langle 1|\}$.

(Gleason)

Chapter 4

How can the Quantum Hamiltonian operator be Formed From the Classical Hamiltonian?

Thanks to Viktor T. Toth for this explanation on Quora. The Hamiltonian of a system is given by

$$H = p^2/2m + V(q),$$

where p is the momentum, and q the position. If we set

$$\psi = e^{(i/\hbar)(p \cdot q - Ht)},$$

then we notice that we can rewrite the Hamiltonian system above as

$$0 = (H - p^2/2m - V(q))\psi = H\psi - \frac{p^2}{2m}\psi - V(q)\psi.$$

Note that

$$H\psi = i\hbar \frac{\partial \psi}{\partial t} \quad \text{and} \quad p\psi = -i\hbar \nabla_q \psi,$$

and so

$$i\hbar \frac{\partial \psi}{\partial t} + \frac{\hbar^2}{2m} \Delta_q \psi - V(q)\psi = 0,$$

or equivalently,

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \Delta_q \psi + V(q)\psi,$$

which begins to look like the Schrödinger equation. In *classical physics*, we restrict ourselves to solutions of this equation of the form $e^{(i/\hbar)(p \cdot q - Ht)}$, from which we can extract out the particular positions and frequencies of objects. *Quantum mechanics* begins when we now allow *linear combinations* of solutions of this form, i.e. 'mixed states'.

Chapter 5

Problems Connecting Relativity and Quantum Mechanics

Consider the Schrödinger Equation

$$\partial_t f = Hf,$$

where H is the Hamiltonian operator that measures the energy of f , which we can write as

$$Hf = Vf - \frac{\hbar^2}{2m}\Delta,$$

where Vf is pointwise multiplication of f with V . In the Schrödinger equation, we have one derivative in the time variable, and two derivatives in the spatial variable. This causes problems when relating quantum mechanics to relativity, where space and time are meant to be interchangeable. This means that an analogue of the Schrödinger equation in relativity theory must also have first order terms in space and time, which is a problem because it can give the Hamiltonian negative spectrum.

One example of an attempt to define this is the relativistic Dirac equation on Minkowski space, which takes the form

$$(i\hbar \sum \gamma^\mu \partial_\mu - mc)\psi = 0,$$

where $\psi : \mathbb{R}_t \times \mathbb{R}_x^3 \rightarrow \mathbb{C}^4$ is called a *Dirac spinor field*, and where $\gamma^0, \gamma^1, \gamma^2, \gamma^3$ are some set of four complex matrices such that $[\gamma^\mu, \gamma^\nu] = 2\eta^{\mu\nu}I_4$, where η is the Minkowski metric. A common choice are

$$\gamma^0 = \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \quad \gamma^j = \begin{pmatrix} 0 & \sigma^j \\ \sigma^j & 0 \end{pmatrix}$$

where σ^j are the Pauli matrices. The presence of a negative spectrum here causes an incompatibility with quantum mechanics, but leads to interesting physical effects; an analysis of the Dirac equation lead to the speculation of the existence of *antiparticles*.