Basic Books in Science

Book 11

The Quantum Revolution



Roy McWeeny

BASIC BOOKS IN SCIENCE

- a Series of books that start at the beginning

Book 11

The quantum revolution

Roy McWeeny

Professore Emerito di Chimica Teorica, Università di Pisa, Pisa (Italy)

The Series is maintained, with regular updating and improvement, at http://www.learndev.org/ScienceWorkBooks.html and the books may be downloaded entirely free of charge



This work is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License

(Last updated 10 November 2011)

BASIC BOOKS IN SCIENCE

Acknowledgements

In a world increasingly driven by information technology no educational experiment can hope to make a significant impact without effective bridges to the 'user community' – the students and their teachers.

In the case of "Basic Books in Science" (for brevity, "the Series"), these bridges have been provided as a result of the enthusiasm and good will of Dr. David Peat (The Pari Center for New Learning), who first offered to host the Series on his website, and of Dr. Jan Visser (The Learning Development Institute), who set up a parallel channel for further development of the project. The credit for setting up and maintaining the bridgeheads, and for promoting the project in general, must go entirely to them.

Education is a global enterprise with no boundaries and, as such, is sure to meet linguistic difficulties: these will be reduced by providing translations into some of the world's most widely used languages. Dr. Angel S. Sanz (Madrid) is preparing Spanish versions of the books and his initiative is most warmly appreciated.

We appreciate the interest shown by universities in Sub-Saharan Africa (e.g. University of the Western Cape and Kenyatta University), where trainee teachers are making use of the Series; and that shown by the Illinois Mathematics and Science Academy (IMSA) where material from the Series is being used in teaching groups of refugee children from many parts of the world.

All who have contributed to the Series in any way are warmly thanked: they have given freely of their time and energy 'for the love of Science'.

Roy McWeeny (Series Editor)

BASIC BOOKS IN SCIENCE

About this book

This book, like the others in the Series¹, is written in simple English – the language most widely used in science and technology. It builds on the foundations laid in earlier Books, which have covered many areas of Mathematics and Physics.

The present book continues the story from Book 4, which introduced the science of Mechanics and showed how it could account for the motion of objects ranging from particles, of small mass, to very massive bodies like the planets that move around the sun. The *force*, which makes such bodies move, is due to 'gravity' – the invisible **attraction** between massive objects. We said there was a **field** of force at any point in space and it was proportional to the **mass** m of the body on which it acted. But in Book 10 we learnt that, besides having a mass m, a body may carry an **electric charge** q; and that another force may then act on the body when it is placed in an **electric field**. Book 10, which developed the science of **Electromagnetism**, was largely about fields in general (whether gravitational, electric, or magnetic).

By the end of the nineteenth century it seemed that the foundations of Physics were almost completely understood. But then came a few great surprises, which saw the end of "classical physics" and the beginning of "relativistic and quantum physics". The theory of relativity was touched on in Books 2 and 4, where it was noted that classical physics couldn't explain what happened when things moved at enormous velocity; but then in the 1920's classical *mechanics* failed completely to account for the behaviour of particles as small as electrons and protons – which define the natural 'units' of negative and positive charge. Classical mechanics had to be replaced by **quantum mechanics**, which required another 'revolution' in scientific thinking. The 'quantum revolution' led to a period of unprecedented progress in science and its applications, which has changed the lives of all of us.

¹The aims of the Series are described elsewhere, e.g. in Book 1.

Looking ahead –

This book is devoted to the second great scientific revolution of the last century, the development of Quantum Mechanics as the necessary replacement of Newtonian dynamics in dealing with the motion of the ultimate constituents of matter – particles like electrons and protons, which may carry electric charges and interact with each other (even when they are not in contact) through the *fields* they produce. You've learnt a lot about particles and fields in Book 10. In the present Book 11 you'll find out why such small particles must be described in a totally different way from the ones we can see and handle in the laboratory; and you'll be able to follow in detail the development (which took the better part of 50 years!) of the new principles and theories that you'll need before going on to thhe study of atoms and molecules (Book 12).

Like relativity theory, quantum theory has its roots in the recognition that the **observer** and the **system** he-or-she is observing are *inseparable*. Physics involves *measuring* things and you can't do that without disturbing what you are measuring: the disturbance can usually be made small for an 'everyday' object like a football; but not for something too small even to see – an electron would be knocked sideways if you looked for it using a beam of light!

So **Chapter 1** really starts 'at the very beginning' by asking what is meant by an ideal measurment. By sending a beam of electrons into a strong magnetic field you can deflect the beam 'up' or 'down': each electron carries a 'two-valued observable' which we can call **spin**; and sending the beam through the field sorts the electrons into two groups, those with up-spin and those with 'down-spin'. If you throw away the down-spin electrons you can be sure the other beam contains only up-spin particles: you've made a measurement! Just by thinking about this experiment, carried out by clever experimental physicists years before quantum mechanics really got started, it's possible to formulate many of the principles we're looking for. So quantum mechanical **operators**, along with **expectation values** and **eigenvalues** all fall out in the first few pages – with the help of only a few pictures that give a geometrical representation of what you're doing.

Chapter 2 shows how these basic ideas can be generalized by putting them into a mathematical 'framework' that you may remember from earlier books in the Series (starting with Book 2). Going from 2-valued observables to many-valued observables just means going from 2-dimensional vector space to n-dimensional vector space. States are still represented by vectors, but now there are n components instead of 2 and the n axes indicate states in which an observable has a definite value. An n-component vector has components that indicate the relative **probabilities** of finding the observable you're measuring with one definite value or another. In the beam-splitting experiemnt of Section 1.1 the state of any particle is represented by a 2-space vector with equal components, indicating a 50:50 probability of its coming out in the up-spin beam or the down-spin. With an n-valued observable, the n components will indicate the probabilities of finding the observable down-spin. With an n-valued observable, the n components will indicate the probabilities of finding the observable as a system with any of the n definite values open to it. And with every observable there is a corresponding operator with properties which can be 'echoed' in vector space, as operations that rotate the state vector. By the end of the chapter you'll know what

these properties are.

In Chapter 3 you catch up with history, arriving at 1925 when Schrödinger wrote his famous equation, representing the state vector as a wave function and the 'rotations' as differential operators – about which you know quite a lot from Books 3, 4 and 10. The square of the wave function $|\Psi(x, y, z)|^2$, for a one-particle system, gives the *probability* of finding the particle at a point with coordinates x, y, z. Then you learn more about expectation values and uncertainty; ending with Heisenberg's Uncertainty Principle. The chapter ends with a small box holding all you need to start using Wave Mechanics – Schrödinger's formulation of Quantum Mechanics.

And in **Chapter 4** you'll find simple examples of how to calculate the states of definite energy (the 'allowed' or **quantized**) states of a single particle moving in a given potential field. These are **eigenstates** of the **Hamiltonian operator** and if you remember a bit of calculus from Book 3 you'll be able to get the solutions easily. By contrast, there's one example of how the job can be done **symbolically**, using only the symbols representing the operators and their properties – without even a simple differential equation, though you'll need to think hard! In another example you find the lowest energy state of the electron in an atom of Hydrogen. The chapter ends with a review of general properties of eigenstates.

So far, the "states of definite energy" have all been **stationary states**, not changing as time passes. But real states are not like that – they must depend on the time t in some way. **Chapter 5** takes the next step by finding how the wave function can develop in time, which leads up to "Schrödinger's equation *including the time*". It first faces the great mystery of **wave-particle duality** – how can a particle like an electron behave sometimes like a 'real' *point* particle, but other times like a 'spread-out' wave? You'll find how the mystery was cleared up by inventing **wave packets**, strongly localized around the point where you expect to find the particle, but travelling through space – just like a classical particle. In fact, the time-dependent wave equation allows for such possibilities and gives us the 'bridge' between classical physics and quantum physics, together with a **Correspondence Principle** linking classical and quantum concepts. As the mass of the particle increases, its behaviour becomes more and more 'classical' – you even find expressions for force and acceleration of a *wave packet*, and discover they are related in a way that reminds you of Newton's second law!

By the end of the chapter you're ready to start work on a real system – the **Hydrogen atom** – and to see how Bohr's semi-classical treatment passes over into quantum mechanics.

Chapter 6 begins a more detailed study of Hydrogen as the simplest possible example of a **central field system**, containing one electron moving in the field provided by a heavy nucleus. From the last chapter, we know what to expect: that there will be eigenstates in which the energy of the electron can have definite quantized eigenvalues; and that, in any of these, other observables with associated operators that commute with H may simultaneously take definite values. The most completely defined electronic state you can find is one in which the energy E, the square of the angular momentum L^2 and one of its components e.g. L_z are all simultaneously known. This example is important because it provides a basis for classifying (and, in some approximation, even calculating the electronic structures of all the atoms in the Periodic Table (see Book 5). This is now all within your reach. The great importance of central field systems depends on their **spherical symmetry**: however the system may be rotated, around the central nucleus, it would look the same!

Having established the importance of symmetry, **Chapter 7** begins to build up the mathematical machinery for making use of it. The set of symmetry *operations*, such as rotations in space, that leave a physical system looking the same, form a **group**. And the effect of the operations can be described in terms of what they do to a set of **basis vectors** associated with the system. (You've seen many examples in earlier chapters.) Here you learn more about the **matrix representations** of a group. You use them first for systems like atoms, where the infinite set of rotations that leave the Hamiltonian unchanged form a **continuous group**. Then you need them again in dealing with molecules, where the symmetry operations generally form finite **point groups** (which leave just one point in the system fixed). Many properties of a sytem depend on its symmetry, in particular the **degeneracies** among its energy levels and the way it may be removed, or 'broken', when the symmetry is disturbed.

The next two chapters go much deeper, leading you to the end of the first phase in the development of the basic principles of quantum mechanics. Most of the spectacular applications of the theory to many-particle systems can be understood on the basis of what you've learnt so far; and this may be a good point for taking a break before looking at some of these applications (in Book 12). Just take a look at Chapters 8 and 9, to see where they're going, and come back to them when, or if, you need them.

Chapter 8 shows you that a study of the 3-dimensional rotation group – even without any 'high-brow' mathematics – can lead you to a deeper understanding of the properties of angular momentum than you got by solving eigenvalue equations in their Schrödinger forms. Symmetry of the central field is enough to determine not only the degeneracies among atomic energy levels but even the spherical harmonic forms of the atomic orbitals – without even setting up and solving a single partial differential equation!

Chapter 9 deals with the final unification of the formulations of quantum mechanics given by Born, Jordan, Heisenberg, Schrödinger and many others, achieved largely by Dirac about 80 years ago. The remaining parts of the theory relate to the proper inclusion of the **electromagnetic field** – which provides the interaction between the particles, but has not been included in Book 11. Quantization of the field takes you to the 'frontiers' of present day physics; but fortunately is not essential in most of the applications of the Quantum Mechanics of particles.

CONTENTS

Cha	apter 1 Observation and measurement	1
1.1	Starting from nothing	1
1.2	Putting it into pictures	4
1.3	Operators and expectation values	5
1.4	Properties of the spin operators	8
Cha	apter 2 How to generalize the basic ideas	13
2.1	From 2 dimensions to n dimensions	13
2.2	Some other spin systems	15
2.3	Vector spaces with a Hermitian metric	19
2.4	What can you do with all that mathematics?	21
Cha	apter 3 And now to Schrödinger!	24
3.1	Five years on –	24
3.2	Pictorial representation: the wave function	26
3.3	The definition of uncertainty	29
3.4	Heisenberg's Uncertainty Principle	30
3.5	Schrödinger's Wave Mechanics	33
3.6	Going into 3-space!	35
Chapter 4 Energy eigenstates of a single particle		37
4.1	Setting up the equations	37
4.2	Particle in a box	38
4.3	Particle in a parabolic potential field	45
4.4	The 3-dimensional oscillator	50
4.5	The hydrogen atom ground state	51
4.6	Back to the free particle	54
4.7	Some general properties of eigenstates	56
4.8	But what happens to a state as time passes?	58
Chapter 5 The 'bridge' between 'classical' and 'quantum' physics		60
5.1	How can we bring in the time?	60
5.2	The Gaussian wave packet	64
5.3	Schrödinger's equation including the time	67
5.4	What makes a wave packet move?	69
5.5	Commutators and 'constants of the motion'. Angular momentum	74
Chapter 6 Hydrogen: the simplest atom of all		
6.1	An electron in the central field of a nucleus	79
6.2	The energy levels of a 1-electron atom	82

6.3	Eigenfunctions with angle dependence	87
Cha	apter 7 More on Symmetry	94
7.1	Operators and matrix representations	94
7.2	What makes the central field so special?	98
7.3	A first glimpse of the theory og groups	102
Cha	apter 8 The three-dimensional rotation group	106
8.1	Representations in spin-space	106
8.2	The standard representations	110
8.3	The infinitesimal operators	118
8.4	The connection with Physics	121
Chapter 9 Finishing touches – the Dirac formalism		125
9.1	The bra-ket notation	125
9.2	Heisenberg's 'matrix mechanics'	128
9.3	Generalizations. Observables with continuous values	130
9.4	The p-representation: momentum space	131

Chapter 1

Observation and measurement

1.1 Starting from nothing

When we *observe* something, such as a moving particle, and try to *measure* its position, say, or its speed, we meet several problems. The observer who makes the measurement must *interact* in some way with the particle and this may disturb the things he is trying to measure. For a **macroscopic** object, like a falling pebble, the disturbance can usually be neglected (we could shine a light on the pebble and observe its moving shadow), but you can't do that for a single atom – you can't even find one! And even if you could, a beam of light (which contains energy, as you know from Book 10) would knock it sideways: so to talk about measurements on a **microscopic** object you have to do everything in your imagination.

When you 'start from nothing' the first thing to do is to define what you're going to call an "ideal measurement":

An **ideal measurement** of any quantity is one which produces *no change* in the quantity being measured, even though it may produce changes in other quantities belonging to the system being observed.

This definition means that if the measuring experiment is immediately repeated there will be no change in the result from the first experiment.

We must also keep in mind that experiments are normally performed on an enormous number of systems, for example the many millions of electrons forming an electron beam (like the one that produces a spot of light on a television screen). We'll call the collection of all such particles, all observed under exactly the same conditions, an **assembly**; and the simplest observation we can make is *whether or not* something has a given value – "yes" or "no" is called a "two-valued observable".

An example of such an observable is found in an experiment made by Stern and Gerlach in 1921. They used a beam of silver atoms, produced by heating a silver wire and letting the evaporated atoms pass through a small hole, as in Figure 1 (below) – all in a vacuum of course, so that the beam wouldn't be 'scattered' by the molecules in the air. After

passing through the hole, the narrow beam of atoms passes between the poles of a powerful **electromagnet** (see Book 10, Chapter 3). In the Figure this is called an **analyser**

You'll remember from Book 4 that when a body is turning around its axis it is said to have **angular momentum**. Here the 'body' is a single atom, rotating around an axis through its nucleus; and the 'cloud' of electrons which surround the nucleus will provide a **circulating current** of negative charge. This produces a magnetic moment, proportional to the angular momentum, and each silver atom then behaves like a tiny magnet. We'll represent the 'spin angular momentum' and its associated magnetic moment by a small arrow – a **spin vector**. The length of the arrow indicates its *magnitude* while its direction is that of the axis of rotation.

The poles of the magnet in Fig.1 are specially shaped, so as to produce a very non-uniform field between them, and in this case the atoms that come through are pushed one way or the other according to their magnetic moment: they are 'deflected'. In Fig.1(a), the magnetic field is switched off: the atoms go straight on, with their spin vectors oriented at random. But Fig.1(b) shows what happens when the field is switched on:



Figure 1 The Stern-Gerlach experiment (schematic)

After passing through the analyser the atoms come out in two different beams; using the field direction as the z axis, the z-component of the spin vector (S_z) takes only two distinct values, λ_1 and λ_2 say. The 'up-spin' beam (with spin vectors pointing upwards) contains only atoms with $S_z = \lambda_1$, the down-spin beam only those with $S_z = \lambda_2$.

The spin component S_z is an observable which, as experiment shows, can take only two measurable values (λ_1 and λ_2). Before applying the field the spin vectors can point in *any* direction and the value of S_z is completely **uncertain**. But as soon as the field is applied,

the atoms can be found only in one **state** or the other: we'll use the symbol ψ (the Greek letter "psi") to denote a general state of the system (i.e. of an atom with spin), α for the state of an atom in the up-spin beam with $S_z = \lambda_1$, and β for an atom in the down-spin beam with $S_z = \lambda_2$.

The analyser in Fig.1 thus gives us a way of measuring the component of angular momentum of the spinning atom along the direction of the applied field: for example, if we take an atom from the top beam coming out, we can be sure it has the value λ_1 – like all the other atoms in the beam. We have made a measurement!

Another important result comes out of this experiment: the up-spin and down-spin beams contain the same number of atoms, so if you think of one particular atom there's a fifty-fifty chance that it goes through the analyser and comes out with spin up or spin down. You can't say definitely what value you will observe, only that each of the possible results will be found with a **probability** $\frac{1}{2}$. If you change the experiment a bit the numbers may change; but in general the results you get from experiments on microscopic systems are going to be **statistical** – each *possible* result will have a certain *probability* of being found, and you can't say more than that. (If you need reminding about probabilities look back at Book 5 Section 4.1)

To confirm that this is an 'ideal' measurement, as defined in the last Section, we can do exactly the same experiment on the atoms in the 'up-spin' beam. We let them go into a second analyser, just like the first one. In Figure 2 the lower beam, of atoms with spin λ_2 is blocked by a 'stop', but the atoms of the upper beam pass into the second analyser and *nothing happens to them* – they come out with the same value $S_z = \lambda_1$ as they had when they went in.



Figure 2. Confirmation of $S_z = \lambda_1$ by 2nd Analyser

To summarize: Fig.2 illustrates an ideal measurement. The atoms coming out in the upper beam from the left-hand analyser (which is part of the Stern-Gerlach experiment), are known to have spin component $S_z = \lambda_1$: it has just been 'measured'. And the Figure shows that this measurement is confirmed by the second analyser, which makes no change. The two 'allowed' values of the observable S_z , namely λ_1 and λ_2 , are called its **eigenvalues**; and the corresponding states, which we'll denote by α and β , are called **eigenstates**. ("eigen" is the German word meaning "self" or "its own", so these are the observable's own special states.)

Now let's change the setup in Fig.1 by turning the whole experiment through 90° around the y-axis, the direction in which the silver atoms were originally going, with the field switched off. In a 'right-handed' system of coordinates, the z-axis of the analyser (and therefore the magnetic field) will now point in the (original) x-direction. The result is indicated in Figure 3, where the spin vectors in the 'up-spin' beam point directly towards you and are shown as bold dots instead of arrows.



Figure 3. Effect of rotating the Stern-Gerlach experiment

The Figure also shows what happens to the beam that goes into the *second* analyser, which remains fixed with its field still in the original z-direction. This analyser, fixed in the laboratory, splits the beam again into parts with $S_z = \lambda_1$ ('up-spin') and $S_z = \lambda_2$ ('down-spin') – ignoring the previous result, which measured a *different* component, S_x .

1.2 Putting it into pictures

Let's review what we've found so far. Observation of a spin component S_z , in the direction of an applied magnetic field (taken as the z-axis), always gives one of two values, λ_1 , or λ_2 . The *state* of a system with any given S_z will be denoted by ψ and will be represented geometrically by an arrow of unit length. The particular states with $S_z = \lambda_1$ and $S_z = \lambda_2$ will then be denoted by $\psi_1 = \alpha$, and $\psi_2 = \beta$, respectively. We can put all this into a picture, as in Figure 4.



The eigenstates α and β are represented as **orthogonal** unit vectors (pointing at rightangles to each other) and *any* other state vector ψ is expressed as a linear combination

$$\psi = c_1 \alpha + c_2 \beta. \tag{1.1}$$

This is a sensible convention because probabilities are essentially positive quantities with values between 0 and 1. By choosing $p_1 = c_1^2$, $p_2 = c_2^2$, any state vector ψ , no matter which way it points, will have unit length,

$$|\psi|^2 = c_1^2 + c_2^2 = 1. \tag{1.2}$$

(Remember that the 'length' or 'modulus' of a vector ψ is usually denoted by $|\psi|$ and its square is the sum of the squares of its components – if you don't you'd better go back to Book 2!) So the spin state of a particle (more precisely a 'spin- $\frac{1}{2}$ particle', like the ones we're talking about) can be represented by a unit vector in a two-dimensional 'spin-space'; every way the vector points indicates a possible spin state.

You may also remember from Book 2 that two vectors have a scalar product defined as the product of their lengths times the cosine of the angle between them, indicated by putting a dot between them. For the unit vectors α and β (inclined at 90°) the cosine is zero and therefore

$$\alpha \cdot \alpha = \beta \cdot \beta = 1, \quad \alpha \cdot \beta = \beta \cdot \alpha = 0. \tag{1.3}$$

It follows that the vector ψ can be rotated in any direction, without changing its length, because the square of its length is (say why!)

$$\psi \cdot \psi = (c_1 \alpha + c_2 \beta) \cdot (c_1 \alpha + c_2 \beta)$$

= $c_1^2 \alpha \cdot \alpha + c_2^2 \beta \cdot \beta + c_1 c_2 \alpha \cdot \beta + c_2 c_1 \beta \cdot \alpha$
= $c_1^2 + c_2^2 = 1,$ (1.4)

according to (1.2). Notice that in rearranging expressions that contain both vectors and numbers you're only using the 'associative' and 'distributive' laws, which you first met in Book 1, and the fact that numbers and vectors 'commute' – it doesn't matter which way round you do the multiplications. And the vector components c_1, c_2 , which appear in (1.4) as 'expansion coefficients', can be either positive or negative – the state vector ψ can be rotated any way you please and nothing has to be changed.

In other books in the Series you've used symbols to stand for **operations** such as rotations, which change the 'state' of some kind of object into a new state (e.g. in Book 1, Chapter 7); or which change a function ψ of some variable x into a new function ψ' (e.g. the differential operator D, which leads from $\psi(x)$ to $\psi'(x)$ by following certain rules). Differential operators were used throughout Book 3 and again, in more advanced ways, in Book 10. So by now you're ready to start using operators in still other ways – for they form the 'backbone' of **quantum theory**.

1.3 Operators and expectation values

In the Stern-Gerlach experiment, particles in a spin state ψ (represented in Fig.4 by a vector) were sent into states with vectors α or β when they passed through the analyser. With the *operation* of selecting up-spin particles from those in an unknown state ψ we can associate a **projection operator** P₁ which picks out the α part of ψ given in (1.4).

And in a similar way, for selecting down-spin particles, we can define a second operator P_2 which will pick out the β part of ψ . Thus

$$\mathsf{P}_1 \, \psi = c_1 \alpha, \quad \mathsf{P}_2 \, \psi = c_2 \beta. \tag{1.5}$$

If you go back to Book 1 (Section 6.2) you'll see we played the same game with a market containing animals of various kinds (Asses, Bullocks, Cows, etc.). The state of the market, before we observed it, was represented by a symbol S, describing its composition (e.g. S = 6a + 4b + 8c + ..., meaning 6 asses, 4 bullocks, 8 cows, etc.). And the operations of selecting animals of one kind or another were represented by A (for asses), B (for bullocks), and so on. To observe the 'bullock component' you operate on the market with B and find BS = 4b, which tells you there are four bullocks. Evidently the mathematical 'machinery' we're using is very general; it applies whenever things are being sorted into categories. The number of categories and their nature are not important – it's only the *interpretation* of the symbols that counts.

Let's get back to Physics! Here we used the convention that the coefficients c_1, c_2 in (1.4) would be related to the probabilities p_1, p_2 of a particle being observed to have spin up or spin down: $p_1 = c_1^2$, $p_2 = c_2^2$. To get the squares of the expansion coefficients directly from (1.5) you'll need a second ψ . By taking a scalar product with ψ , from the left, the two equations in (1.5) will give

$$\psi \cdot \mathsf{P}_{1}\psi = (c_{1}\alpha + c_{2}\beta) \cdot (c_{1}\alpha) = c_{1}^{2} = p_{1},$$

$$\psi \cdot \mathsf{P}_{2}\psi = (c_{1}\alpha + c_{2}\beta) \cdot (c_{2}\beta) = c_{2}^{2} = p_{2},$$
(1.6)

from the results given in (1.3).

Now if you make a number of observations of S_z , sometimes finding the value λ_1 (up spin) and sometimes finding λ_2 (down spin) you can't say which it will be in any single observation; but you can say what value to expect. The **expectation value** of S_z is defined as the average value obtained in a large number N of measurements. If n_1 particles come out of the analyser in the upper beam and n_2 in the lower beam, the average value of S_z for the emerging particles will be

$$\langle S_z \rangle = (n_1 \lambda_1 + n_2 \lambda_2)/N_z$$

where putting a quantity between pointed brackets means taking its average value (as used in Book 5 and elsewhere). Thus, since $n_1/N = p_1$ is interpreted as the probability of finding any particle, taken at random, with spin $S_z = \lambda_1$, and similarly for n_2/N , the average may be expressed as

$$\langle S_z \rangle = p_1 \lambda_1 + p_2 \lambda_2. \tag{1.7}$$

The last step is to use the expressions for p_1 and p_2 , given in (1.6), to write this result in terms of the state vector ψ and a single operator S_z associated with the observable quantity S_z . The result (check it!) is

$$\langle S_z \rangle = \psi \cdot \mathsf{S}_z \psi, \tag{1.8}$$

where

$$\mathsf{S}_z = \lambda_1 \mathsf{P}_1 + \lambda_2 \mathsf{P}_2. \tag{1.9}$$

The possibility of associating an operator (X, say) with any observable quantity X and using it to get the expectation value $\langle X \rangle$, in any state with state vector ψ , is so important that we'll put it in a 'box':

> The **expectation value** of any observable quantity X, in state ψ , is given by $\langle X \rangle = \psi \cdot X \psi$ where the operator X associated with the observable is $X = X_1 P_1 + X_2 P_2 + \dots$.

> > (1.10)

The formula (1.9) is a specially simple example of this result, where X is the spin operator S_z , X_1 and X_2 are the **eigenvalues**, namely λ_1 and λ_2 , and P_1, P_2 are the projection operators onto the **eigenvectors**, α , β . But (1.10) is completely general: it doesn't matter *how many* special states, with their eigenvalues and eigenvectors, there may be; if there are more than three you just have to use a *many*-dimensional space; and you've done that many times already, starting from Chapter 7 of Book 2, where instead of the two eigenvectors α, β you had *n* basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$.

Example 1.1 A property of the eigenvectors.

The spin operator S_z normally produces from any state vector ψ a new vector ψ' pointing in a different direction. Make a test, using the operator (1.9) and the vector $\psi = \frac{1}{2}\alpha + \frac{1}{2}\beta$.

Since $P_1 \alpha = \alpha$ (the projection of α on itself) and $P_1 \beta = 0$ (projection of β on the α -axis), the first term in $S_z \psi$ will be $\frac{1}{2}\lambda_1 \alpha$. A similar argument leads to a second term $\frac{1}{2}\lambda_2 \beta$ and therefore $\psi' = S_z \psi = \frac{1}{2}(\lambda_1 \alpha + \lambda_2 \beta)$. But the spin components for the α and β states are $\lambda_1 = +\frac{1}{2}$ and $\lambda_2 = -\frac{1}{2}$, so

$$\psi' = \mathsf{S}_z \psi = \frac{1}{4} (\alpha - \beta).$$

This is certainly *different* from the original ψ : it is in fact orthogonal to it, as you can verify by noting that the scalar product of the factors $\alpha + \beta$ and $\alpha - \beta$ is zero.

There are only two cases in which ψ and ψ' lie on the same axis: they are obtained by taking one of the coefficients c_1, c_2 to be 1 and the other to be 0, leading to

$$\psi_1 = \alpha, \qquad \psi_2 = \beta.$$

These are the **eigenvectors** of the operator S_z , with eigenvalues λ_1 and λ_2 , respectively. They are solutions of the **eigenvalue equation** $S_z \psi = \lambda \psi$.

With all the quantities we observe and measure in physics we can associate **operators**: any state of a system is indicated by a **state vector** ψ and in this state the observable X, with corresponding operator X, will have an **expectation value** $\langle X \rangle = \psi \cdot X \psi$.

In certain particular states, with **eigenvectors** (ψ_1, ψ_2, \dots) , the quantity X will have a *definite* value X_1, X_2, \dots) called an **eigenvalue**. The eigenvalues and eigenvectors are solutions of the **eigenvalue equation** $X \psi = X \psi$. (Remember that X stands for the physical quantity being measured, a *number* of units, while X, in special type, is the *operator* associated with it.)

In this Chapter, we're just putting down the mathematical foundations of **quantum** mechanics – the discipline that replaces Newton's 'classical' mechanics when we deal with really small systems like electrons and protons. Although we started by talking about silver atoms in the Stern-Gerlach experiment, the simplest 'spin- $\frac{1}{2}$ particle' of all is a single electron; and experiment shows that the eigenvalues λ_1, λ_2 are in reality $+\frac{1}{2}\hbar, -\frac{1}{2}\hbar$ where \hbar is a quantity with physical dimensions of angular momentum (MLT⁻¹, or energy \times time) and numerical value 1.0545×10^{-34} J s.

We can't go any further into quantum mechanics until we know something about the operators themselves, so that will be our next task.

1.4 Properties of the spin operators

As the angular momentum is a 3-component quantity, we expect it will have other components S_x, S_y as well as S_z ; but experiment has shown that only *one* of them can be measured and we've called that the z-component: any attempt to measure S_x or S_y will then spoil the result of the first measurement – we may find a value of S_x , say, but will then lose the values of S_z . We continue to talk about *ideal* measurements in the sense of Section 1.1.

All we can say so far is that (using units of \hbar everywhere for angular momentum)

$$S_z \alpha = \frac{1}{2} \alpha, \qquad S_z \beta = -\frac{1}{2} \beta,$$
 (1.11)

for the spin states defined in earlier Sections. This simply says that α and β are particular eigenvectors of S_z . which multiplies α by the eigenvalue $\frac{1}{2}$ and β by $-\frac{1}{2}$. A second application of the operator gives $S_z S_z \alpha = \frac{1}{4} \alpha$ and similarly for β . So the operator S_z^2 multiplies both eigenvectors, and any combination (), by $\frac{1}{4}$:

$$\mathsf{S}_z^2 \psi = \frac{1}{4} \psi. \tag{1.12}$$

In other words, S_z^2 is equivalent to $\frac{1}{4}I$, where I is the 'identity operator' which leaves any vector ψ of the spin space unchanged.

To find some other properties of the spin operators it's enough to note that ordinary physical space is **isotropic**: what's good for one direction is good for any other direction.

Therefore the squares of all three vector components will have associated operators just like (1.12): and the operator associated with the squared magnitude of the spin S will be

$$S^{2} = S_{x}^{2} + S_{y}^{2} + S_{z}^{2} = \frac{1}{4}I.$$
 (1.13)

And when this operator works on *any* spin state vector ψ the result will be $S^2\psi = \frac{3}{4}\psi$, which can be written in the form

$$S^{2}\psi = \frac{3}{4}\psi = \mathbf{s}(\mathbf{s}+\mathbf{1})\psi, \qquad (1.14)$$

where $s = \frac{1}{2}$ is the maximum allowed value of any spin component. The value s(s + 1), where in classical mechanics you would have expected s^2 , is characteristic of quantum mechanics and is found for vectors of all kinds.

To find other properties of the spin operators we can again use the fact that physical space is isotropic: we expect that the component of spin along any other direction (let's call it the z'-axis) will have the same properties as S_z .

Now the component S'_z of a vector along a *rotated* z-axis, pointing in the direction with **direction cosines** (see Book 2) l, m, n, will be (In case you've forgotten, l, m, n are the cosines of the angle the new z-axis makes with the old ones; and whichever way it points $l^2 + m^2 + n^2 = 1$)

$$S_{z}' = lS_{x} + mS_{y} + nS_{z}.$$
(1.15)

What we're supposing now is that the associated operator S'_z , when squared. will be equivalent to $\frac{1}{4}I$, just as it was for s_z . But if you multiply out you'll find

$$\begin{aligned} \mathsf{S}_{z}^{\,\prime\,2} &= l^{2}\mathsf{S}_{x}^{\,2} + m^{2}\mathsf{S}_{y}^{\,2} + n^{2}\mathsf{S}_{z}^{\,2} + lm(\mathsf{S}_{x}\mathsf{S}_{y} + \mathsf{S}_{y}\mathsf{S}_{x}) + \dots \\ &= \frac{3}{4}\mathsf{I} + lm(\mathsf{S}_{x}\mathsf{S}_{y} + \mathsf{S}_{y}\mathsf{S}_{x}) + \dots \end{aligned}$$

and if this is to be equivalent to $\frac{3}{4}I$, for any values of the numbers l, m, n, then we must have

$$(\mathsf{S}_x\mathsf{S}_y + \mathsf{S}_y\mathsf{S}_x) = (\mathsf{S}_y\mathsf{S}_z + \mathsf{S}_z\mathsf{S}_y) = (\mathsf{S}_z\mathsf{S}_x + \mathsf{S}_x\mathsf{S}_z) = 0.$$
(1.16)

Thus, if we apply $S_x S_y$ to a ψ representing any spin state, and then $S_y S_x$, and finally add the results, nothing will be left. The spin operators for different components are said to **anticommute**. (They would *commute* if you had a *minus* sign between the two products.)

It remains only to ask what the operators S_x, S_y, S_z will do to the **basis vectors** α and β of spin space. The following Example shows how to get the answer.

Example 1.2 Effects of the spin operators

Using the last anticommutator in (1.16), we can write

$$S_z(S_x\alpha) = -S_xS_z\alpha = -\frac{1}{2}(S_x\alpha)$$

We start from one spin operator, S_z , and the definitions $S_z \alpha = \frac{1}{2} \alpha$, $S_z \beta = -\frac{1}{2} \beta$ – which come straight from the Stern-Gerlach experiment.

as follows from $S_z \alpha = \frac{1}{2} \alpha$. This means $S_x \alpha$ is an eigenvector of S_z with eigenvalue $-\frac{1}{2}$; but that is the property defining β – or any multiple of it, $k\beta$ say. In the same way, working on β instead of α , you find $(S_x\beta)$ is an eigenvector of S_z with eigenvalue $\frac{1}{2}$; so you can write

$$\mathsf{S}_x \alpha = k\beta, \qquad \mathsf{S}_x \beta = k'\alpha$$

Now we know that $S_x^2 \alpha = \frac{1}{4} I \alpha = \frac{1}{4} \alpha$ and have just found that $S_x \alpha = k\beta$ and $S_x \beta = k'\alpha$. So we can say

$$\mathsf{S}_{x}^{2}\alpha = k\mathsf{S}_{x}\beta = kk'\alpha$$

and, by comparison, it follows that $kk' = \frac{1}{4}$. A similar argument, but starting from the second anticommutator in (

Choosing the numbers k and k' is more difficult: in general they contain a 'phase factor' and can be written $k = \frac{1}{2}e^{i\theta}$, $k' = \frac{1}{2}e^{-i\theta}$, where $i^2 = -1$ and θ is a real number. The usual choice is $\theta = 0$ in defining the S_x operator, giving $S_x \alpha = \frac{1}{2}\beta$ $S_x \beta = \frac{1}{2}\alpha$; and $\theta = 1$ in defining S_y , which leads to $S_y \alpha = \frac{1}{2}i\beta$. $S_y \beta = -\frac{1}{2}i\alpha$. (The reasons for choosing the phases in this way will become clear later.)

In summary, the conclusions from Example 1.2 may be collected as:

$$\begin{array}{ll} \mathsf{S}_{x}\alpha = \frac{1}{2}\beta, & \mathsf{S}_{y}\alpha = \frac{1}{2}i\beta, & \mathsf{S}_{z}\alpha = \frac{1}{2}\alpha\\ \mathsf{S}_{x}\beta = \frac{1}{2}\alpha, & \mathsf{S}_{y}\beta = -\frac{1}{2}i\alpha, & \mathsf{S}_{z}\beta = -\frac{1}{2}\beta \end{array}$$

(1.17)

As a third Example let's find *commutation* relations for the spin operators: these must apply to the operators themselves, without reference to any particular spin states.

Example 1.3 The commutation relations for spin

Let's look at the first anticommutator in (1.16) and ask what it does when it acts on each of the eigenvectors α and β . Take it step by step, using the results we've just obtained:

$$(\mathsf{S}_x\mathsf{S}_y - \mathsf{S}_y\mathsf{S}_x)\alpha = \mathsf{S}_x(\frac{1}{2}i\beta - \mathsf{S}_y(\frac{1}{2}\beta))$$
$$= \frac{1}{2}i(\frac{1}{2}\alpha) - \frac{1}{2}(-\frac{1}{2}i\alpha) = \frac{1}{2}i\alpha$$
$$(\mathsf{S}_x\mathsf{S}_y - \mathsf{S}_y\mathsf{S}_x)\beta = \mathsf{S}_x(-\frac{1}{2}i\alpha - \mathsf{S}_y(\frac{1}{2}\alpha))$$
$$= -\frac{1}{2}i(\frac{1}{2}\beta) - \frac{1}{2}(\frac{1}{2}i\beta) = -\frac{1}{2}i\beta.$$

So the operator on the left, acting on α , simply multiplies it by $\frac{1}{2}i$; whereas, acting on β , it multiplies by $-\frac{1}{2}i$. But the first equation of the pair therefore says that

$$(\mathsf{S}_x\mathsf{S}_y - \mathsf{S}_y\mathsf{S}_x)\alpha = i\mathsf{S}_z\alpha,$$

while the second equation says that

$$(\mathsf{S}_x\mathsf{S}_y - \mathsf{S}_y\mathsf{S}_x)\beta = i\mathsf{S}_z\beta.$$

In other words the anticommutator has exactly the same effect as iS_z , when it works on either of the two basic eigenvectors, α and β , and therefore on any combination $\psi = c_1 \alpha = c_2 \beta$. Two other identities can be found (try it!), using the other anticommutators in (1.16).

The conclusion from Example 1.3 is that

$$\begin{split} \mathsf{S}_x\mathsf{S}_y - \mathsf{S}_y\mathsf{S}_x &= i\mathsf{S}_z,\\ \mathsf{S}_y\mathsf{S}_z - \mathsf{S}_z\mathsf{S}_y &= i\mathsf{S}_x,\\ \mathsf{S}_z\mathsf{S}_x - \mathsf{S}_x\mathsf{S}_z &= i\mathsf{S}_y. \end{split}$$

(1.18)

These are the **commutation relations for spin**. Such relations are easy to remember, as noted before, because the subscripts follow the cyclic order $xyz \rightarrow yzx \rightarrow zxy$ in going from one equation to the next. The first equation says that the S_xS_y -anticommutator gives you S_z and to get the next you just change the labels x,y,z to y,z,x: the yz anticommutator gives you the x-component. (The other choice of phase factor in Example 1.2, putting $\theta = -1$ instead of +1, would spoil the nice symmetry of the equations in (1.18). If you work through it you find that similar equations come out, but with $-S_z$ in place of S_z : reversing the sign of the z-component corresponds to changing from a right-handed to a left-handed coordinate system. But our convention has been always to use a right-handed system – so if we change it in just one place we'll be in trouble!)

To end this Chapter let's use the properties in (1.18) to get two new operators which let you pass easily from one spin eigenvector to another, with a different value of the z-component. These are often called "step-up" and "step-down" operators, since they increase or decrease the value of S_z by one unit. They are very useful and important because they apply to *all* kinds of angular momentum – not only to that of a single spin- $\frac{1}{2}$ particle:

Example 1.4 Step-up and step-down operators

To do that we note that, from (1.18), the first term in $S_z S^-$ can be replaced by $S_z S_x = S_x S_z + iS_y$; and the second term (again using (1.18)) by $-iS_z S_y = -S_x - iS_y S_z$ On adding the two terms together you find

$$\mathsf{S}_z(\mathsf{S}_x - i\mathsf{S}_y) = (\mathsf{S}_x - i\mathsf{S}_y)\mathsf{S}_z - (\mathsf{S}_x - i\mathsf{S}_y)$$

As in (1.17), we'll go on using S_x, S_y, S_z for angular momentum operators, even in the general case: what matters is only that they have the same commutation properties and that we can start from a state of given S_z – an eigenvector of the operator S_z , with an eigenvalue we can call m.

From the operators for the x-, y- and z-components we'll form two new operators $S^+ = S_x + iS_y$ and $S^- = S_x - iS_y$, called "step-up" and "step-down" operators, respectively.

To see what S^- , for example, does to a state vector ψ_m , we must 'test' $S^-\psi_m$ by applying S_z (which simply multiplied the original ψ_m by the eigenvalue m). So we need an expression for $S_zS^- = S_zS_x - iS_zS_y$; and we want to get the S_z over to the right, closest to the ψ_m that it's going to work on.

and, since $(S_x - iS_y) = S^-$ (the step-down operator defined above) this may be re-written as

$$S_z S^- = S^- S_z - S^-$$

Now suppose there is one eigenstate in which the spin z-component has its largest value of s units and a state vector ψ_m with m = s. And let the operator just defined work on ψ_m . The result will be

$$\mathsf{S}_z\mathsf{S}^-\psi_m=\mathsf{S}^-\mathsf{S}_z\psi_m-\mathsf{S}^-\psi_m$$

and, since S_z just multiplies ψ_m by the eigenvalue m, this may be re-written as

$$\mathsf{S}_{z}(\mathsf{S}^{-}\psi_{m}) = m(\mathsf{S}^{-}\psi_{m}) - (\mathsf{S}^{-}\psi_{m}) = (m-1)(\mathsf{S}^{-}\psi_{m}).$$

What does this tell us? It shows that when the 'step-down' operator works on an eigenvector of S_z , with any eigenvalue m (not only m = s), it changes it into an eigenvector with m replaced by m - 1. This of course is the reason for the name. A similar argument (try to work through it) shows that the operator S^+ leads from an eigenvector with eigenvalue m to one with m 'stepped up' to m + 1.

In summary, the "step-up" and "step-down" operators,

$$\mathsf{S}^+ = \mathsf{S}_x + i\mathsf{S}_y, \qquad \mathsf{S}^- = \mathsf{S}_x - i\mathsf{S}_y, \tag{1.19}$$

have the properties

$$S^+\psi_m = \psi_{m+1}, \qquad S^-\psi_m = \psi_{m-1}.$$
 (1.20)

Note that the eigenstates should be labelled by *two* numbers, which we've denoted by s and m (s being the maximum value that m can take, here being fixed and not shown. These are the **quantum numbers** for spin and determine the allowed values of S^2 and S_z in any spin state.

The 'spectrum' of *m* values goes from -s up to +s and is *discrete*: for a single electron $s = \frac{1}{2}$ and $m = \pm \frac{1}{2}$, but many other systems exist with a larger spin, of *S* units. In general, *S* (denoted by a capital letter) may only take values which are equal to **half an integer**. For example $S = 1(=\frac{1}{2}2)$, $S = \frac{3}{2}(=\frac{1}{2}3)$, $S = 2(=\frac{1}{2}4)$,

More about that later: for now it's enough to know that for given S there are 2S + 1 quantum states, with M values $-S, -S + 1, \ldots + S$ – all very similar to the spin- $\frac{1}{2}$ case, but with quantum numbers (s, m) replaced by S, M. Note also that there's always a 'top' state, with M = S, and a 'bottom' state, with M = -S. In Chapter 2 we'll discover where these mysterious rules come from.

Chapter 2

How to generalize the basic ideas

2.1 From 2 dimensions to *n* dimensions

The Stern-Gerlach experiment has given us a lot to think about, but what else can we get from it? So far we've been talking about measuring a physical quantity that can take only two possible values, like 'up' $(S_z = +\frac{1}{2})$ or 'down' $(S_z = -\frac{1}{2})$. But the quantities measured in Physics usually have an infinite number of possible values; and these values may be separate from each other (i.e. *discrete*) or may be infinitely close together (i.e. *continuous*). So we have to generalize everything to include such cases, starting from what we've discovered already.

We'll follow the same steps we took in talking about the Stern-Gerlach experiment – but just changing the notation a bit so it will apply to other kinds of measurement and will, at the same time, remind you of the things you've done already in other books.

If we use X for the thing we're measuring (instead of S_z) and use $\psi = c_1\mathbf{e}_1 + c_2\mathbf{e}_2$ for the state vector describing our results (instead of $\psi = c_1\alpha + c_2\beta$), then we can easily go to the case where X can be found with possible values $X_1, X_2, X_3, \dots X_i, \dots X_n,$ the number n being as big as we wish. Note that we are now going to talk about ndimensional vector space, like the the ones you first met long ago in Section 7.1 of Book 2. We've kept $\mathbf{e}_1, \mathbf{e}_2, \dots \mathbf{e}_n$ for the basis vectors to remind you that in Section 1.1 $\alpha = \mathbf{e}_1$ and $\beta = \mathbf{e}_2$ were **eigenvectors** (e-vectors).

An n-dimensional 'state vector' will be

$$\psi = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3 + \dots = \sum_{i=1}^n c_i \mathbf{e}_i,$$
 (2.1)

where the coefficients c_i are, as usual, vector components. (The same word – "component" – is used for both purposes: as the numerical *coefficient* c_i of a basis vector \mathbf{e}_i , or as the vector $c_i \mathbf{e}_i$, which is one part of the whole vector ψ . And although i, j, k are often used as general labels you mustn't get confused when i is used for the complex number $i = \sqrt{-1}$. With each basis vector \mathbf{e}_i is associated a **projection operator** P_i , which 'projects out' the part of any vector ψ along the 'direction' of \mathbf{e}_i . (Look back at Fig.4 to see what this means.) Thus

$$\mathsf{P}_i\psi = c_i\mathsf{e}_i.\tag{2.2}$$

Remember also the interpretation of the components c_1, c_2 in Section 1.1: $c_1^2 = p_1$ was the fractional number of times an atom was found to come out of the Analyser (Fig.1) with $S_z = \frac{1}{2}$, in a large number of observations. It measures the **probability** of finding that value if you observe a single atom, taken 'at random', undergoing the same experiment. In Fig.1, the fractional number p_1 was $\frac{1}{2}$; and the same conclusion holds for the other possible result, so $p_1 = p_2 = \frac{1}{2}$ and $p_1 + p_2 = 1$. In Fig.4, the components of the state vector ψ were chosen to match this result: $c_1^2 + c_2^2 = p_1 + p_2 = 1$ and the vector is then said to be **normalized to unity**. In the picture (Fig.4), ψ is represented as a vector of unit length.

The same interpretation can be used in the *n*-dimensional case, where the quantity X can take *n* possible values. Measurement means doing an experiment to find which possible value (X_i, say) the observable X has: you must do it a large number of times (N) (without changing anything between experiments), then count the number of times (n_i) you record that value (X_i) and finally put $p_i = n_i/N$. That's how you get an average or 'expectation' value of the quantity you're measuring: it's simply

$$\langle X \rangle = \left(\frac{n_1 X_1}{N} + \frac{n_2 X_2}{N} + \dots\right) = \sum_{i=1}^n p_i X_i.$$
 (2.3)

To get a geometrical picture of this, like the one in Fig.4, you only have to define the components c_i of the state vector ψ , choosing them so that $p_1 = c_1^2, p_2 = c_2^2$, etc., and the vector will be normalized to unity because

$$\frac{n_1}{N} + \frac{n_2}{N} + \dots \frac{n_n}{N} = \frac{n_1 + n_2 + \dots + n_n}{N} = \frac{N}{N} = 1.$$

(Of course you can't *draw* the picture, because your paper only has *two* dimensions. But by now that shouldn't worry you. If the mathematics is OK you can use it!)

The next step is to associate an **operator** X with the observable X. In Chapter 1 we defined a spin operator S_z in terms of the eigenvectors α and β and the projection operators P_1 and P_2 , which picked out the α - and β -components of the state vector ψ . And then we found that $\langle S_z \rangle$ could be obtained from ψ in the form (1.8) – namely $\langle S_z \rangle = \psi \cdot S_z \psi$. We can do the same in n-dimensional space; but first need to say a word about the scalar **product**, indicated by the dot between two vectors. This was first defined in Section 5.4 of Book 1, for a space with basis vectors which are mutually perpendicular (**orthogonal**) and of unit length: such spaces are called "Cartesian", after the French mathematician Descartes. We still use them most of the time, owing to their simplicity. (For other examples see Section 7.1 of Book 2.). The following part in small type is just to remind you of things you should already know!

The simplest scalar product is the square of the length of a vector, which in two dimensions is $|\psi|^2 = \psi \cdot \psi = c_1^2 + c_2^2$. In *n* dimensions the corresponding form is the sum of squares of all components:

$$|\psi|^2 = c_1^2 + c_2^2 + c_3^2 + \dots = \sum_{i=1}^n c_i \mathbf{e}_i,$$

But for two *different* vectors, ψ and ψ' , the scalar product becomes

$$\psi \cdot \psi' = c_1 c'_1 + c_2 c'_2 + c_3 c'_3 + \dots = \sum_{i=1}^n c_i c'_i,$$

where the 'primed' components are those of ψ' , the second vector. $\psi \cdot \psi'$ is then the product of the lengths of ψ and ψ' times the cosine of the angle between the vectors.

Perhaps you remember (Section 7.1 of Book 2) that for a basis of orthogonal unit vectors the distinct scalar products take a very simple form

$$\mathbf{e}_i \cdot \mathbf{e}_i = 1 \text{ (all } i), \qquad \mathbf{e}_i \cdot \mathbf{e}_j = \mathbf{e}_j \cdot \mathbf{e}_i = 0 \text{ } (i \neq j).$$

You now have most of the mathematical 'tools' needed in Chapter 2 – and in much of the whole book!

The last step in generalizing what came out of the Stern-Gerlach experiment is to write down an expression for the operator X to be associated with observable X: this will be similar to the operator S_z defined in (1.9) and follows on changing from 2 to n dimensions (it's as easy as that!). Instead of (1.9) we're going to get

$$\mathsf{X} = X_1 \mathsf{P}_1 + X_2 \mathsf{P}_2 \dots = \sum_{i=1}^n X_i \mathsf{P}_i,$$
(2.4)

where, you'll remember, the projection operator P_i simply picks out the component $c_i \mathsf{e}_i$ from any state vector ψ it works on – as in (2.2), above.

Having found this 'geometrical' form of the operator, let's confirm that the expression

$$\langle X \rangle = \psi \cdot (\mathsf{X}\psi) \tag{2.5}$$

leads correctly to the expectation value of X in the form (2.3). This follows when the state vector is expressed in terms of its components, as in (2.1). Thus, using (2.2) and (2.4), we find

$$\begin{aligned} \mathsf{X}\psi &= \mathsf{X}(c_1\mathsf{e}_1 + c_2\mathsf{e}_2 + c_3\mathsf{e}_3 + \dots) \\ &= (X_1\mathsf{P}_1 + X_2\mathsf{P}_2 \dots)(c_1\mathsf{e}_1 + c_2\mathsf{e}_2 + c_3\mathsf{e}_3 + \dots) \\ &= X_1c_1\mathsf{e}_1 + X_2c_2\mathsf{e}_2 \dots + X_nc_n\mathsf{e}_n \end{aligned}$$

and, on putting this in (2.5), the expectation value becomes

$$\begin{aligned} \langle X \rangle &= \psi \cdot (\mathsf{X}\psi) \\ &= (c_1 \mathsf{e}_1 + c_2 \mathsf{e}_2 + c_3 \mathsf{e}_3 + \dots) (X_1 c_1 \mathsf{e}_1 + X_2 c_2 \mathsf{e}_2 + X_3 c_3 \mathsf{e}_3; \dots) \\ &= X_1 c_1^2 + X_2 c_2^2 + X_3 c_3^2 + \dots \end{aligned}$$

- in agreement with (2.3). This confirms the general principle stated in (1.10).

2.2 Some other spin systems

So far, the only actual system considered has been a single 'spin- $\frac{1}{2}$ ' particle, typically an electron. But the original Stern-Gerlach experiment used atoms of silver, each one containing many electrons (47 of them!), buzzing about like a swarm of bees around a heavy **nucleus** (as you may remember from Book 5). The reason why (as we now know) the system behaves like a single spin- $\frac{1}{2}$ particle is that most of the electrons 'stick together' in pairs, with their spin vectors coupled together to give a *resultant* spin zero. So the next thing we have to think about is how spins can be coupled to a resultant *total* spin with quantum number $S = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$, and with a two-particle state vector, which might be called $\Psi(1, 2)$, using capital letters for states of two particles. For example, we might use $\Psi(1, 2) = \alpha(1)\beta(2)$ to show that Particle 1 is in the up-spin state and Particle 2 in the down-spin state.

And how would you describe an observable, such as the z-component of total spin? In classical physics, each spin component would be the sum of components for the separate particles and would be found by measuring $S_z(1) + S_z(2)$. But in quantum physics $S_z(1)$ and $S_z(2)$ have associated operators $S_z(1), S_z(2)$. Now we've just supposed that Particle 1 was in up-spin state $\alpha(1)$ – in which case $S_z(1)\alpha(1) = \frac{1}{2}\alpha(1)$ – and similarly Particle 2 was in down-spin state $\beta(2)$, for which $S_z(2)\beta(2) = -\frac{1}{2}\beta(2)$. In classical physics angular momentum vectors can be added by using the laws of vector addition, which you can visualize by putting arrows together 'head to tail' and joining the first 'tail' to the last arrow head to find the resultant vector. We can use the same picture in quantum physics, thinking of two spins coupled to zero resultant when they point in opposite directions; or coupled to give S = 1 when they point the same way. But that's only to help our imagination: the pictures just remind us of the things we've done in classical physics (Book 4) and set us off in a similar direction. In the end we'll be looking for numbers that relate to real experiments.

We're starting with a system of two spin- $\frac{1}{2}$ particles, Particle 1 and Particle 2, with spin angular momentum having a z-component $S_z(1)$ for the first and $S_z(2)$ for the second. Suppose each observable can be independently measured, so we can say, for example, $S_z(1)$ has the value $+\frac{1}{2}$ and $S_z(2)$ has the value $-\frac{1}{2}$. (You don't have to say how you could get such results – this is a "thought experiment"!). Another way of describing this is to say that Particle 1 is in State 1, with spin up, and Particle 2 in State 2 (spin down). To put this in symbols we can say the situation is described by the two-particle state vector $\Psi(1,2) = \alpha(1)\beta(2)$. We can also define a two-particle operator S_z (without the labels) for the whole system:

$$S_z = S_z(1) + S_z(2) \tag{2.6}$$

and can verify that $\Psi(1,2)$ is an eigenvector of S_z with eigenvalue $S_z = S_z(1) + S_z(2) = \frac{1}{2} + (-\frac{1}{2}) = 0$. Thus

$$\mathsf{S}_{z}\alpha(1)\beta(2) = [\frac{1}{2}\alpha(1)]\beta(2) + \alpha(1)[-\frac{1}{2}\beta(2)] = (\frac{1}{2} - \frac{1}{2})\alpha(1)\beta(2) = 0 \times \alpha(1)\beta(2),$$

where each term in (2.6) works only on the spin factor (α or β with the same particle label.

Using the S^+ and S^- operators

First we need to show that the two-particle operators like (2.6) behave in the same way as those for a single particle. For example, do the commutation relations (1.18) also apply to the *total* spin operators S_x, S_y, S_z ? To show that they do you'll have to put $S_x = S_x(1) + S_x(2)$ and $S_y = S_y(1) + S_y(2)$ in the first equation of (1.18 and work out the result. If it turns out to be the same as iS_z , where $S_z = S_z(1) + S_z(2)$ then you can guess that the two-particle operators *will* satisfy the commutation relations (1.18), because all three equations follow simply from a 'cyclic permutation' of the labels x, y, z. We'll try it in an Example:

Example 2.1 Commutation relations for 2-particle operators

The first commutation rule is $S_x S_y - S_y S_x = i S_z$ and on putting in the 2-particle forms of the operators we get

$$\begin{aligned} \mathsf{S}_x\mathsf{S}_y - \mathsf{S}_y\mathsf{S}_x &= [\mathsf{S}_x(1) + \mathsf{S}_x(2)][\mathsf{S}_y(1) + \mathsf{S}_y(2)] - [\mathsf{S}_y(1) + \mathsf{S}_y(2)][\mathsf{S}_x(1) + \mathsf{S}_x(2)] \\ &= [\mathsf{S}_x(1)\mathsf{S}_y(1) + \mathsf{S}_x(1)\mathsf{S}_y(2)] + [\mathsf{S}_x(2)\mathsf{S}_y(1) + \mathsf{S}_x(2)\mathsf{S}_y(2)] \\ &- [\mathsf{S}_y(1)\mathsf{S}_x(1) + \mathsf{S}_y(1)\mathsf{S}_x(2)] + [\mathsf{S}_y(2)\mathsf{S}_x(1) + \mathsf{S}_y(2)\mathsf{S}_x(2)]. \end{aligned}$$

It all looks like a lot of hard work! But the spin operators with labels '1' only work on factors $\alpha(1)$ and $\beta(1)$, not touching those for Particle 2: so, for example, $S_x(1)S_y(2)$ on the second line can be written in reverse order as $S_y(2)S_x(1)$ and is cancelled by the same term on the following line with a negative sign. It's only when the operators in a term refer to the *same* particle that they have a non-zero sum. That makes life much easier: the only terms left in the last expression for the commutator $S_xS_y - S_yS_x$ are $S_x(1)S_y(1) - S_y(1)S_x(1)$ and $S_x(2)S_y(2) - S_y(2)S_x(2)$. They are identical except for particle number, being $iS_z(1)$ and $iS_z(2)$, respectively, and their sum is just iS_z .

The conclusion from Example 2.1 applies, as you can confirm, to any number of particles:

If the spin operators for Particle *j* satisfy commutation relations $S_x(i)S_y(i) - S_y(i)S_x(i) = iS_z(i)$ etc., then the corresponding *total* spin operators $S_x = \sum_i S_x(i)$ have exactly similar properties.

(2.7)

As a first example of how to use this conclusion let's look at the step-up and step-down operators in (1.20): they're examples of "shift operators" (shifting the eigenvalue m up or down) and, being expressed in terms of S_x and S_y , will have the same properties as the corresponding 1-particle operators. So we'll start from the 2-particle spin state with $\Psi(1,2) = \alpha(1)\alpha(2)$.

Example 2.2 Spin states for a 2-particle system

If we apply the operator (2.6) to $\Psi(1,2) = \alpha(1)\alpha(2)$ the result will clearly be $S_z\Psi(1,2) = (\frac{1}{2} + \frac{1}{2})\Psi(1,2)$, a state with spin z-component M = 1. We suspect that the total spin, with eigenvalue S will have S = 1 (the maximum value of the z-component M), but we should prove it.

For a single spin- $\frac{1}{2}$ particle we have $m = s = \frac{1}{2}$, but for two 'parallel-coupled' spins the maximum zcomponent is M = 1. What can we say about the total spin, which (in classical physics) has a square given as $S^2 = S_x^2 + S_y^2 + S_z^2$? In quantum physics, this will have an associated operator $S^2 = S_x^2 + S_y^2 + S_z^2$ and to find its eigenvalues we must express it in terms of 'known' operators. But we only know the properties of S_z , S^+ , S^- ; so let's play with the last two:

$$S^{-}S^{+} = (S_{x} - iS_{y})(S_{x} + iS_{y}) = S_{x}^{2} + S_{y}^{2} + i(S_{x}S_{y} - S_{y}S_{x}),$$

$$S^{+}S^{-} = (S_{x} + iS_{y})(S_{x} - iS_{y}) = S_{x}^{2} + S_{y}^{2} - i(S_{x}S_{y} - S_{y}S_{x}),$$

noting that both products lead to $S_x^2 + S_y^2$. But the first form gives an extra bit $i \times iS_z$, while the second gives an extra $-i \times iS_z$. The unwanted 'extra bits' (when you're aiming for $S_x^2 + S_y^2 + S_z^2$) can be removed by adding S_z , to the first form, or subtracting S_z from the second form. Then, by adding S_z^2 to both, you'll find the two alternative forms of S^2 :

$$S^2 = S^-S^+ + S_z + S_z^2$$
, $S^2 = S^+S^- - S_z + S_z^2$.

Example 2.2 has given us two alternative expressions for the total spin operator S^2 in terms of operators we've already used. They are

$$S^2 = S^-S^+ + S_z + S_z^2,$$

 $S^2 = S^+S^- - S_z + S_z^2$

(2.8)

and they hold good for any number of particles, not just 2.

If we have a spin state $\Psi_{S,M}$, where S is the maximum value that the quantum number M can take, we know that S_z simply multiplies $\Psi_{S,M}$ by the quantum number M – which is an *eigenvalue* of S_z . On the other hand S^- reduces the value of M by 1, while S^+ increases it by 1, so

$$S^{-}\Psi_{S,M} = k_{M}^{-}\Psi_{S,M-1}, \qquad S^{+}\Psi_{S,M} = k_{M}^{+}\Psi_{S,M+1},$$
(2.9)

where the numerical constants have been put in because we don't know if the shift operators also change the *length* of a state vector - just multiplying it by a number. (We'll have to check that later.)

Since $\Psi_{S,M=S}$ (with M = S) was taken to be the 'top' state, it can't be stepped up: S^+ must destroy the state vector, multiplying it by zero. But it can be stepped down to give a whole series of eigenstates of S_z with quantum numbers M = S, S - 1, S - 2, ... until you reach the 'bottom' state, with M = -S, which must be destroyed by the operator S^- . The spectrum of possible states is *bounded* by those with $M = \pm S$. Examples of the states for a 2-particle system, reached by using the step-down operator can be collected in a 'picture' as below, where the values of S, M are shown on the left, followed by the expressions for $\Psi_{S,M}$ obtained from the top state with M = S).

• (1,1) $\Psi_{1,1} = \alpha(1)\alpha(2)$

• (1,0)
$$\Psi_{1,0} = \beta(1)\alpha(2) + \alpha(1)\beta(2)$$
 • (0,0) $\Psi_{0,0} = \beta(1)\alpha(2) - \alpha(1)\beta(2)$

• (1,-1) $\Psi_{1,-1} = \beta(1)\beta(2)$

Everything looks fine; but where did $\Psi_{0,0}$ come from? It doesn't belong to the same 'family', corresponding instead to S = 0. You can easily check that by working on it with either of the operators in (2.8) to get the eigenvalue of S^2 (do it for yourself!). If you do the same for any of the eigenvectors with S = 1 (e.g. $\Psi_{1,1} = \alpha(1)\alpha(2)$) the answer will be S(S + 1), not the S^2 you were probably expecting. But if you go back to Section 1.1, you'll see that even for a single spin-half system the eigenvalue of S^2 was given by the same rule, but with $S = \frac{1}{2}$, as $\frac{1}{2}(\frac{1}{2} + 1)$. So there are still some mysteries to clear up. You should also check that the eigenvectors on the same line (i.e. with different eigenvalues of S^2 are orthogonal. (Find their scalar product, using the properties of the α s and β s – namely $\alpha \cdot \alpha == \beta \cdot \beta = 1$, and $\alpha \cdot \beta = \beta \cdot \alpha = 0$ for each particle: when this is zero the

vectors are orthogonal.)

As another example let's take a 3-particle system, starting from the state vector for all spins parallel-coupled with $S = M = \frac{3}{2}$. (The particle labels can be left out if we agree to keep them always in the order 1,2,3, so $\alpha\beta\alpha$ will be short for $\alpha(1)\beta(2)\alpha(3)$.)

• $\left(\frac{3}{2}, \frac{3}{2}\right)$ $\Psi\left(\frac{3}{2}, \frac{3}{2}\right) = \alpha \alpha \alpha$

•
$$\left(\frac{3}{2}, \frac{1}{2}\right)$$
 $\Psi\left(\frac{3}{2}, \frac{1}{2}\right) = \beta\alpha\alpha + \alpha\beta\alpha + \alpha\alpha\beta$

- $\left(\frac{3}{2}, -\frac{1}{2}\right)$ $\Psi\left(\frac{3}{2}, -\frac{1}{2}\right) = \beta\beta\alpha + \beta\alpha\beta + \alpha\beta\beta$
- $(\frac{3}{2}, -\frac{3}{2})$ $\Psi(\frac{3}{2}, -\frac{3}{2}) = \beta\beta\beta$

Note that the *M*-values still go down in unit steps as you go down the page, but the total spin quantum number is now *half an odd integer* and there is no state for M = 0

Again it's possible to find another family of eigenvectors, this time with $S = \frac{1}{2}$ instead of $\frac{3}{2}$, namely $\Psi(\frac{1}{2}, \frac{1}{2}), \Psi(\frac{1}{2}, -\frac{1}{2})$. In general, the *S* quantum number has allowed values going down in unit steps until you reach the lowest value 0 (for an even number of spins) or $\frac{1}{2}$ (for an odd number) and for any given *S* there are 2S + 1 states of different *M*.

It remains only to find a general formula for the eigenvalues of S^2 and to 'normalize' the eigenvectors. This is done in the next Section.

2.3 Vector spaces with a Hermitian metric

That sounds like quite a mouthful of strange words – but in fact we've been using vectors and the operators that send one vector into another ever since Book 2. And we know that two vectors have a **scalar product**, a number that determines their *lengths* and the *angles* between them, and a **vector product**, which is another vector. We've introduced sets of **basis vectors** $\mathbf{e}_1, \mathbf{e}_2, \dots \mathbf{e}_n$ as in (2.1), to define an *n*-dimensional space, and we remember from Book 2 (Section 7.1) that a square array of all possible unit vector scalar products, such as

$$\begin{pmatrix} \mathbf{e}_1 \cdot \mathbf{e}_1 & \mathbf{e}_1 \cdot \mathbf{e}_2 & \mathbf{e}_1 \cdot \mathbf{e}_3 \\ \mathbf{e}_2 \cdot \mathbf{e}_1 & \mathbf{e}_2 \cdot \mathbf{e}_2 & \mathbf{e}_2 \cdot \mathbf{e}_3 \\ \mathbf{e}_3 \cdot \mathbf{e}_1 & \mathbf{e}_3 \cdot \mathbf{e}_2 & \mathbf{e}_3 \cdot \mathbf{e}_3 \end{pmatrix}$$
(2.10)

completely determines the **metric** of the space – the lengths of all the vectors it contains and the angles betwen them. All that, was for vectors of the form $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{e}_i$ with *real* components $c_1, c_2, ..., c_n$. But now, in quantum mechanics, the 'imaginary unit' (*i*), with $i^2 = -1$, is beginning to appear (e.g. in the basic equations (1.16)) and sooner or later we'll have to deal with vectors whose components are **complex numbers** (Book 2, Section 5.2). The next Example shows how we can get round this difficulty:

Example 2.3 Vectors with complex components

When vectors were first introduced, in Book 2 of the Series, space was taken to be 3-dimensional with a metric defined in (2.10). The 'length' v of any vector $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$ was then determined by the sum of squares of the components: $v^2 = v_1^2 + v_2^2 + v_3^2$ and any length (the distance between two endpoints) could thus be defined as a *positive number*. This is one of the fundamental axioms of Euclidian geometry: $\mathbf{v} \cdot \mathbf{v} > 0$ for all non-zero vectors. But when complex numbers are admitted this is no longer true. For example, if \mathbf{v} is multiplied by i the square of its length becomes $-\mathbf{v} \cdot \mathbf{v}$, which is *negative*. From any complex number, however, a real number can be obtained by introducing the *complex conjugate*: if z = x + iy where x and y are real, the conjugate is defined as $z^* = x - iy$, simply by reversing the sign of i. Similarly, the number associated with the squared length of vector \mathbf{v} may be re-defined in terms of its components as $v_1^*v_1 + v_2^*v_2 + v_3^*v_3$ instead of $v_1^2 + v_2^2 + v_3^2$. More generally, for any two vectors \mathbf{u} and \mathbf{v} , their scalar product will be defined as $u_1^*v_1 + u_2^*v_2 + u_3^*v_3$ and instead of calling it " $\mathbf{u} \cdot \mathbf{v}$ " we'll have to use a new name for it. Several notations are commonly used, but here we'll just put the vectors side by side, with a vertical bar between them: thus, ($\mathbf{u} | \mathbf{v}$).

For vectors with complex components, the left-hand vector in any scalar product will be replaced by one with -i in place of i in all its components.

The result obtained in Example 2.3 – where $\mathbf{u} \cdot \mathbf{v}$ had to be re-defined as $(\mathbf{u}|\mathbf{v})$ to admit complex components – means that the **metric matrix** (2.10) should be replaced by

$$\begin{pmatrix} (\mathbf{e}_{1}|\mathbf{e}_{1}) & (\mathbf{e}_{1}|\mathbf{e}_{2}) & (\mathbf{e}_{1}|\mathbf{e}_{3}) \\ (\mathbf{e}_{2}|\mathbf{e}_{1}) & (\mathbf{e}_{2}|\mathbf{e}_{2}) & (\mathbf{e}_{2}|\mathbf{e}_{3}) \\ (\mathbf{e}_{3}|\mathbf{e}_{1}) & (\mathbf{e}_{3}|\mathbf{e}_{2}) & (\mathbf{e}_{3}|\mathbf{e}_{3}) \end{pmatrix}$$
(2.11)

for a space in which the vectors are allowed to have complex components. In general $(\mathbf{u}|\mathbf{v})$ will stand for the **Hermitian scalar product**: the left-hand vector in any such product, by convention, will have *i* replaced by -i in all complex components. Such spaces may also be *n*-dimensional and two vectors, $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \ldots + v_n\mathbf{e}_n$ and $\mathbf{v}^* = v_1^*\mathbf{e}_1 + v_2^*\mathbf{e}_2 + \ldots + v_n^*\mathbf{e}_n$, are often said to be "dual" to each other.

Spaces of this kind are described as having a **Hermitian metric**, being named after the French mathematician Hermite.

Finally, note that the order of the vectors in a Hermitian scalar product is no longer

unimportant. For any two vectors, \mathbf{u} and \mathbf{v} ,

$$(\mathbf{u}|\mathbf{v}) = (\mathbf{v}|\mathbf{u})^* \tag{2.12}$$

as you can easily see because the scalar product on the right-hand side is short for $v_1^*u_1 + v_2^*u_2 + v_3^*u_3 + \dots$ Since $(z^*)^* = z$ for any number z (change i to -i and do the same again!) its complex conjugate in (2.12) becomes $v_1u_1^* + v_2u_2^* + v_3u_3^* + \dots$ and this is the same as $(\mathbf{u}|\mathbf{v})$.

2.4 What can you do with all that mathematics?

We're starting to work with vectors and operators: in quantum mechanics the vectors are usually defined in a space of the kind introduced in the last Section and the operators, like S_x, S_y, S_z , are 'recipes' for getting from one vector to another. For example, the operators defined in (1.20) for a single spin- $\frac{1}{2}$ particle and generalized for more than one particle in (2.9), work on vectors of 'spin-space' according to

$$S^+ \Psi_{S,M} = k_M^+ \Psi_{S,M+1}, \qquad S^- \Psi_{S,M} = k_M^- \Psi_{S,M-1}.$$
 (2.13)

It's now possible to find what values the constants k_M^+ and k_M^- must have in order for the new states to be normalized to unity. For example, we require $\Psi_{S,M-1} = \mathsf{S}^- \Psi_{S,M}/k_M^-$ to have unit modulus, supposing that this is true for the 'un-shifted' state $\Psi_{S,M}$. This looks a bit difficult because the scalar product we want to set equal to 1 has the operator $\mathsf{S}^$ working on both the left- and right-hand vectors: it is $(\mathsf{S}^- \Psi_{S,M} | \mathsf{S}^- \Psi_{S,M})/(k_M^{-*} k_M^-)$, where the k-factors have been moved out of the scalar product (remembering that the left-hand one must be replaced by its complex conjugate).

The next short Example shows how we can deal with things like that.

Example 2.4 The adjoint of an operator

Moving things around inside a scalar product is easy if you use the basic property (2.12). Suppose we have state vectors Ψ and Ψ' and two operators A and B.

First put $\Psi' = A\Psi$, so you can say from (2.12) that $(\Psi|A\psi) = (A\Psi|\Psi)^*$. By taking the complex conjugate of both sides this gives also $(A\psi|\Psi) = (\Psi|A\Psi)^*$. So you can move the operator across the bar, changing the scalar product only into its complex conjugate. If there's no change at all, the scalar product is a real number and the operator is said to be a **Hermitian operator**.

Next look at the operator C = A + iB, where both A and B are Hermitian operators. From above it follows that $(\Psi|C\Psi) = (\Psi|A\Psi) + i(\Psi|B\Psi) = (A|iB\Psi) = (\Psi A|\Psi) + (-iB\Psi|\Psi)$ where *i* has been replaced by -i in moving the operator *iB* to the left. If we define a new operator, usually denoted by $C^{\dagger} = A - iB$ and called the **adjoint** of C, the last result may be written more neatly as $(C\Psi|\Psi) = (\Psi|C^{\dagger}\Psi)$. Clearly, Hermitian operators can also be described as "**self-adjoint**".

The rule just discovered also works the other way round: an operator of the form C = A + iB, where A and B are both Hermitian, can be moved from one side of a scalar product

to the other provided it is replaced by its adjoint operator $C^{\dagger} = A - iB$. Thus,

$$(\mathsf{C}\Psi|\Psi) = (\Psi|\mathsf{C}^{\dagger}\Psi). \tag{2.14}$$

We can now go back and finish the job we started in (2.13): we wanted to use the stepdown operator S^- , starting from a 'top' state with M = S (the maximum value) to get states with M = S - 1, M = S - 2. and so on. But we wish to choose the constant k_M^- , in the definition $S^-\Psi_{S,M} = k_M^-\Psi_{S,M-1}$. so that the state vector stays normalized in every step. This requires that the scalar product of the vector with itself is equal to unity for all values of M. In other words we must choose the constant so that

$$(\mathsf{S}^{-}\Psi_{S,M}|\mathsf{S}^{-}\Psi_{S,M})/(k_{M}^{-*}k_{M}^{-}) = 1.$$
 (2.15)

Now the 'state labels' S, M are determined by the eigenvalues of the operators S^2 and S_z , so to get the value of the scalar product in (2.15) let's try to express it in terms of them. First, note that the operators $S^- = S_x - iS_y$ and $S^+ = S_x + iS_y$ are adjoint: $S^- = (S^+)^{\dagger}$. So S^- on the left in the scalar product can be moved to the right (jumping over the bar) if we replace it by S^+ . Thus

$$(\mathsf{S}^{-}\Psi_{S,M} \,|\, \mathsf{S}^{-}\Psi_{S,M}) = (\Psi_{S,M} \,|\, \mathsf{S}^{+}\mathsf{S}^{-}\Psi_{S,M}).$$

But then, from (2.8), it is known that $S^+S^- = S^2 + S_z - S_z^2$ so altogether

$$(\mathsf{S}^{-}\Psi_{S,M} | \mathsf{S}^{-}\Psi_{S,M}) = (\Psi_{S,M} | [\mathsf{S}^{2} + \mathsf{S}_{z} - \mathsf{S}_{z}^{2}]\Psi_{S,M}).$$

And we're finished! The three operators inside the square brackets just multiply the state vector $\Psi_{S,M}$ by the eigenvalues, S(S+1), M, and M^2 , respectively. On going back to (2.15), putting in these results, and removing the unnecessary $\Psi_{S,M}$ -factors, the final result is

$$|k_M^-|^2 = (\mathsf{S}^-\Psi_{S,M}|\mathsf{S}^-\Psi_{S,M}) = S(S+1) + M - M^2.$$
(2.16)

This can also be written (check it!) as (S+M)(S-M+1) and, on putting this value in (2.13), the result of the operation becomes $S^-\Psi_{S,M} = \sqrt{(S+M)(S-M+1)} \Psi_{S,M-1}$.

The step-up operator in (2.13) can be handled in the same way. Both results are important, applying to angular momentum in general (not only to spins), so we'll put them together in a box:

$$\begin{split} \mathbf{S}^{-}\Psi_{S,M} &= \sqrt{(S+M)(S-M+1)} \, \Psi_{S,M-1}, \\ \mathbf{S}^{+}\Psi_{S,M} &= \sqrt{(S-M)(S+M+1)} \, \Psi_{S,M+1}. \end{split}$$

(2.17)

Before starting on Chapter 3 remember what you've already done: you're already deep into quantum mechanics and, even if you need a bit more time for so many new ideas to sink in, you now have all the basic mathematical tools for going on to the end of the book.

But the most miraculous thing about what you've done so far is that it all came out of so little – a single experiment that was done several years before quantum mechanics really got started, together with a few general ideas about 'space' and 'measurement'. Of course you needed to look back at other Basic Books (especially Books 1 and 2) – and you had to do quite a lot of thinking. But it's all something you *can* do.

Chapter 3

And now to Schrödinger!

3.1 Five years on -

A note to the reader. Some parts of this Chapter will be difficult on first reading them; so when you find it hard just skip over the bits you don't understand, getting an idea of the way the argument is going and of the results that are coming out. You can come back to the proofs and other details when you feel ready – you're not preparing for an exam!

Chapters 1 and 2 took us from the Stern-Gerlach experiment, made in 1921, deep into quantum mechanics, showing how things *might* have happened. But that's not the way they *did* happen. It was not until five years later, that Erwin Schrödinger formulated his famous equation, which still provides the most commonly used form of quantum mechanics. About the same time, Werner Heisenberg made another, completely different approach to the subject; and by around 1930 Paul Dirac had shown that the alternative approaches were mathematically equivalent. At that point mathematicians like Weyl and von Neuman entered the field and turned attention towards the *interpretation* of quantum mechanics – about which, 80 years later, there is still much argument. But the fact is that the theory works beautifully and allows us to make predictions of things like the structure and properties of matter in its many forms – from atoms to molecules and solids – at a level of accuracy set only by the power of our computers. And even down inside the nucleus, where new particles are still being discovered, quantum mechanics (with tiny corrections to include the effects of radiation) still provides the best way we have for predicting what goes on.

Dirac's formulation is not far from the one already uncovered in the earlier chapters, so you already have most of the background you need for understanding quantum mechanics in its present form.

At the same time it should not be forgotten that quantum mechanics has a long and fascinating history, starting about 1900 when Planck, de Broglie, Einstein and others recognised that 'classical' physics could no longer be applied in dealing with radiation (see Book 10) and its interaction with matter. And continuing with attempts by Bohr and his collaborators to lay the foundations of an early form of quantum theory.

In Book 11 we're wanting to 'get there fast' by using what we know *now*, but some day perhaps you'll be able to read the whole story. Meanwhile, let's continue to build up from the foundations laid in Chapters 1 and 2, starting from the idea that measuring some physical quantity or 'variable' is like observing its value a large number of times and then taking an average as the 'best estimate': if the experiment is repeated yet again, without any kind of change, this estimate is taken as the **expectation value** of the variable. (Alternatively, instead of doing the same experiment a large number of times on a single system, we can do it on a large 'assembly' of identical copies of the system: that's what happens in the Stern-Gelach experiment, where the beam of silver atoms is separated by the Analyser in Fig.4 into two smaller assemblies of up-spin and down-spin atoms. It doesn't matter which interpretation is used.)

So far our 'system' has been a single particle, spinning around the z-axis, and the variable we've been talking about has been the angular momentum S_z . This is a very special observable because it is found to have only two possible values, $\pm \frac{1}{2}$ in units of a fundamental quantity \hbar . The observable values are *discrete* and there are only *two* of them: the spin is a "two-valued observable".

However, most of the quantities we measure in science are not like that: they can take any number of possible values, usually in the range 0 to ∞ , or $-\infty$ to $+\infty$; and they are *continuously* variable. (Examples are the distance x of a particle from a fixed point on the x-axis, defined in the interval $(-\infty, +\infty)$; and its distance from a fixed point in space, defined in the interval $(0, +\infty)$, which is always positive).

If we could measure the value of some such quantity x (e.g. the x-coordinate of a particle's position at some given instant) we might record an approximate value $x \approx x_1$, meaning that the value lies in a small interval $(x_1, x_1 + \delta)$. Here we imagine the x-axis is marked out in points at $x_1, x_2, x_3, \ldots x_k$, ... separated by spaces of width δ – the smallest distance we can measure. And when we say $x \approx x_k$ we'll mean that, as far as we can judge, the value of x lies within the interval $(x_k, x_k + \delta)$. This gives us a mathematical way of saying where the particle is at a given time, allowing for the fact that our measurements may be **uncertain**. And this allows us to generalize what was done in Sections 1.1 and 1.2.

Every attempt at measuring x means deciding which category it belongs to: if we say $x \approx x_k$ we'll mean its value appears to be in the tiny interval (x_k, x_{k+1}) and by counting the number of times (n_k, say) we find it there, in a very large number (N) of experiments, we can assign the variable a probable value x_k , with a probability $p_k = n_k/N$. You know something about probabilities from Book 5 (Chemistry), and even about **probability** *densities* (probabilities per unit range of what you're talking about); if you've forgotten, read Chapter 4 of Book 5 again.

Now for every value of k we're asking if the recorded value of x lies in the kth interval. If it does, we increase n_k by 1; so n_k is a 'counting index', which we use to get the probability of finding a result in this category, namely $p_k = n_k/N$. But the value you get for p_k will also depend on how big δ is; if you double δ , the number of measurements you record will also be approximately doubled. So we can say $p_k \approx p(x_k)\delta$ where $p(x_k)$ becomes a continuous function of x, evaluated at the point $x = x_k$. In the notation of the differential calculus (Book 3) we take an *infinitesimal* interval $\delta = dx$ and write $p_k = p(x)dx$ for the

probability of finding a result in the kth category. The continuous function p(x) is the probability *density* for finding x in unit range 'at' the point $x = x_k$.

3.2 Pictorial representation: the wave function

Remember Figure 4, where we pictured the state vector ψ for a two-state system in terms of unit vectors α and β along two orthogonal axes. The results of observing the spin component S_z fell into two categories, 'spin-up' and 'spin-down', and this was represented mathematically by writing $\psi = c_1 \alpha + c_2 \beta$, where the coefficients c_1 and c_2 had the properties $c_1^2 = p_1$, $c_2^2 = p_2$, being the probabilities of finding $S_z = \frac{1}{2}$ (spin-up) or $S_z = -\frac{1}{2}$ (spin-down). We introduced 'projection operators' P_1, P_2 for picking out the α and β parts of the vector ψ and were led – without really noticing – to one of the fundamental principles (1.10) of quantum mechanics.

In Section 2.1 we found that the basic results could easily be extended to the observation of other measurable quantities, where the possible results fell into any number of categories. But now we want to extend the whole mathematical framework to admit the measurment of *continuously* variable quantities, where the probabilities $p_1, p_2, ..., p_k...$ of finding discrete results $x_1, x_2, ..., x_k...$ are replaced by probability *densities* – which are continuous functions p(x) of the variable being measured. This is a big step: it leads to Schrödinger's great invention, the **wave function** $\psi(x)$, whose square modulus gives the probability density $p(x) = |\psi(x)|^2$ for finding a particle 'at' point x. (We're not going to do this the way a Mathematician would do it – but after all Schrödinger started life as an engineer, and you don't have to be a great Mathematician to do good Science.)

To avoid difficult mathematics we'll go on using k as a discrete index labelling the categories into which the observed value of x may fall. Category k will refer to the set of n_k observations in the interval $(x_k, x_k + \delta)$ and we can then take over all that was done in Section 2.1, noting only that the total number of categories n and basis vectors \mathbf{e}_k may now tend to $\pm \infty$ (as long as it is 'countable').

We'll use lower-case letters, $x_1, x_2, ..., x_k, ..., x_n$ for values of the x-coordinate of a particle's position and write, as in (2.1),

$$\Psi = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3 + \dots = \sum_{k=1}^n c_k \mathbf{e}_k, \qquad (3.1)$$

but using from now on an upper-case Ψ for the general state vector.

Again, with each basis vector there will be a projection operator P_k with the property $\mathsf{P}_k \Psi = c_k \mathsf{e}_k$; and the coefficients c_k will be chosen so that $c_k^2 = p_k = n_k/N$ (i.e. the probability, or fractional number of times, that an observed value of x falls in the interval $(x_k, x_k + \delta)$.

Note that the set of projection operators for all intervals has the following properties:

• $\mathsf{P}_k \Psi = c_k \mathsf{e}_k$ describes the operation of selecting results with x in the range $(x_k, x_k + \delta)$ from the assembly of all results contained in Ψ . The selected results form a subassembly with x having the observed value $x \approx x_k$: it is described by the vector $c_k \mathbf{e}_k$ for a **pure state**, while Ψ itself describes a **mixed state** which may contain results in all the categories.

Repeating the selection can give nothing new, so $\mathsf{P}_k\mathsf{P}_k\Psi = \mathsf{P}_k\Psi$ and since Ψ is any state vector this is an *operator* property:

$$\mathsf{P}_k^2 = \mathsf{P}_k$$
, for all k

Operators of this kind are said to be **idempotent** and we've met them before in Book 1 Section 6.2 – where the different categories referred to animals in the market place. (We've come a long way since then! but still it's worth looking back.) We now see that when an operator is associated with an *ideal* measurement it must be an idempotent operator.

• Selecting values in category j, however, from an already selected category k can give nothing:

$$\mathsf{P}_i \mathsf{P}_k \Psi = \mathsf{0} \Psi$$
, for all $j \neq k$

where the zero operator 0 destroys any vector it works on. Again this is true for any Ψ and is an operator property – after selecting only results in category k you won't find any in category j among them. Operators with this property are said to be **exclusive**.

• If you take the sum of *all* selected categories, you'll get back the original assembly, represented by the state vector Ψ . Thus

$$\sum_{k=1}^n \mathsf{P}_k \Psi = \Psi$$

and as this must be so for any Ψ it follows that

$$\sum_{k=1}^{n} \mathsf{P}_{k} = \mathsf{I},$$

the 'identity' operator which leaves any state vector unchanged. The set of all P_k is said to be a **complete set**.

We can now define a 'complete observation' of the variable x: it is one in which a fractional number $n_k/N(=p_k)$ of observations of x yield results $x \approx x_k$, for all the infinitesimal intervals, of width δ , into which the whole range is divided. Figure 5 will help to fix things in our minds.



Figure 5. Values of x in three intervals (shown by dots in Interval k) Ordinates at x_{k-1}, x_k, x_{k+1} show corresponding p_k -values.
The whole range of possible values of x is indicated by a horizontal line (the x-axis), divided into intervals that go from $x = x_k$ to $x_k + \delta$. Interval k is shown, greatly magnified, between Interval k - 1 and Interval k + 1 (with no spaces between them), and the bold dots mark the n_k recorded x values for particle positions found in Interval k.

When we talk about a particle 'at' point $x = x_k$ we'll be thinking of one of those dots. (Doesn't this remind you of Figure 16 in Book 5, where the dots marked 'hits' when you were aiming at a target?). If we draw a vertical line of height p_k , for each interval, we'll have a picture of the **probability distribution**, p = p(x), shown as a broken line.

Now let's turn to the geometrical representation of Ψ in terms of basis vectors \mathbf{e}_k , along the lines of Fig.4. But here the bold arrow represents $\Psi = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 + \dots$ and the coefficient c_k is the projection of the state vector along the direction of \mathbf{e}_k .

Don't get mixed up! This is a **representation space** with an infinite number of dimensions – nothing to do with a 'real' 3-space in which x, y, z may be coordinates labelling the position of a point. It just happens that x is a physical variable whose value we're interested in measuring. The representation space is something we've made up in our minds to give a 'picture' of the measurements we make.

Here c_k will represent the value of $|\Psi|$ at x = k (in units of δ) and when the unit becomes small enough $|\Psi| \to |\Psi(x)|$, a continuous function of the x we're measuring 'in real space'. To see how nicely things work out let's suppose Ψ is a unit vector as in Fig.4.

Example 3.1 Going from discrete to continuous

For a *discrete* space, with the integer k running over all n components of a vector, we must put the squared length of Ψ equal to unity:

$$|\Psi|^2 = c_1^2 + c_2^2 + c_3^2 + \dots = \sum_{k=1}^n c_k^2 = 1.$$

But for a *continuous* space the corresponding condition becomes

$$|\Psi|^2 = \int_{-\infty}^{+\infty} |\Psi(x)|^2 \mathrm{d}x = 1.$$

The components of Ψ become continuous functions of x, namely $\Psi(x)$, and the square modulus of the vector becomes a definite integral (see Book 3) over all possible values of x.

The 'length', or modulus, of a state vector Ψ thus follows from the continuous function $\Psi(x)$ representing it by integration. When the observable values of the variable x run over the range $-\infty$ to $+\infty$ the square modulus is given by the integral

$$|\Psi|^2 = \int_{-\infty}^{+\infty} |\Psi(x)|^2 \mathrm{d}x = 1.$$
(3.2)

A key feature of Schrödinger's wave mechanics is that $\Psi(x)$ itself is **not an observable quantity**: only the square modulus $|\Psi(x)|^2$ has physical meaning. It gives the probability per unit range of finding the particle at point x:

$$p(x) = |\Psi(x)|^2.$$
(3.3)

And (3.2) means simply that the probability of finding it *somewhere* along the x-axis is unity, corresponding to certainty. When we study motion of a particle in 'everyday' 3space we'll find that these conclusions are general, applying to a wave function $\Psi(x, y, z)$ of all three position coordinates.

3.3 The definition of uncertainty

So far, the only physical variable we've been studying is the x-coordinate of a particle moving along the x-axis. If it is found 'at' x (i.e. in the range x to x + dx), with a probability p(x)dx, then the probability of finding it in the finite interval (x_1, x_2) will be, with the normalization of (3.2).

$$P(x_1, x_2) = \int_{x_1}^{x_2} p(x) \mathrm{d}x = \int_{x_1}^{x_2} |\Psi(x)|^2 \mathrm{d}x.$$
(3.4)

The average or 'expectation' value of x, namely $\langle x \rangle$ (often denoted by \bar{x}), follows directly from the distribution function p(x). To get it we must multiply each measured value (x_k) by the fractional number of times it is found (p(x)dx for $x = x_k)$ and take the sum of all contributions as x goes from $-\infty$ to $+\infty$ – which becomes the integral

$$\langle x \rangle = \int_{-\infty}^{\infty} x p(x) \mathrm{d}x$$
 (3.5)

One very common form of probability distribution is described by the Gaussian function, shown below in Figure 6, as used in Chapter 4 of Book 5



Figure 6. The Gaussian curve

As the curve is symmetrical about the mid-point x = 0, the average value is in this case $\bar{x} = \langle x \rangle = 0$. But there is clearly an 'uncertainty' in any measurement of x we might make: the curve indicates a considerable 'scatter' of observed values around the mean value \bar{x} . We need a single number to show how big the uncertainty is – and find it in the following Example.

Example 3.2 The uncertainty in a measured value of x

To measure the **uncertainty** in x, one can ask for the average *deviation* of x from its average value \bar{x} . If we used $x - \bar{x}$ for this quantity, we should again find zero – every positive value being exactly cancelled by a negative value (just look at the Figure). But by "deviation" we really mean the difference between x and \bar{x} without the \pm sign i.e. $|x - \bar{x}|$. In practice it is easier to use

$$\bar{D} = \langle \Delta x^2 \rangle = \langle (x - \bar{x})^2 \rangle$$
$$= \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x) dx$$
$$= \langle x^2 \rangle - 2\bar{x} \langle x \rangle + \langle \bar{x}^2 \rangle$$
$$= \langle x^2 \rangle - \bar{x}^2$$

- since \bar{x} is just a numerical factor and the last term comes from $\bar{x}^2 \langle 1 \rangle$ when p(x) in the definition (3.5) is normalized to unity. This is called the **mean square deviation** and is an essentially positive quantity. The square root of \bar{D} is called the **root mean square deviation**: it is indicated by the broken lines in Fig.6 and shows the extent of the 'spread' of observed x values around the mean value \bar{x} .

Example 3.2 has given us a precise numerical measure of the uncertainty in position of a particle moving along the x-axis and described by a wave function $\Psi(x)$; it is the square root of

$$\langle \Delta x^2 \rangle = \langle x^2 \rangle - \bar{x}^2 = \int_{-\infty}^{\infty} (x^2 - \bar{x}^2) \Psi(x) |^2 \mathrm{d}x.$$
(3.6)

At the microscopic level we can never get rid of uncertainty: if we manage to make it very small in one of the observables we're trying to measure, then we'll make it bigger and bigger in others. In the next Section we'll find how the different uncertainties are *related* and how this allows us to discover the properties of the operators that we must use in quantum mechanics.

3.4 Heisenberg's Uncertainty Principle

Note. You will find some of this Section difficult: read earlier Chapters again before starting on it (especially Example 2.8 of Chapter 2).

We noted in Chapter 1 that the commutator of two operators had a special significance: see, in particular, the commutation relations for spin operators. For any two Hermitian operators, A and B say, we could define their commutator by C = AB - BA; but this would not be Hermitian. The adjoint of C would be

$$(AB - BA)^{\dagger} = (B^{\dagger}A^{\dagger} - A^{\dagger}B^{\dagger}) = (BA - AB)$$

– since Hermitian operators are *self*-adjoint $(A^{\dagger} = A, B^{\dagger} = B)$.

In that case C would be *anti*-Hermitian, $C^{\dagger} = -C$, and would not represent anything observable. It's better then to define the commutator as

$$i\mathsf{C} = \mathsf{A}\mathsf{B} - \mathsf{B}\mathsf{A},\tag{3.7}$$

where the factor i ensures that the operator iC is Hermitian.

We start from the fact that any non-zero vector Ψ has a 'length' $|\Psi|$ which is real and positive, so for any operator X the new vector $X\Psi$ will have the property

$$(\mathsf{X}\Psi|\mathsf{X}\Psi) \ge 0 \tag{3.8}$$

Now let's choose $X = A + \lambda B$ and try to find a condition relating expectation values of A^2 , B^2 and AB, BA – which are not far from the things we're looking for. And then work through the next two Examples.

Example 3.3 What follows from $(X\Psi|X\Psi) \ge 0$? (Step 1)

On substituting $X = A + \lambda B$, where λ is any complex number, we get (remembering that λ on the left side of a scalar product must be changed to λ^*)

$$\begin{aligned} (\mathsf{X}\Psi|\mathsf{X}\Psi) &= & ([\mathsf{A}\Psi + \lambda^*\mathsf{B}\Psi]][\mathsf{A}\Psi + \lambda\mathsf{B}\Psi]) \\ &= & (\mathsf{A}\Psi|opA\Psi) + |\lambda|^2(\mathsf{B}\Psi|\mathsf{B}\Psi) + \lambda(\mathsf{A}\Psi|\mathsf{B}\Psi) + \lambda^*(\mathsf{B}\Psi|\mathsf{A}\Psi) \\ &= & (\Psi|\mathsf{A}^2\Psi) + |\lambda|^2(\Psi|\mathsf{B}^2\Psi) + \lambda(\Psi|\mathsf{A}\mathsf{B}\Psi) + \lambda^*(\Psi|\mathsf{B}\mathsf{A}\Psi) \geq 0, \end{aligned}$$

since, for Hermitian operators, the operator on the left in each scalar product can be moved to the right.

Next we have to choose λ to get a more useful form of the last result: we take it to be $\lambda = a + ib$, where a and b are arbitrary real numbers. And then we'll use a trick you learnt years ago (Book 1) in dealing with quadratic equations. It's usually called "completing the square". All this will be done in the next Example.

Example 3.4 What follows from $(X\Psi|X\Psi) \ge 0$? (Step 2)

On putting $\lambda = a + ib$ the result from Example 3.3 gives

$$\begin{aligned} (\mathsf{X}\Psi|\mathsf{X}\Psi) &= & (\Psi|\mathsf{A}^{2}\Psi+) + (a^{2}+b^{2})(\Psi|\mathsf{B}^{2}\Psi) + a(\Psi|\mathsf{A}\mathsf{B}\Psi) + a(\Psi|\mathsf{B}\mathsf{A}\Psi) + ib(\Psi|opA\mathsf{B}\Psi) - ib(\Psi|opB\mathsf{A}\Psi) \\ &= & (\Psi|\mathsf{A}^{2}\Psi+) + (a^{2}+b^{2})(\Psi|\mathsf{B}^{2}\Psi) + a(\Psi|\mathsf{D}\Psi) - b(\Psi|\mathsf{C}\Psi), \\ & \text{where } \mathsf{D} = (\mathsf{A}+\mathsf{B}), \ i\mathsf{C} = (\mathsf{A}\mathsf{B}-\mathsf{B}\mathsf{A}). \end{aligned}$$

The term $(\Psi|\mathsf{D}\Psi)$ is of no interest and (since *a* is any real number) can be removed by putting a = 0. At this point we're left with an expression for $(\mathsf{X}\Psi|\mathsf{X}\Psi)$ that can be written $(\mathsf{X}\Psi|\mathsf{X}\Psi) = P b^2 + Q b + R$, which is quadratic in the real variable *b*. The constant coefficients in this expression are

$$P = (\Psi | \mathsf{B}\Psi), \quad Q = (\Psi | \mathsf{C}\Psi), \quad R = (\Psi | \mathsf{A}^2\Psi)$$

and if you set $(X\Psi|X\Psi)$ equal to zero you get a quadratic equation of the kind you studied in Book 1 (Section 5.3). Let's rearrange it in the same way by 'completing the square'. It becomes

$$Pb^{2} + Qb + R = P\left[b^{2} + \frac{Q}{P} + \frac{R}{P}\right]$$
$$= P\left[\left(b + \frac{Q}{2P}\right)^{2} - \left(\frac{Q^{2}}{4P^{2}}\right) + \frac{R}{P}\right],$$

where the second term inside the square brackets cancels the unwanted $Q^2/4P^2$ that comes from 'completing the square' in the first term. Finally, since b is any real number, we can choose it to make the second term zero; and on putting in the values of the constants P, Q, R and multiplying throughout by $4(\Psi|\mathsf{B}^{2}\Psi)$ the whole expression becomes (check it for yourself!)

$$4(\Psi|\mathsf{B}^{2}\Psi)(\mathsf{X}\Psi|\mathsf{X}\Psi) = 4(\Psi|\mathsf{B}^{2}\Psi)[(\Psi|\mathsf{A}^{2}\Psi)(\Psi|\mathsf{B}^{2}\Psi) - \frac{1}{4}(\Psi|\mathsf{C}\Psi)].$$

Since the expectation value of the operator B^2 is real and positive, the inequality we started from $(X\Psi|X\Psi) \ge 0$ gives directly the end result

$$(\Psi|\mathsf{A}^{2}\Psi)(\Psi|\mathsf{B}^{2}\Psi) - \frac{1}{4}(\Psi|\mathsf{C}\Psi) \ge 0.$$

The result from the last Example is a very fundamental and important inequality: it is true for any two Hermitian operators, A and B, with a commutator iC = AB - BA, working on a state vector Ψ :

$$(\Psi|\mathsf{A}^{2}\Psi)(\Psi|\mathsf{B}^{2}\Psi) - \frac{1}{4}(\Psi|\mathsf{C}\Psi) \ge 0.$$
(3.9)

Remember that the first two factors are closely related to the mean square deviations in the measurable values of quantities represented by the operators A, B. The **uncertainty** Δx in position of a particle moving along the x-axis was defined in Example 3.2 as the square root of $(\Delta x)^2 = \langle x^2 \rangle - \bar{x}^2$ (average value of x^2 minus average value of x). The same definition is adopted for other observables, such as A and B, except that here we don't know how to measure such quantities; and if we agree to use the basic principle (1.10 we still don't know what operators, A and B, to associate with them. So we're working in the dark! We must depend on the symbols alone until we discover their properties – as we did for the spin operators. So we go on using A, B, ... for the operators and the ordinary letters A, B, ... for the observables we associate them with. The uncertainties in the quantities A and B will thus be taken as

$$(\Delta A)^2 = \langle A^2 \rangle - \bar{A}^2, \quad (\Delta B)^2 = \langle B^2 \rangle - \bar{B}^2, \tag{3.10}$$

where the expectation values are determined by the usual rule (1.10):

$$\bar{A} = \langle A \rangle = (\Psi | A \Psi), \quad \langle A^2 \rangle = (\Psi | A^2 \Psi)$$

$$(3.11)$$

and similarly for other operators.

Example 3.5 The Uncertainty Principle

 $(\mathsf{A}-\bar{A})(\mathsf{B}-\bar{B})-(\mathsf{B}-\bar{B})(\mathsf{A}-\bar{A})$

The first term gives

 $AB - A\bar{B} - \bar{A}B + \bar{A}\bar{B}$

while the second gives a similar result with A and B exchanged, namely

$$\mathsf{B}\mathsf{A} - \mathsf{B}\bar{A} - \bar{B}\mathsf{A} + \bar{B}\bar{A}.$$

The expectation values $(\Psi|A^2\Psi)$ and $(\Psi|B^2\Psi)$ in (3.9) are not the uncertainties themselves; but since $\Delta A^2 = (\Psi|(A - \bar{A})^2\Psi)$, and similarly for ΔB^2 , they are easily seen to have exactly the same commutator. Thus, putting $(A - \bar{A}) = A'$, and similarly $(B - \bar{B}) = B'$, the commutator iC' = A'B' - B'A' becomes the difference

The difference is simply AB - BA, all other terms cancelling in pairs.

In summary, iC' = iC and we may use A' and B' in the first two terms of (3.9) instead of A and B. When that is done the result may be re-arranged to give

$$(\Psi | {\mathsf{A}'}^2 \Psi) \times (\Psi | {\mathsf{B}'}^2 \Psi) \ge \frac{1}{4} (\Psi | \mathsf{C} \Psi)^2$$

and from (3.10) it follows that $\Delta A^2 \times \Delta B^2 \geq \frac{1}{4} (\Psi, \mathsf{C}\Psi)^2$.

Example 3.5 completes the derivation of Heisenberg's famous Uncertainty Principle in its general form. On taking the square root of the terms on both sides of the inequality \geq , in the last equation, it follows that

$$\Delta A \times \Delta B \ge \frac{1}{2} (\Psi, \mathsf{C}\Psi). \tag{3.12}$$

This gives a precise lower bound to the product of the uncertainties in any two observables. Before it can be used we must know what operators to associate with the various physical variables we want to study – components of position, velocity, energy, etc. and then we'll be able to pass finally from Classical Mechanics to Quantum Mechanics.

3.5 Schrödinger's Wave Mechanics

What are the physical observables we need to study in dealing with, say, a single particle, moving along the x-axis? In classical mechanics we'd probably think first of position (x) and velocity (v_x) ; and things that can be derived from them, like momentum $(p_x = mv_x,$ where m is the particle mass) or energy $(E = \frac{1}{2}mv_x^2 + V(v_x))$. (If you've forgotten about these things you'd better read again the first parts of Book 4, because all the same concepts are carried over into quantum theory.)

In Quantum Mechanics we're going to associate *operators* with the basic variables which, for a 1-dimensional system will be x and p_x (p also denoting *momentum*). Operators for other quantities (e.g the energy) will then be set up by starting from the classical expressions for them, in terms of x and p_x , and substituting the corresponding *operators* \times and $p_x - as$ soon as we know what they are!. The possibility of doing this is 'guesswork' and was first put forward as the **Correspondence Principle** by the Danish physicist Bohr: it just ensures that classical mechanics goes over 'smoothly' into quantum mechanics when we go down to systems of 'atomic' dimensions. Introduced as a hypothesis more than 80 years ago, it has never received a complete justification; but it was an essential step towards building quantum theory in its current form.

The first task, then, is to identify the basic operators x and p_x , through their action on a 'state vector' Ψ , which is in fact a **wave function** – a function of position, $\Psi = \Psi(x)$ in Schrödinger's form of quantum mechanics (or "wave mechanics"). Remember from Chapter 1 that the variable x was taken to be a particle coordinate, with a measurable value which was interpreted as an eigenvalue of the associated operator x – meaning that $x\Psi(x) = x\Psi(x)$. The effect of the operator x is thus simply to *multiply* the operand $\Psi(x)$ by the variable x. The second operator of the basic pair, x and p_x , is more interesting: it must describe some kind of 'recipe' for getting a new function $\Psi'(x)$ from any $\Psi(x)$ when it acts on the given function. Now ever since Book 3 we've been using **differential operators** such as D_x , the operator which *differentiates* any function of the variable x. The property of the operator is

$$\mathsf{D}_x\Psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\Psi(x) = \Psi'(x),$$

where $\Psi'(x)$ is the *first derivative* of the function $\Psi(x)$. The two operators x and D_x clearly do not commute, for

$$\mathsf{D}_x \mathsf{x} \Psi(x) = \left(\frac{\mathrm{d}}{\mathrm{d}x}x\right) \Psi(x) + \mathsf{x}\left(\frac{\mathrm{d}}{\mathrm{d}x}\Psi(x)\right) = \Psi(x) + \mathsf{x}\left(\frac{\mathrm{d}}{\mathrm{d}x}\Psi(x)\right),$$

where the operand has been differentiated as a product. Thus

$$\mathsf{D}_x \mathsf{x} \Psi(x) = \Psi(x) + \mathsf{x} \mathsf{D}_x \Psi(x).$$

This must be true for any 'well-behaved' functions, so the $\Psi(x)$ may be removed to give $xD_x - D_x x = -I$ where the identity operator I has been inserted in place of the deleted $\Psi(x)$. The remaining *operator identity* is thus essentially a *commutation relation* between x and D_x .

It seems possible that D_x could be the right operator to associate with the momentum component p_x . But unfortunately that wouldn't work, because D_x is not Hermitian. (Check that by using 'integration by parts' (see Book 3) to show that $(D\Psi, \Psi) = -(\Psi, D\Psi)$, provided that $\Psi(x)$ goes to zero at the limits $x \to \pm \infty$ of the definite integrals – where there must be zero probability of finding the particle.)

To put things right D_x must be multiplied by $\pm i$: then both D_x and i will change sign when they are moved from one side of a scalar product to the other. The resultant operator is Hermitian and we'll try it out, supposing that the correct momentum operator is

$$\mathbf{p}_x = \pm ic \frac{\mathrm{d}}{\mathrm{d}x},$$

where a constant c has been included to allow for choice of units.

Now p_x is a physical quantity with dimensions $[p_x] = M L T^{-1}$ (see earlier Books in the Series) and those of the constant c must therefore be $[c] = M L T^{-1} L = M L^2 T^{-1}$, or $[energy] \times [time]$, which is called **action**. The only fundamental constant of this kind that we've met is Planck's constant h, divided by 2π , which has been called \hbar and is also the 'natural' or 'atomic' unit of *angular momentum*. So let's write $c = \hbar$ and take the momentum operator to be

$$\mathbf{p}_x = \pm i\hbar \frac{\mathrm{d}}{\mathrm{d}x} \tag{3.13}$$

We can't yet decide on the positive or negative sign, but after going a bit further we'll be able to say which one to throw away. The eigenvalue equation to determine possible values of the momentum p_x will then be $p_x \Psi = p_x \Psi$, where Ψ is a function of x, and thus

$$\mathbf{p}_x \Psi = \pm i\hbar \frac{\mathrm{d}}{\mathrm{d}x} \Psi(x) = p_x \Psi(x). \tag{3.14}$$

This is a differential equation you've met many times, with the solution

$$\Psi = N \exp\left(\pm \frac{p_x}{i\hbar} x\right),\tag{3.15}$$

N being a normalizing factor, usually chosen to make $\int |\Psi|^2 dx = 1$.

This solution describes a **wave**, of the kind you studied in Book 10 – where it referred to something 'physical' like an electric or magnetic field propagating through space. Here, instead, Ψ is totally different because only $|\Psi|^2$ is a physical observable: the wave, pictured in Book 10 Figure 40, only satisfied a similar differential equation – something that often happens in Science!

Remember that the arbitrary constant N is often chosen to **normalize** the wave function, so that the integral of $|\Psi|^2$ over all x values is unity – meaning that the particle whose momentum we're talking about is certain to be found *somewhere* in space. We'll come back to such questions in the next Chapter, when we apply Schrödinger's equations to a few real physical systems.

To end this Chapter we extend the association of operators with observables (position, momentum, energy, etc.), so as to pass from 1-dimensional systems to their 3-dimensional counterparts.

3.6 Going into 3-space!

In the 3-dimensional space we live in ('3-space' for short) there are, of course, many more of these **dynamical variables** to deal with, even for a single particle – and even more for a many particle system like an atom, in which many electrons move around a central nucleus. It may seem that things are going to get impossibly complicated and difficult. But that's not true: for a particle moving in 3-space we'll usually need only six variables, 3 to describe its position (which can be taken as the Cartesian coordinates,x, y, z) and 3 more for the corresponding components of momentum, p_x, p_y, p_z . (For a particle with spin, 3 more are needed (S_x, S_y, S_z) , but you've already learnt how to handle them in Chapter 1 and for the present we'll leave them out.)

The variables can be grouped in "conjugate pairs", such as (x, p_x) (y, p_y) (z, p_z) , and we've discovered how to associate operators (x, p_x) with the first pair: these don't commute, which means that x and p_x cannot have simultaneously observable values (measuring one accurately will disturb the other). The variables within each pair are said to be **incompatible**; their associated operators do not commute and in the last Section we found how their **uncertainties**, such as Δx and Δp_x , are related by Heisenberg's Uncertainty Principle. But what about the variables in *different* pairs? It's clear that, as long as Schrödinger's prescription for setting up the operators holds good, the variables in different pairs must be **compatible** – because the operations of multiplication *commute* and so do those of differentiation (as long as they work on 'well-behaved' functions). In principle, then, it would seem possible to define states of a particle in which the coordinates x, y, z were all simultaneously 'knowable' (i.e. the *position* was definite) or in which the same was true of p_x, p_y, p_z (the momentum had a definite magnitude and direction). Each of these choices defines a **maximal set** of compatible observables; and each defines a **representation** in which corresponding operators can be set up. The first choice is the one due to Schrödinger; the second is called the "momentum representation". Here we'll stick to Schrödinger's representation and try to complete the formulation without more mathematical details: in brief, the operators for each conjugate pair of variables have exactly similar commutation rules; and those for more complicated quantities, like energy or angular momentum, can be obtained from their classical counterparts simply by substituting the corresponding quantum mechanical operators. The only other things to note are that the 'state vector' Ψ now becomes a wave function $\Psi(x, y, z)$, where x, y, zare 'compatible variables', and that the operation of differenting must then be understood as partial differentiation, $D_x = \partial/\partial x$.

Let's put together in a single box nearly all you need to know about the Schrödinger representation:

The state of a single particle, moving in 3-space, is determined by a wave function $\Psi(x, y, z)$. The conjugate pairs of variables (x, p_x) (y, p_y) (z, p_z) have operators associated with them according to the following scheme $x \to \mathbf{x} = x$, $y \to \mathbf{y} = y$, $z \to \mathbf{z} = z$ $p_x \to \mathbf{p}_x = \frac{\hbar}{i} \frac{\partial}{\partial x}$, $p_y \to \mathbf{p}_y = \frac{\hbar}{i} \frac{\partial}{\partial y}$, $p_z \to \mathbf{p}_z = \frac{\hbar}{i} \frac{\partial}{\partial z}$. The operators for other quantities (e.g. the energy E) follow from their classical definitions on substituting the basic operators above.

(3.16)

Note that the momentum operators have been defined with the correct choice of the \pm sign in (3.13). The minus sign ensures that, when a particle is moving, the wave that describes it is going in the same direction – it would be nonsense if the wave went off to $-\infty$ when the particle was on its way to $+\infty$! (You'll find the details in the next Chapter.)

Now you're all set to go on to Chapter 4 and make some calculations on real physical systems to see how everything works out.

Chapter 4

Energy eigenstates of a single particle

4.1 Setting up the equations

In Book 4 you studied systems of one or many particles, moving according to Newton's laws: that was Classical Mechanics. Now we're ready to use Quantum Mechanics to study some of the same systems, but on the 'microscopic' scale where Newton's laws no longer apply.

We'll start with a single particle moving in a field, such that its (classical!) **potential** energy would be V = V(x, y, z), a function of the Cartesian coordinates x, y, z, of its position in space. In Quantum Mechanics, it will be described by a wave function $\Psi(x, y, z)$, also a function of position, whose square gives us the probability/unit volume of finding it in a volume element dxdydz at point x, y, z. This is all so different from the classical picture, where the particle would move along a definite path (like the projectiles and planets you studied in Book 4) – which you could calculate by using Newton's laws, provided you knew how it started (the *initial conditions*).

The concepts of momentum, energy, etc. will still be used (in line with Bohr's "correspondence principle") but the methods of calculation will be totally different: there will be no definite observable path to calculate and to get even the *probability* of finding the particle at a point you must first set up and solve a **differential equation**, using the rules in (3.16). And when you come to *interpret* the solutions you must be ready for surprises!

The first surprise is that the energy of a moving particle can't take just *any* value: it must be one of certain "allowed values", namely the **eigenvalues** of an energy *operator*. And the system stays in such a state 'forever' – until something is done to change it. The eigenstate is a **stationary state**, reminding us of Newton's principle that a particle in a state of uniform motion continues in that state as long as it's not disturbed. In this chapter, then, we'll be concerned with finding the stationary eigenstates of a few simple systems and with calculating the quantized energy values they are allowed to take.

The first thing to do is to write down the classical expression for the energy E in terms of

position (x, y, z) and momentum components (p_x, p_y, p_z) . If the particle is not acted on by forces and is free to move anywhere in space E will not depend on its position, but only on the speed with which it is moving: the energy will be wholly **kinetic energy**, usually denoted by $T = \frac{1}{2}mv^2 = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)$, and you will be dealing with a 'free particle'. But more generally there will also be a **potential energy** V = V(x, y, z), arising from forces acting on the particle. In much of Physics and Chemistry the particles of interest are electrons and the forces are produced by electric and magnetic fields, which you've studied in Book 10.

As we're working in terms of position and *momentum* components, like $p_x = mv_x$, this means the total energy will have to be written

$$E = H(x, y, z, p_x, p_y, p_z) = \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) + V(x, y, z)$$
(4.1)

The energy expression written in this way, as a function of the position and momentum variables is called the **Hamiltonian function**, after the Irish mathematician Hamilton. We can now follow Schrödinger's prescription (3.16) for setting up the *operator* to associate with the energy E. This will be the **Hamiltonian operator**

$$H = \frac{1}{2m} (\mathbf{p}_x^2 + \mathbf{p}_y^2 + \mathbf{p}_z^2) + V(x, y, z)$$

= $-\frac{\hbar^2}{2m} \nabla^2 + V(x, y, z).$ (4.2)

Here the momentum components have been replaced by differential operators, according to (3.16), and the resultant KE operator contains ∇^2 , often called "del-squared" or "the Laplacian" (see, for example Book 10 (Chapter 6 of Book 10).

To get the allowed energy eigenstates we have to solve $H\Psi = E\Psi$, often just called "the **Schrödinger equation**". Here we'll start with the simplest problem of all, the 'particle in a box'.

4.2 Particle in a box

We take a rectangular container or 'box' with sides of lengths l_x, l_y, l_z and suppose that V = 0 inside i.e. as x goes from 0 to l_x , y from 0 to l_y , z from 0 to l_z . The eigenvalue equation for the energy will then read

$$\mathsf{H}\Psi = -\frac{\hbar^2}{2m}\nabla^2\Psi = -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\Psi = E\Psi.$$
(4.3)

Here, H apparently commutes with all the momentum operators, since (for well-behaved functions) the order in which the partial differentiations are performed doesn't matter (look back at Section 6.1 in Book 2). Thus, for a particle moving freely through space the energy E and all three components of momentum (p_x, p_y, p_z) can have simultaneously definite values – the eigenvalues of the corresponding operators. To get the 'feel' of things,

suppose first the system is 1-dimensional, with the particle moving along the x-axis, and let's take this as our first Example:

Example 4.1 Particle moving in a straight line

In this case there's only one position variable to think of and the two eigenvalue equations to solve are $p_x \Psi = p_x \Psi$, $H\Psi = E\Psi$. The first one is (using now ordinary d's instead of ∂ 's, as there is only one independent variable)

$$(i): \quad \frac{\hbar}{i}\frac{\mathrm{d}}{\mathrm{d}x}\Psi = p_x\Psi,$$

while the second is

(*ii*):
$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Psi = E\Psi.$$

We'll be interested mostly in the energy E, but the eigenvalue equation for p_x is even simpler: it is

$$\frac{\mathrm{d}}{\mathrm{d}x}\Psi = \frac{ip_x}{\hbar}\Psi,$$

which says "differentiating Ψ with respect to x must give ip_x/\hbar . On putting the constant $(p_x/\hbar) = k$, the solution (as you know from your first exercises on the calculus, in Book 3) is simply $\Psi = \exp(ikx)$. This is a **periodic function** which repeats whenever the 'angle' (kx)*increasesby* 2π .(Think of the Argand diagram, Fig.47 in Book 10, for $z = re^{i\theta}$.) The momentum eigenfunction with eigenvalue p_x then becomes $\Psi = \exp ikx$, $(k = p_x/\hbar)$, for any real value of p_x – positive or negative.

This solution describes a wave (look at Fig.40 of Book 10) – except that so far no variation with *time* has been included. Values of the **wave function** Ψ repeat whenever kx increases by 2π – or x increases by $2\pi/k$. This distance, during which Ψ runs through a complete cycle of values (i.e. a whole 'wave') is called the "wavelength" and is usually denoted by $\lambda = 2\pi/k$. Thus $k = 2\pi/\lambda$ and we can say

Wavelength:
$$\lambda = 2\pi/k = 2\pi\hbar/p_x$$
.

With this definition of λ , which is standard,

$$\Psi = \exp ikx = \exp 2\pi i(x/\lambda).$$

Now the number of complete waves per unit length is $1/\lambda$ and in the days before quantum mechanics was often called the "wave number" and also denoted by k. This has led to much confusion. Nowadays the term is usually reserved for $k = p_x/\hbar$, which is $2\pi \times$ the reciprocal wavelength λ^{-1} . We therefore use the definition

Wave number $k = p_x/\hbar$.

In terms of this now-standard definition, the momentum eigenfunction becomes simply $\Psi = \exp ikx$. The wave number is thus a convenient measure of the momentum p_x in units of \hbar : $p_x = k\hbar$.

(Note that some authors still use a different definition of k, without including the 2π factor: the freeparticle eigenfunction then becomes $\Psi = \exp 2\pi i k x$. So watch out! It's important to decide which definition you're going to use and then *stick to it*! Here we'll be using the first form, with $\Psi = \exp i k x$.) On turning to the energy eigenvalue equation (ii) it is clear that this is also satisfied by $\Psi = \exp i k x$ because $H\Psi = (1/2m)\mathbf{p}_x^2\Psi = (p_x^2/2m)\Psi$. Evidently, when the particle is free, its energy and momentum are related *classically*: $p_x = k\hbar$, $E = p_x^2/2m$. And both are free to take any real values we please: they are continuously variable. Let's summarize what came out from the Example. We started from the Hamiltonian operator (4.2), for a particle possessing only kinetic energy T, derived from the classical energy expression E = T + V by using Schrödinger's prescription (3.16).

For the simplest case of 1-dimensional motion along the x-axis, the linear momentum operator \mathbf{p}_x commutes with H and there will then be states Ψ in which both p_x and E can have simultaneously definite eigenvalues. The two eigenvalue equations are

Momentum :
$$\mathbf{p}_x \Psi = \frac{\hbar}{i} \frac{\mathrm{d}}{\mathrm{d}x} \Psi = p_x \Psi,$$

Energy : $\mathbf{H}\Psi = -\frac{\hbar^2}{2m} \frac{\mathrm{d}^2}{\mathrm{d}x^2} \Psi = E\Psi,$

$$(4.4)$$

with a common solution

$$\Psi = \exp i(p_x/\hbar)x = \exp ikx \tag{4.5}$$

All this seems mysterious, because we started by talking about a "particle in a box" and perhaps we expected to find discrete, 'quantized' energy levels. In fact, we found that energy and momentum were related *classically*: $p_x = k\hbar$, $E = p_x^2/2m$. And both were free to take any values we please; they were continuously variable. So what went wrong? *The dimensions of the box were not used in any way.* But even a 1-dimensional box has ends e.g. at x = 0 and $x = l_x$. Perhaps quantization will result when we admit the finite dimensions of the box? We must put in **boundary conditions**. Ψ is supposed to be zero outside the box, where there's no chance of finding the particle, so let's try putting $\Psi(x) = 0$ at x = 0 and $x = l_x$.

Example 4.2 Putting in the boundaries

We know from the last Example that $\Psi = e^{ikx}$ still satisfies the energy eigenvalue equation in (4.4) as long as x lies in the interval $(0, l_x)$ – for any value (negative or positive) of the real constant k. But the only solutions that are now *acceptable* are those which make $\Psi = 0$ at x = 0 and at $x = l_x$. How can these conditions be satisfied?

You've used **periodic functions**, like $\exp ikx$, before and know that they appear whenever you study any kind of wave motion, such as the vibrating string of a musical instrument (Chapter 6 of Book 3) or the oscillating field of a radio wave (Chapter 6 of Book 10). You also know that $e^{i\theta}$ can always be expressed in terms of the sine and cosine of an angle (θ) in the form $e^{i\theta} = \cos \theta + i \sin \theta$.

Now the same functions have come up again in quantum mechanics! And since the energy eigenvalue equation is *linear* the solutions $e^{\pm ikx}$, which both correspond to the same value of E, can be combined into $\Psi = Ae^{ikx} + Be^{-ikx}$ (with arbitrary constants A, B). Alternatively, from above, the same solution can be expressed as $\Psi = C \sin kx + D \cos kx$, with new values of the constants (C, D).

All we have to do now is choose the constants so that $\Psi = 0$ for x = 0 and $x = l_x$. Let's take the 'trigonometric' form, which may be more familiar to you. The term with coefficient C is fine when $x \to 0$, because the sine of angle zero is also zero; but the term with coefficient D goes to D for $x \to 0$, since the cosine of angle zero is 1. The D-term must therefore be 'thrown away' by putting D = 0. The wave function that satisfies the boundary condition at x = 0 must have the form $\Psi = C \sin kx$.

But what about the boundary at $x = l_x$? The sine of any angle $\theta = kl_x$ is zero only when it has the form $n \times \pi$, where n is any integer. It follows that that the only energy eigenfunctions that satisfy both

boundary conditions must have $k = n\pi/l_x$ and that the allowed eigenfunctions are thus

$$\Psi_n = C\sin kx = C\sin n\pi x/l_x.$$

We also know that $E = (\hbar^2/2m)k^2$, from which it follows that the *n*th energy eigenvalue is $E_n = (\hbar^2/2m)\pi^2(n^2/l_x^2)$

To summarize the conclusions from Example 4.2, when the box boundaries are taken into account and the wave function is excluded from the region outside, the allowed energy levels become

$$E_n = \frac{\pi^2 \hbar^2}{2m} \frac{n^2}{l_r^2}$$
(4.6)

with corresponding wave functions

$$\Psi_n(x) = C \sin \pi n x / l_x. \tag{4.7}$$

The integer *n* is called a **quantum number**: it labels the allowed quantum states, with energies E_n and associated wave functions Ψ_n . The constant *C* is usually chosen to **normalize** the function, so that the probability density $|\Psi|^2$ gives unit probability of finding the particle 'in the box' (in this case between x = 0 and $x = l_x$): it has the value $C = \sqrt{2/l_x}$ (Can you prove it by doing the integration? Chapter 4 of Book 3 will remind you of how.)

Now – as usual – there's nothing special about the names of the three directions in space: so you won't be surprised to be told that for a real 3-dimensional box, with sides of lengths l_x, l_y, l_z , the eigenvalues and eigenfunctions take very similar forms. The allowed energies are given by

$$E_{n_x,n_y,n_z} = \frac{\pi^2 \hbar^2}{2m} \left(\frac{n_x^2}{l_x^2} + \frac{n_y^2}{l_y^2} + \frac{n_z^2}{l_z^2} \right), \qquad (4.8)$$

where there are now three quantum numbers for each allowed state, n_x, n_y, n_z for the box sides of length l_x, l_y, l_z .

The corresponding wave functions are

$$\Psi_n(x) = C_x C_y C_z \sin \pi \frac{n_x}{l_x} x \sin \pi \frac{n_y}{l_y} y \sin \pi \frac{n_z}{l_z} z, \qquad (4.9)$$

a product of three similar factors, one for each box side.

If you want to prove these results yourself it's not too hard, so take it as another Example and use the methods already used in Book 3 (Chapter 6) and Book 10 (Chapter 6).

Example 4.3 The rectangular box in 3-space

Start from the energy eigenvalue equation for a free particle inside a box (where V = 0) in 3-dimensional space, namely (4.3):

$$\mathsf{H}\Psi = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \Psi = E \Psi.$$

and use the method of 'separation of the variables' by looking for a solution of the form $\Psi(x, y, z) = X(x)Y(y)Z(z)$. On putting this form into the equation and dividing both sides by the same product you'll get

$$\frac{1}{X}\left(\frac{\partial^2}{\partial x^2}\right)X + \frac{1}{Y}\left(\frac{\partial^2}{\partial y^2}\right)Y + \frac{1}{Z}\left(\frac{\partial^2}{\partial z^2}\right)Z = -\frac{2mE}{\hbar^2}.$$

Now each term on the left is a function of one variable only: the first depends only on x, so however you change y and z it will remain unchanged; as regards variation of x it will be a *constant*. If you call the constant E_x then you can say (using d/dx for the ordinary derivative)

$$\frac{1}{X} \left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} \right) X = E_x.$$

But this is the same equation you had in (*ii*) of Example 4.1 and has the solution (4.3); and the same argument holds for the functions Y(y) and Z(z). From (4.6) and (4.7), putting a subscript x on the quantum number n, the energy E_x becomes

$$E_{n_x} = \frac{\pi^2 \hbar^2}{2m} \left(\frac{n_x^2}{l_x^2} \right)$$

and the sum of three similar terms gives the energy $E = E_x + E_y + E_z$ in the form (4.8). The corresponding wave function, with quantum numbers n_x, n_y, n_z , is the one given in (4.9).

We can learn a lot from this simple model of a particle in a box. First you should read Chapter 1 of Book 5, which will remind you of what kind of particles matter is made up from (electrons, protons, atoms, molecules, etc.) and how big they are and so on. If you imagine a single electron in a very small box, that could give you a very crude model of a hydrogen atom; but if you take instead a whole molecule and put it in a much bigger box, that could give you a model for one of the molecules of a gas, containing many millions of molecules, confined within a container. In each case the first thing you'll want to know is the *spacing of the energy levels*, which gives an idea of how important quantization is likely to be: if the levels are very close together the system will behave 'classically' (we saw in Book 5 how the properties of gases could be well described by supposing the molecules moved according to Newton's laws); but when the levels are far apart the system is non-classical, the energy changing only in observable 'quantum jumps'.

Another thing we can learn from the simple model of a particle in a box is how the change from classical to quantum behaviour is related to the boundary conditions. In Example 4.2, the wave function e^{ikx} described a particle with positive momentum $p_x = k\hbar$ (i.e. moving along the positive x-axis); while e^{-ikx} decribed one with negative momentum $p_x = -k\hbar$ of the same magnitude, but going in the opposite direction. Neither of the two functions by itself satisfies the required boundary conditions that $\Psi = 0$ at both x = 0and $x = l_x$; but the combination $\Psi = \sin kx = -\frac{1}{2}ie^{ikx} + \frac{1}{2}ie^{-ikx}$ will do so, provided k is chosen to make $kl_x = n\pi$, and this is just the combination used in (4.7). What does this mean?

If you look back at Chapter 1 Section 1.2, you'll be reminded that when a state Ψ is not an eigenstate of an operator X but can be expressed as a linear combination of states Ψ_1, Ψ_2 which are, then the squared magnitudes of the coefficients in $\Psi = c_1\Psi_1 + c_2\Psi_2$ have the meaning

- $|c_1|^2$ = probability that observation of X will yield $X = X_1$ (i.e. that the system will be found in state Ψ_1)
- $|c_2|^2$ = probability that observation of X will yield $X = X_2$ (i.e. that the system will be found in state Ψ_2 .)

(In Chapter 3, the same interpretation followed more generally, for states represented by vectors in a space with a 'Hermitian' metric, where the components are complex numbers and $|c_i|^2 = c_i^* c_i$. Now, in the Schrödinger representation, the eigenvectors Ψ_i are replaced by wave functions and the scalar product $(\Psi|\Psi)$ becomes an integral $\int \Psi^*(x)\Psi(x)dx$. It's always the same! – the left-hand member in a scalar product is the one that carries the star. And you see that the probability interpretation extends also to continuous functions, provided you think of $\Psi(x)$ as a sort of 'x-component' of the state 'vector' Ψ ; for then $\Psi^*(x)\Psi(x)$ then gives the probability density that the variable is found with a particular value x. Think about it and after a while you'll see that all these abstract ideas and conventions combine to make good sense! More about all this in Chapter xx.

Here we've supposed that Ψ_1, Ψ_2 represent (normalized) states of definite momentum, for a particle going to the right or left, respectively, with momentum of magnitude $|p_x| = k\hbar$. It therefore follows that the first *energy* eigenstate with $E_1 = (\pi^2 \hbar^2/2m)(1/l_x^2)$ is a combination

$$\Psi_{(E=E_1)} = \frac{1}{\sqrt{2}} (\Psi_1 + \Psi_2) \tag{4.10}$$

- a mixture of two states, with equal weights $p_1 = p_2 = \frac{1}{2}$, one describing the particle going to the right, the other showing it going to the left. The same argument applies for all values of the quantum number n in (4.7), so every state with definite energy E_n can be thought of as a 50:50 mixture of momentum eigenstates in which the particle is going to right or left. You might say that the particle in the box is moving back and forth, with a definite energy, being *reflected* when it meets the boundary: but this is just a classical picture to help you to visualize the situation – not to be taken seriously!

To end this Section, let's think for a moment of the boundaries of the box. How do we describe the walls of a container in quantum mechanics? To do that, in general, we must admit the potential energy function V = V(x, y, z) in the classical model we start from.

Energy diagrams

For a particle moving in space the classical energy expression is E = T + V, where V = V(x, y, z) is the potential energy of the particle at point (x, y, z), and inside the 'box' V has been given the value zero so that E = T and the energy is wholly *kinetic*. Again, in classical physics, the energy of an isolated system remains *constant* as the particle moves around: this is the principle of **energy conservation** (Book 4, Chapter 3). The relationship between kinetic and potential energies (KE and PE) can be pictured in an **energy diagram**: this is obtained by first plotting the PE as a function of one or more coordinates. For the particle in a 1-dimensional box this takes the form shown in Figure 7:





Figure 7a. Energy diagram Bold line shows PE function V(x)E indicates energy level for n = 2

Figure 7b. Wave function Ψ_2 – corresponding to $E = E_2$

Here the 'box', in which the particle is confined, has 'walls' at x = 0 and $x = l_x$ where we put $\Psi = 0$ as a boundary condition. In the energy diagram these points correspond to a sudden 'jump' in the PE, from zero (chosen by convention) inside the box to a large value (V > E) outside. The particle would *never* be found outside the box, according to classical physics, because that would mean T = E - V < 0 – contradicting the fact that the KE is essentially positive.

One of the strange things about quantum mechanics is that particles can be found in **classically forbidden** regions of space, where Ψ must be non-zero, giving a positive probability density $p = |\Psi|^2 > 0$. To see how this comes about you need to look at the energy eigenvalue equation (4.3) for values of the coordinates which make E - V negative. In Example 4.1, however, the kinetic energy operator (T) on the left-hand side of the eigenvalue equation must be replaced by H = T + V in regions where $V \neq 0$. So outside the box

$$\mathsf{H}\Psi = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Psi + V\Psi$$

and the eigenvalue equation becomes

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Psi = (E-V)\Psi$$

where (E - V) is now a *negative* constant.

The solution is easily found: looking at (4.5), you might again expect an exponential form $\exp k'x$, where k' is a new constant (in place of the *ik* in Example 4.1). So let's try it: each differentiation d/dx multiplies the exponential by k' and the last equation then becomes

$$-\frac{\hbar^2}{2m}(k')^2\Psi = (E-V)\Psi.$$

The solution in (4.5) is thus replaced by $\Psi = \exp k'x$, where (taking out the common factor Ψ) $(\hbar^2/2m)(k')^2 = (V - E)$. The new constant k' therefore has a value such that $\hbar^2(k')^2 = 2m(V - E)$. This means that (taking the square root of both sides),

$$k' = \pm \sqrt{2m(V-E)}/\hbar, \qquad (4.11)$$

where the \pm sign has been put in because only the square of the constant is defined. The general solution in a region where V > E is accordingly of the form

$$\Psi = A \exp k' x + B \exp -k' x, \qquad (4.12)$$

where A and B are arbitrary constants, which must be chosen to satisfy any boundary conditions (as in Example 4.2). Here things are more difficult, however, because it's not just a question of choosing Ψ at the boundaries: you have to remember that Ψ is supposed to be 'well behaved' i.e. it mustn't become infinite at some points and it must be everywhere smooth and continuous. But one thing is clear, even without calculation: for negative values of x the first term in (4.12) is OK because it goes smoothly to zero on the left of the box for large x, while the second is not – because it goes to infinity. The only satisfactory function on the left is thus $\Psi = A \exp k' x$, obtained by putting B = 0. A similar argument for $x > l_x$ shows that on the right of the box we can only use $\Psi = B \exp -k' x$.

The conclusion is that a wave function that is non-zero inside the box will have 'tails' extending beyond the box boundaries – *into the classically forbidden regions*. To complete the solution you have to choose the parameters, A, B, E so that the exponentially decaying tails join smoothly onto the wave-like form inside the box. That requires a lot of numerical work, matching function values and slopes at the boundaries, and it's not surprising that physically acceptable solutions can only be found for special values of E – the **energy eigenvalues**. The only simple solutions you can get are those that result when V - E is infinitely large, in which case the exponential tails fall off so quickly that, in effect, Ψ can be taken as zero at the boundaries. We made a good choice in the Examples above! In the following Sections we'll find other cases in which nice 'clean' solutions can be obtained.

4.3 Particle in a parabolic potential field

The 'potential box' in Figure 7 was limited by 'hard' walls, where the function V(x) rose sharply from zero to a very high value – which means a 'classical' particle would have to be given a large amount of energy (E > V) to jump over the wall. Usually, however, the PE varies smoothly and although it can still form a container, keeping the particle inside, the walls are 'softer'. A typical case is the **harmonic oscillator** where a particle is bound to a centre (usually taken as origin) by a restoring force proportional to its displacement x. In this case the classical Hamiltonian is

$$H = T + V = (1/2m)p^2 + \frac{1}{2}kx^2, \qquad (4.13)$$

where p = mv is the momentum and here k is a **force constant**, the restoring force per unit displacement. The classical PE function is then a parabola (prove it!) and you might like to jump ahead and look at Figure 8, which shows the energy diagram and some of the allowed energy levels. The **frequency** of the classical oscillator is given by $\nu = \sqrt{k/m}/2\pi$, giving the *angular* frequency $\omega = 2\pi\nu = \sqrt{k/m}$. The next Example will remind you of all that.

Example 4.4 The harmonic oscillator in classical mechanics

The energy diagram in Fig.8a shows what you can expect. If you give the particle a certain total energy E, indicated by the horizontal line at height E, it will oscillate between the two points where the line

crosses the parabola. At these points the energy is purely *potential* because E = V and the KE is T = E - V = 0: the particle is instantaneously at rest (it stops to turn round!). But V is large and rapidly changing, giving a force F = -(dV/dx), which we've called kx, and this makes it move into the **classically allowed** region between the crossing points. (Note that here k denotes a **force constant** and has nothing to do with the *wave number* used in Section 4.1)

The classical equation of motion (force = mass times acceleration) then gives

$$\mathrm{d}x^2/\mathrm{d}t^2 = F/m = kx/m,$$

so the function x = f(t), when differentiated twice, must give (k/m)x. And, as you surely know, the functions with this property are $\sin \omega t$ and $\cos \omega t$, with $\omega^2 = k/m$. The general solution is $x = A \cos \omega t + B \sin \omega t$, where A, B are any constants – to be determined. If the particle is released from rest at x = A, when t = 0, it will oscillate between the points $x = \pm A$, which is called the **amplitude**, with 'angular frequency' $\omega = \sqrt{k/m}$ – corresponding to frequency $\nu = 2\pi\omega$ of the particle's back-and-forth motion. B must be zero in order to meet the boundary conditions.

The quantum mechanical operator corresponding to (4.13) can be written

$$\mathbf{H} = (1/2m)\mathbf{p}^2 + \frac{1}{2}k\mathbf{x}^2 \tag{4.14}$$

where the classical variables are replaced by their equivalent operators, and the eigenvalue equation to determine the allowed energy values becomes $H\Psi = E\Psi$.

In the last Section we started from Schrödinger's prescription, replacing the abstract operators x and p by x and $(\hbar/i)d/dx$, and the abstract state vector Ψ by the **function** $\Psi(x)$. This gave a new 'representation' of the operators and the things they work on – in terms of quantities we knew how to handle.

In that way the eigenvalue equation $H\Psi = E\Psi$ was replaced by a **differential equation**, like the ones we met in Books 3 and 3A, which could be solved by standard methods to get the energy eigenvalues and corresponding eigenfunctions. The last Section showed how nicely everything worked out when the differential equation was easy to solve. Now we'll do the same for the harmonic oscillator.

With the same substitutions, the eigenvalue equation $H\Psi = E\Psi$ becomes (check it out!)

$$\left(-\frac{1}{2m\hbar^2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}k\mathbf{x}^2\right)\Psi = E\Psi.$$
(4.15)

After multiplying throughout by $2m/\hbar^2$, putting $(2m/\hbar^2)E = \lambda$, and $mk/\hbar^2 = \bar{\alpha}^2$, this differential equation takes the form

$$\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} + (\lambda - \bar{\alpha}^2 x^2)\Psi = 0. \tag{4.16}$$

This looks very simple, but in fact it's rather difficult to get a general solution which will give all the eigenstates – as the next Example will show.

Example 4.5 The harmonic oscillator: Schrödinger's method

How can we find a function $\Psi = \Psi(x)$ satisfying (4.16)? All we have to do is to find a function which, when differentiated twice with respect to x, will give the result $-(\lambda - \bar{\alpha}^2)$ – so that (4.16) will become zero for all values of x. The function should be 'well-behaved' and should lead to a finite result when $|\Psi(x)|^2$ is integrated over the whole range of $x (-\infty, +\infty)$, for only in that way can it be normalized to unity. This requires that $\Psi(x) \to 0$ as $x \to \pm\infty$.

Let's start by trying the Gaussian function $\Psi = \exp -\frac{1}{2}\bar{\alpha}x^2$, which satisfies these conditions and gives an acceptable probability density $|\Psi|^2 = \exp -\bar{\alpha}x^2$, diminishing smoothly at large distances from the origin. We then have, putting $\frac{1}{2}\bar{\alpha} = a$, and $x^2 = u$

$$\frac{\mathrm{d}}{\mathrm{d}x}\exp-ax^{2} = \frac{\mathrm{d}}{\mathrm{d}u}\exp-au \times \frac{\mathrm{d}u}{\mathrm{d}x} = -a\exp-au \times 2x = -2ax\exp-ax^{2}$$

and repeating the operation (do it yourself!), to get the second derivative, gives

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\exp-ax^2 = -2a\exp-ax^2 + (2ax)(2ax)\exp-ax^2 = -2a\exp-ax^2 + (2a)^2x^2\exp-ax^2.$$

(Here you had to differentiate the product of two terms, using the result (d/dx)AB = A(d/dx)B + B(d/dx)A - go back to Book 3 if you're stuck.)

We're lucky! Because this result agrees exactly with the final term in (4.16), except for a change of sign (which gives $(\lambda - \bar{\alpha}^2 x^2)\Psi$, or $\lambda \exp -\frac{1}{2}\bar{\alpha}x^2 + (\bar{\alpha})^2 x^2 \exp -\frac{1}{2}\bar{\alpha}x^2$), provided we identify the unknown constant λ as $\lambda = 2$. So $\Psi = \exp -\frac{1}{2}\bar{\alpha}x^2$ is indeed an eigenfunction of the operator H, given in (??), with $\lambda = (2m/\hbar^2)E = 2a$. In other words the corresponding energy eigenvalue is $E = \frac{1}{2}(\alpha\hbar/m)$.

Example 4.5 showed how the energy and wave function for *one* eigenstate could be found by 'trial and error' – using a trial function that 'looks right' and choosing constants (like E and $\bar{\alpha}$ to make it fit the differential equation. But this is no good if one wants to find *all* possible quantized energy levels – often an infinite number. The first few levels, found by the method you're going to use in Example 4.6, are indicated in Figure 8b



Figure 8a Energy diagram (oscillator) Bold line shows PE function V(x)



Figure 8b –first few quantized levels Arrows indicate possible 'quantum jumps'

How can one find all the other solutions? $E = \hbar^2/m$ is in fact the *lowest allowed energy* of the harmonic oscillator, the **ground state**. Other states can be found by choosing a different form of the wave function and again adjusting it until it fits the differential equation. But this must be done systematically and the standard method is to start from the ground state function $\Psi(x)$ and multiply it by another function, v(x) say, which will allow you to modify it. So you can take

$$\Psi(x) = v(x) \exp{-\frac{1}{2}\alpha x} \tag{4.17}$$

and ask how to choose v(x). For example, choosing v(x) = cx would 'split' the Gaussian function into two halves, giving $\Psi(x)$ a form something like the function in Figure 7(b). To find v(x) more generally, the trial function (4.17) can be substituted in the Schrödinger equation (4.16), which then gives another differential equation that must be satisfied by v(x). What you find (try it if you're brave, but it's hard work!) is that v must be a solution of

$$\frac{\mathrm{d}^2 v}{\mathrm{d}x^2} - 2\alpha x \frac{\mathrm{d}v}{\mathrm{d}x} + (\lambda - \alpha)v = 0.$$
(4.18)

This doesn't look any easier to solve than the equation you started from; but equations of this form were studied long ago by the French mathematician Hermite. The solutions can be written as infinite series, but for certain values of λ these series *terminate*, giving finite polynomials (the **Hermite polynomials**). So you can get there in the end – but first you'd have to study a whole book on the "differential equations of mathematical physics"!

Instead, let's go back to first principles and start from the **operator form** of the eigenvalue equation, based on the classical function (4.13). In the next Example you'll do something clever, and very beautiful, finding the whole eigenvalue spectrum by a purely symbolic method.

Example 4.6 The harmonic oscillator: a symbolic method

The eigenvalue equation to be solved is

$$\mathsf{H}\Psi = E\Psi$$
, with $\mathsf{H} = (1/2m)(\mathsf{p}^2 + m^2\omega^2\mathsf{x}^2)$.

where Ψ is an element of an abstract vector space while the operators x, p obey the commutation relation

$$xp - px = (i\hbar)I.$$

Nothing else is needed!

It's useful to 'factorize' 2mH by introducing a new operator $\eta = (\mathbf{p} + im\omega \mathbf{x})/\sqrt{2m\hbar\omega}$ and its adjoint operator $\eta^{\dagger} = (\mathbf{p} - im\omega \mathbf{x})/\sqrt{2m\hbar\omega}$

(Look back at Chapter 2, noting that Hermitian operators like p, x are *self*-adjoint and that the constants are all real.)

The product $\eta \eta^{\dagger}$ doesn't give exactly the Hamiltonian operator H) – but it does *nearly* (that's why we defined η with that nasty denominator). In fact,

$$\hbar\omega\eta\eta^{\dagger} = (2m)^{-1}(\mathbf{p} + im\omega\mathbf{x})(\mathbf{p} - im\omega\mathbf{x})$$
$$= (2m)^{-1}[\mathbf{p}^{2} + m^{2}\omega^{2}\mathbf{x}^{2} + im\omega(\mathbf{x}\mathbf{p} - \mathbf{p}\mathbf{x})]$$

And a similar result follows for the product $\eta^{\dagger}\eta$, namely

$$\hbar \omega \eta^{\dagger} \eta = (2m)^{-1} (\mathbf{p} + im\omega \mathbf{x}) (\mathbf{p} - im\omega \mathbf{x})$$

= $(2m)^{-1} [\mathbf{p}^2 + m^2 \omega^2 \mathbf{x}^2 + im\omega (\mathbf{x}\mathbf{p} - \mathbf{p}\mathbf{x})]$

Together they give

$$\hbar \omega \eta \eta^{\dagger} = \mathsf{H} - \frac{1}{2} \hbar \omega \mathsf{I}, \qquad (A)$$

$$\hbar \omega \eta^{\dagger} \eta = \mathsf{H} + \frac{1}{2} \hbar \omega \mathsf{I}. \qquad (B)$$

On multiplying (A) and (B) by η^{\dagger} , from right and left respectively, it follows that

$$\hbar\omega\eta^{\dagger}\eta\eta^{\dagger} = \eta^{d}ag\mathsf{H} - \frac{1}{2}\hbar\omega\eta^{\dagger}, \qquad \hbar\omega\eta^{\dagger}\eta\eta^{\dagger} = \mathsf{H}\eta^{\dagger} + \frac{1}{2}\hbar\omega\eta^{\dagger}.$$

Finally, taking the difference of the two expressions, we get a commutation rule for η^{\dagger} and H:

$$\eta^{\dagger} \mathsf{H} - \mathsf{H} \eta^{\dagger} = -\hbar \omega \eta^{\dagger}. \qquad (C)$$

Suppose now there is one solution of the eigenvalue equation, with energy E_0 and eigenvector Ψ_0 . Put Ψ_0 on the right in all terms of the last equation and it then follows that $\mathsf{H}(\eta^{\dagger}\Psi_0) - \eta^{\dagger}\mathsf{H}\Psi_0 = -\hbar\omega(\eta^{\dagger}\Psi_0)$. But since $\mathsf{H}\Psi_0 = E_0\Psi_0$ this shows that

$$\mathsf{H}(\eta^{\dagger}\Psi_{0}) = (E_{0} + \hbar\omega)(\eta^{\dagger}\Psi_{0})$$

and this means that $\eta^{\dagger}\Psi_0$ is an eigenvector of H with eigenvalue $E_0 + \hbar\omega$).

In other words, η^{\dagger} is a **step-up operator** like the one you met in Section 1.4 of Chapter 1: it works on *any* eigenvector Ψ_0 to produce one with $E = E_0 + \hbar\omega$.

If you take the adjoint of every term in equation (C) – remembering that $(AB)^{\dagger} = B^{\dagger}A^{\dagger}$ – you'll find (do it!) that η is a **step-down operator**, producing a state vector with eigenvalue $E = E_0 - \hbar\omega$.

To summarise the conclusions from Example 4.6: it is possible to define step-up and stepdown operators for the harmonic oscillator. They are, respectively, η^{\dagger} and its adjoint η , where

$$\eta^{\dagger} = \frac{\mathbf{p} - im\omega\mathbf{x}}{\sqrt{2m\hbar\omega}}, \qquad \eta = \frac{\mathbf{p} + im\omega\mathbf{x}}{\sqrt{2m\hbar\omega}}, \tag{4.19}$$

which operate on any given eigenvector Ψ_0 , with eigenvalue E_0 , to produce one with $E = E_0 \pm \hbar \omega$. In this way it is possible by repetition to find *all* the allowed eigenvalues. There is, however, a *minimum* energy. This follows from the expectation value formula (1.10) $E_0 = \langle \Psi_0 | H \Psi_0 \rangle$ on using the operator η to get a state $\eta \Psi_0$ of lower energy. The lower energy would be $\langle \eta \Psi_0 | \eta \Psi_0 \rangle = \langle \Psi_0 | \eta^{\dagger} \eta \Psi_0 \rangle$ (η in the left-hand side of a scalar product being equivalent to its adjoint on the right). But in the Example we found

$$\hbar\omega\eta\eta^{\dagger} = \mathsf{H} - rac{1}{2}\hbar\omega\mathsf{I}, \quad \hbar\omega\eta^{\dagger}\eta = \mathsf{H} + rac{1}{2}\hbar\omega\mathsf{I},$$

and therefore, since the squared modulus of the vector $\eta \Psi_0$ must be positive,

$$\langle \eta \Psi_0 | \eta \Psi_0 \rangle = \langle \Psi_0 | \eta^{\dagger} \eta \Psi_0 \rangle = \langle \Psi_0 | \mathsf{H} \Psi_0 \rangle - \langle \Psi_0 | \hbar \omega \Psi_0 \rangle \ge 0.$$

It follows that $(E_0 - \frac{1}{2}\hbar\omega)\langle\Psi_0|\Psi_0\rangle$ and therefore that

$$E_0 \ge \frac{1}{2}\hbar\omega,\tag{4.20}$$

giving the lowest possible energy $\frac{1}{2}\hbar\omega$ (in agreement with Example 4.5).

Finally, remember that in Example 4.4 ω was identified with the angular frequency of a classical oscillator. Thus, $\omega \hbar = h\nu$, where $h = 2\pi\hbar$ is the 'atomic' unit of **action**, as first used by Planck, and ν is the classical oscillation frequency $\nu = \omega/2\pi$. The quantity $h\nu$ is called a **quantum** of energy.

In summary

The allowed energy values are all positive, the lowest one being $E_0 = \frac{1}{2}h\nu$; but after that they increase without limit in steps of $h\nu$.

(4.21)

The quantum mechanical oscillator is a very important system in Physics: not only does it describe in good approximation the vibrations of many types of observable systems, such as molecules and crystals (see Book 5); it plays a fundamental part in describing the **radiation field** (Book 10) due to the presence of *zero-mass* particles called **photons**. When a photon hits a 'real' oscillator the energy it carries can be absorbed, producing a **transition** from an eigenstate with initial energy E_i to a final state with energy E_f . Examples are indicated by the vertical arrows in Fig.8b: they correspond to $E_0 \to E_3$ (the long up-arrow), for **absorption**, and $E_3 \to E_1$ (the shorter down-arrow), for **emission** of radiation. We'll come back to photons in later Sections.

4.4 The 3-dimensional oscillator

It's easy to extend the methods of the last Section to the case of a particle free to move in three dimensions, the potential field being of the form

$$V = \frac{1}{2}k(x^2 + y^2 + z^2) = \frac{1}{2}kr^2, \quad r^2 = x^2 + y^2 + z^2, \tag{4.22}$$

r being the distance from a fixed attracting centre; for the Schrödinger equation is then separable (as you can see by writing out in full the x-, y- and z-dependent terms in Tand V, along with their equivalent operators). Thus

$$H = (1/2m)(p_x^2 + p_y^2 + p_z^2) + \frac{1}{2}kx^2 + \frac{1}{2}ky^2 + \frac{1}{2}kz^2$$
(4.23)

and

$$\begin{aligned} \mathsf{H} &= \left(-\frac{1}{2m\hbar^2} \right) \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + \frac{1}{2}k\mathsf{x}^2 + \frac{1}{2}k\mathsf{y}^2 + \frac{1}{2}k\mathsf{z}^2 \\ &= \mathsf{H}_x + \mathsf{H}_y + \mathsf{H}_z, \end{aligned}$$

Here H_x is the operator for a 1-dimensional oscillator, as used in (??) for oscillations in the x-direction, and H_y , H_z are defined similarly. The partial differential operator in the big parentheses is the "Laplacian" or "del-squared" operator, as used in Book 10.

The next Example shows how the corresponding eigenvalue equation $H\Psi = E\Psi$ can again be solved by using the 'separation' method.

Example 4.7 Separating the variables

As in Example 4.3, the equation $H\Psi = E\Psi$ can be 'separated' by looking for solutions of the form $\Psi = \Psi(x, y, z) = X(x)Y(y)Z(z)$. Thus, writing

$$(\mathsf{H}_x + \mathsf{H}_y + \mathsf{H}_z)X(x)Y(y)Z(z) = EX(x)Y(y)Z(z)$$

and dividing throughout by X(x)Y(y)Z(z), gives

$$\frac{\mathsf{H}_x}{X} + \frac{\mathsf{H}_y}{Y} + \frac{\mathsf{H}_z}{Z} = E.$$

The first term on the left depends only on the variable x, the second only on y and the third only on z; and their sum has to be equal to the unknown constant E. As the three terms are *independently* variable, it follows that each must be equated to a constant, and that the sum of the constants must be equal to E. Calling the three constants E_x , E_y , E_z , the first equation becomes,

$$\left(-\frac{1}{2m\hbar^2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}kx^2\right)X = E_xX$$

and the other two are similar. (Note that X is a function of only one variable, so the differentiations are now ordinary – not partial.)

From the general result (4.19) the allowed energy levels are $E = E_0, E_1, E_2, ..., E_n, ...,$ where $E_0 = \frac{1}{2}h\nu$ and in general $E_n = E_0 + nh\nu$. The frequency ν is the same (for the 'isotropic' oscillator) for all components and the total vibrational energy will thus be

$$E = \frac{3}{2}E_0 + (n_1 + n_2 + n_3)h\nu,$$

depending on the sum of all three quantum numbers.

The main result to come from this Example refers to the eigenstates of a 3-dimensional isotropic oscillator: the wave function will be a *product* of three factors and the energy will be a *sum* of terms, giving the total

$$E = \frac{3}{2}E_0 + (n_1 + n_2 + n_3)h\nu, \qquad (4.24)$$

which depends on the sum of all three quantum numbers, one for each direction in space. When the sum has the same value for two different states of vibration (e.g. for $(n_1, n_2, n_3) = (2,1,4)$ and (1,4,2)) the states are said to be **degenerate**: the degeneracy may be 'broken' by supposing that the force constants take different values for different directions, spoiling the symmetry of the isotropic oscillator for which $(k_1 = k_2 = k_3 = k)$.

4.5 The hydrogen atom ground state

The simplest chemical element (see Book 5) is **hydrogen**, whose atom consists of a single (negatively charged) **electron** moving in the central field provided by the nucleus, a (positively charged) proton. The proton is nearly 2000 times heavier than the electron and will be taken to be at rest, fixing the origin of coordinates. The classical PE of the electron in the field of the nucleus is thus (see Book 10)

$$V(r) = -\frac{e^2}{\kappa_0 r},\tag{4.25}$$

where r is the radial distance of the electron from the origin $(r^2 = x^2 + y^2 + z^2)$ and $\kappa_0 = 4\pi\epsilon_0$, where ϵ is a constant called the "permittivity of free space" (see Book 10 for *units* used in Electricity and Magnetism).

The Hamiltonian operator for the system is thus

$$\mathsf{H} = -\frac{\hbar^2}{2m}\nabla^2 - \frac{e^2}{\kappa_0 r},\tag{4.26}$$

where ∇^2 is the operator used in (4.3).

In this Section we'll seek only the ground-state solution of the Schrödinger equation $H\Psi = E\Psi$, for which the wave function depends on r alone, $\Psi = \Psi(r)$. More generally polar coordinates (r, θ, ϕ) are needed and things are much more difficult. So first of all we study the simplest case.

Example 4.8 ∇^2 for a spherically symmetrical system

Remember that a function f(r) depending only on the radial distance r is said to be "spherically symmetrical" and that the three terms in ∇^2 then give similar results, so only $d^2 f/dx^2$ will be needed. On putting $r = (x^2 + y^2 + z^2)^{\frac{1}{2}} = u^{\frac{1}{2}}$ it follows that, keeping y, z fixed to get the partial derivatives with respect to x,

$$\frac{\partial f}{\partial x} = \frac{\mathrm{d}f}{\mathrm{d}r}\frac{\partial r}{\partial x}$$
$$\frac{\partial r}{\partial x} = \frac{\partial r}{\partial u}\frac{\partial u}{\partial x} = \frac{1}{2}u^{-\frac{1}{2}} \times 2x = \frac{x}{r}.$$

Thus, remembering that f is a function of r alone,

and

$$\frac{\partial f}{\partial x} = \frac{x}{r} \frac{\mathrm{d}f}{\mathrm{d}r}.$$

Now do the same again to get $\partial^2 f / \partial x^2$, thinking of $\partial f / \partial x$ as the product ABC, with A = x, $B = r^{-1}$, C = df/dr and remembering the rule (Section 2.5 of Book 3) for differentiating a product. Thus

$$\frac{\partial}{\partial x}(ABC) = \left(\frac{\partial A}{\partial x}\right)BC + A\left(\frac{\partial B}{\partial x}\right)C + AB\left(\frac{\partial C}{\partial x}\right)$$

and on putting in the results you already have (do it!) this gives

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{r} \frac{\mathrm{d}f}{\mathrm{d}x} + x \left(\frac{-x}{r^3}\right) \frac{\mathrm{d}f}{\mathrm{d}r} + \frac{x^2}{r^2} \frac{\mathrm{d}^2 f}{\mathrm{d}r^2}.$$

If you now add the two similar terms, with x replaced by y and z in turn, the result will be (check it carefully!)

$$\nabla^2 f = \frac{2}{r} \frac{\mathrm{d}f}{\mathrm{d}r} + \frac{\mathrm{d}^2 f}{\mathrm{d}r^2}.$$

Example 4.8 has given a very simple form for ∇^2 when it operates on a function of the radial distance r alone, namely

$$\nabla^2 f \equiv \frac{2}{r} \frac{\mathrm{d}f}{\mathrm{d}r} + \frac{\mathrm{d}^2 f}{\mathrm{d}r^2}.$$
(4.27)

This can now be used to get any spherically symmetrical solution of the Schrödinger equation with Hamiltonian (4.26). The equation we have to solve determines a wave function $\Psi(r)$ and, since only the single variable r is involved, this will be an *ordinary* differential equation:

$$\mathsf{H}\Psi = -\frac{\hbar^2}{2m}\nabla^2\Psi - \frac{e^2}{\kappa_0 r}\Psi = E\Psi. \tag{4.28}$$

As we've often found wave functions of exponential form, which die away smoothly as the variable goes to infinity and give a $|\Psi|^2$ which stays finite when integrated over all space, we might look for a 'trial' function such as $\Psi(r) = A \exp -br$, where A and b are constants. (If that doesn't fit we can always allow A to vary with r and find another equation to determine it – as we did in (4.17).) On substituting this trial function in the last equation, doing the differentiations, and rearranging things a bit (do it yourself!), you'll find

$$\left[\frac{\hbar^2}{2m}b\left(b-\frac{2}{r}\right) + \frac{e^2}{\kappa_0 r} + E\right] \times A\exp(-br) = 0.$$

Now the factor in square brackets contains r-independent terms, and those with a factor 1/r. These two parts are quite independent (one stays constant, the other varies with r): and you can make the their sum vanish only by choosing the constant b so as to make each part separately zero:

$$(\hbar^2/2m)b^2 + E = 0$$
 and $(\hbar^2/m)b - e^2/\kappa_0 = 0.$

By making this choice, $b = me^2/\hbar^2 \kappa_0^2$, the function $\Psi = A \exp -br$ becomes a solution of the Schrödinger equation, with energy eigenvalue

$$E = -\left(\frac{\hbar^2}{2m}\right)b^2 = -\frac{me^4}{2\hbar^2\kappa_0^2}.$$
 (4.29)

Evidently it's not necessary to allow the factor A to vary with distance of the electron from the nucleus: it is a true constant which can easily be determined to **normalize** the wave function. The integral over all space of $|\Psi|^2$ is unity when

$$\Psi = A \exp{-br} \quad (A = b^{\frac{3}{2}} / \sqrt{\pi}, \quad b = me^2 / \hbar^2 \kappa_0^2).$$
(4.30)

(You should check the integration – remembering that the element of volume is that of a spherical shell of thickness dr.)

A note on units

The energy of the hydrogen atom ground state, given in (4.29), is often used in defining an 'atomic unit' of energy; thus $E = E_0(H - atom) = \frac{me^4}{2\hbar^2\kappa_0^2}$ may be written $E_0(H - atom) = -\frac{1}{2}e_H$, where

$$\mathbf{e}_{\mathrm{H}} = \frac{me^4}{\hbar^2 \kappa_0^2}$$

is "the Hartree unit of energy" named after one of the pioneers in the field.

You'll remember the SI units from classical physics (Book 4), where the standard energy unit is the Joule: 'atomic units' are of course very much smaller, $e_{\rm H} \approx 4.359 \times 10^{-18}$ J. A fully consistent set of units on an 'atomic' scale is obtained by taking the mass and charge of the electron (m, e) to have unit values, along with $\hbar = h/2\pi$ and $\kappa_0 = 4\pi \epsilon_0$. (Here, as already noted, ϵ_0 is the "electric permittivity of free space" (see Book 10) – but more about such things when you need to use them!)

For now, let's just note that $4\pi r^2 \Psi(r) \Psi^*(r) dr$ gives the probability of finding the electron at a distance between r and r + dr from the nucleus. This is the quantity which gives unity when integrated from r = 0 to $r = \infty$ i.e. *certainty* that the electron is *somewhere* in space: without the factor dr, it is often called the **radial probability density** and denoted by $P_r(r)$. It has a peak when $r = a_0$, where

$$a_0 = 1/b = \kappa_0 \hbar^2 / me^2. \tag{4.31}$$

This quantity has the dimensions of length (L) and represents the **most probable distance** of the electron from the nucleus: it gives a rough idea of the 'size' of the hydrogen atom and is called the "Bohr radius" (after the Danish physicist Bohr, who first used a 'semi-classical' form of quantum theory). Clearly a_0 takes the value unity when the quantities on the right of (4.31) are all expressed in atomic units. In other words a_0 provides a useful **atomic unit of length**: expressed in metres, its value is $a_0 \approx 0.5292 \times 10^{-10}$ m.

4.6 Back to the free particle

In Section 4.2 (Example 4.1) we found that a particle moving in a region where its (classical) potential energy V was constant, and could be used as the energy 'zero' V = 0, could be described by a wave function

$$\Psi(x, y, z) = A \exp i(k_x x + k_y y + k_z z) = A \exp i\mathbf{k} \cdot \mathbf{r}, \qquad (4.32)$$

where a vector notation is used (see Book 2) and k_x, k_y, k_z are the components of the wave vector k, while r denotes the **position vector** of a point on the 'wave front'.

The wave vector is a very important quantity: the plane wave (4.32) travels in the direction of k and the values of Ψ and its derivatives repeat whenever x, y, z increase by $2\pi/k_x, 2\pi/k_y, 2\pi/k_z$, respectively. As you'll remember from Section 4.2, the components of k are proportional to the components of linear momentum **p** in the three directions, $p_x = \hbar k_x, p_y = \hbar k_y, p_z = \hbar k_z$. At the same time the energy eigenvalue is

$$E_{\mathbf{k}} = \frac{\hbar^2}{2m} |\mathbf{k}|^2 = \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2).$$
(4.33)

So where in space is the particle likely to be found?

Clearly, from the basic interpretation of the wave function, $|\Psi|^2$ is the probability per unit volume of finding the particle at the point **r** where Ψ is evaluated. But here

$$|\Psi|^2 = \Psi \Psi^* = A^2 e^{i\mathbf{k}\cdot\mathbf{r}} e^{-i\mathbf{k}\cdot\mathbf{r}} = A^2$$

and the probability of finding the particle at any point r in space is therefore everywhere the same! The position of the particle is completely uncertain! With a wave function of this kind the *momentum* p is absolutely definite – it's the value we've supposed it was given – but we've no idea of the particle's *position* – it could be anywhere! This is an extreme case of Heisenberg's uncertainty principle, which we talked about in Section 3.4 in a more general way.

The problem with wave functions of this kind is that we can't normalize them to unity, because the integral of $|\Psi|^2$ over all space is infinite! There is, however, a simple trick for getting round the difficulty: if the particle is free its wave function just goes on repeating over the whole of space and we need think of only one finite 'chunk' of space, called a **fundamental volume**, within which Ψ can be normalized. The state of the particle will be described by Ψ when it is known to be in that part of space – and by an exactly similar function when it has moved into the next fundamental volume. The next Example shows the details.

Example 4.9 Periodic boundary conditions

If we look only at a finite part of space we must decide on its size and shape. The simplest choice will be a rectangular 'box', with sides of length L_x, L_y, L_z and volume $V = L_x L_y L_z$. And we then have to decide on boundary conditions, to make sure that what goes on in that 'fundamental volume' will be repeated over and over again as the box is moved through space. In other words we must impose **periodic boundary conditions**, such that when the x-coordinate of a point is increased by L_x the wave function Ψ (along with its derivatives) begins to repeat in value. Thus, we require (in the 1-dimensional case)

$$\Psi_{(x=L_x)} = \Psi_{(x=0)}, \quad \left(\frac{\mathrm{d}\Psi}{\mathrm{d}x}\right)_{x=L_x} = \left(\frac{\mathrm{d}\Psi}{\mathrm{d}x}\right)_{x=0}$$

and for a free-particle eigenfunction $\Psi(x) = \exp ikx$ this imposes the condition $k = \pm n\pi/L_x$, where n is any integer.

The same argument applies, in the 3-dimensional case, for all three directions. The final result will be, as you can easily verify,

$$\Psi_{\mathsf{k}}(\mathsf{r}) = \frac{1}{\sqrt{V}} \exp i\mathsf{k}\cdot\mathsf{r}, \quad \text{with energy } E_{\mathsf{k}} = \frac{\hbar^2}{2m} |\mathsf{k}|^2,$$

where the allowed values of the wave numbers are, for example,

$$k_x = (\pi/L_x) \times (0, \pm 1, \pm 2, \dots).$$

and similarly for the y- and z-directions.

Boundary conditions of this kind must be applied whenever we deal with a system such as a crystal, where the contents of the 'fundamental volume' are simply repeated in building up the whole crystal (see Book 5, Section 1.4). A particle described by a wave function of the form $\Psi(x, y, z) = A \exp i\mathbf{k} \cdot \mathbf{r}$ has a definite linear momentum, with components $p_x = \hbar k_x, p_y = \hbar k_y, p_z = \hbar k_z$ and travels in the direction of the k-vector. If the particle carries an electric charge there will be a resultant **electric current**. A *real* wave function, on the other hand, can only describe a particle bound within a finite volume – with zero expectation value of the momentum, and therefore zero current.

4.7 Some general properties of eigenstates

All the solutions found so far, to the 1-particle eigenvalue equation $H\Psi = E\Psi$, have certain properties in common – whether they were obtained by using Schrödinger's form of the equation, or by using purely symbolic arguments. In the first case Ψ is represented as a function of position of the particle (i.e. of independent variables such as x, y, z) and the value of $|\Psi|^2$ gives the probability, per unit volume, of finding the particle at that point. The solution Ψ is a wave *function* or eigen*function*. In the second case everything is much more abstract: Ψ is just a symbol, like the operators x, p, H which act on it, and all we need to know are the rules for combining the symbols (e.g. the 'commutation rules' that we first met in Chapter 1). Here Ψ can be visualized as a *vector*, similar to the ones used in geometry, but belonging to a more general kind of space: it is called a "state vector" or "eigen*vector*" as it is no longer a *function* in the sense of mathematical analysis – it is simply an 'element' of the space.

The two kinds of space have much in common: they are both examples of **linear vector spaces** and their elements are commonly just called "vectors". In earlier Chapters (especially Chapter 2) you've already had a taste of these things and have discovered some of the properties that make vector spaces so useful: here we list them and add some new ones (so you can refer back to them when you need to).

• The spaces are *linear* because vectors a, b, ...v, ... can be combined using the rules

 $\mathsf{a} + (\mathsf{b} + \mathsf{c}) = (\mathsf{a} + \mathsf{b}) + \mathsf{c}, \qquad x \mathsf{v} + y \mathsf{v} = (x + y) \mathsf{v}, \qquad x(\mathsf{v} + \mathsf{w}) = x \mathsf{v} + x \mathsf{w}, \qquad x(y \mathsf{v}) = x y \mathsf{v},$

where x and y are any complex numbers. (You did all that in Book 1, long ago!)

• There is a Hermitian scalar product between any two vectors such that

$$\begin{aligned} (\mathsf{a}|\mathsf{b}) &= (\mathsf{b}|\mathsf{a})^*, \qquad (\mathsf{a}|\mathsf{a}) \geq 0, \quad (\text{length}^2 \text{ positive or zero}), \\ (\mathsf{a}|x\mathsf{b}) &= x(\mathsf{a}|\mathsf{b}), \qquad (x\mathsf{a}|\mathsf{b}) = x^*(\mathsf{a}|\mathsf{b}), \end{aligned}$$

(This includes the scalar product of elementary vector theory – where all quantities are real and the star is not needed.)

• The scalar product of a sum of two vectors with a third is the sum of their separate scalar products:

$$(a + b|c) = (a|c) + (b|c).$$

• A space can be defined in terms of a set of **basis vectors**, which will usually be denoted by e_1, e_2, e_3, \dots and a general vector in the space will then be

$$\mathbf{v} = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3 + \dots$$

, where the numerical coefficients are the **components** of the vector. If there is a last basis vector \mathbf{e}_n the space is **finite** and the integer *n* is its **dimension**. The 'ordinary' space we live in is a '3-space' and any fourth vector, like \mathbf{v} is then a linear combination of the first three; the four are **linearly dependent**.

- It is usually most convenient to set up basis vectors of unit length ('normalized' so that length² = $|\mathbf{a}|^2 = (\mathbf{a}|\mathbf{a})$ is unity) and to make them perpendicular, or 'orthogonal', with $(\mathbf{a}|\mathbf{b}) = 0$. A basis of this kind is said to be **orthonormal**.
- For a *function* space the main difference is that the dimension is *infinite* and the basis vectors are not even countable they are so 'close together' that the space is 'continuous' rather than 'discrete'. This means that the methods of differential and integral **calculus** have to be used. Thus, for functions of a single variable x, the scalar product is defined as an integral:

$$(\Psi_1|\Psi_2) = \int \Psi_1^*(x)\Psi_2(x)\mathrm{d}x$$

- with the variable x playing the part of a 'counting index' k in the expression $(\mathbf{a}|\mathbf{b}) = \sum_k a_k^* b_k$, where a_k and b_k are components of two vectors \mathbf{a} and \mathbf{b} . A space of this kind is what mathematicians call a **Hilbert space**.

Some of these ideas you've already met in other books of the Series, even the early ones like Book 3, where reading Chapter 6 again will help you a lot.

In quantum mechanics, the Schrödinger wave functions are defined in a space of continuous functions of one or several real variables e.g. $\Psi = \Psi(x, y, z)$ and the basis functions are often derived as eigenfunctions of a special type of **differential equation**, of the form $L\Psi = \lambda w \Psi$ with a real 'weight factor' w. All the the equations you've studied so far are of this ("Sturm-Liouville") type, without the w-factor, and the important thing about the operator L is that it is **self-adjoint**, having the Hermitian symmetry property $(L\Psi_1|\Psi_2) = (\Psi_1|L\Psi_2)$. The nice thing about the eigenfunctions is that those with different eigenvalues are automatically orthogonal. The next Example gives a simple proof.

Example 4.10 Orthogonality of eigenfunctions

Here we'll leave out the weight factor w and use the symbolic approach, starting from the equations satisfied by any two eigenfunctions Ψ_1 and Ψ_2 . By definition

$$\mathsf{L}\Psi_1 = \lambda_1 \Psi_1, \qquad \mathsf{L}\Psi_2 = \lambda_2 \Psi_2.$$

Take the scalar product of the first equation with Ψ_2 on the left, getting

$$(\Psi_2|\mathsf{L}\Psi_1) = \lambda_1(\Psi_2|\Psi_1) \quad (a)$$

and the scalar product of the second equation with Ψ_1 on the right, getting

$$(\mathsf{L}\Psi_2|\Psi_1) = (\lambda_2\Psi_2|\Psi_1) = \lambda_2^*(\Psi_2|\Psi_1), \quad (b)$$

where the numerical factor multiplying the left-hand function is replaced by its complex conjugate in forming the scalar product. Now the operator in the first term of (b) can 'jump over' the bar if you replace it by its adjoint: but when the operator is Hermitian this makes no change – the operator is *self*-adjoint. In that case, taking the difference of (a) and (b) gives

$$(\lambda_1 - \lambda_2^*)(\Psi_2 | \Psi_1) = 0.$$

There are two conclusions:

(i) On taking $\Psi_2 = \Psi_1$ the scalar product becomes a positive number and therefore $(\lambda_1 - \lambda_1^*) = 0$. This means the eigenvalue λ_1 is *real*.

(ii) If the eigenvalues (λ_1, λ_2) are different then $(\Psi_2 | \Psi_1) = 0$ – the eigenfunctions are orthogonal.

Since Ψ_1 and Ψ_2 in the last Example are arbitrary eigenfunctions of the same Hermitian operator it follows that

(i) A Hermitian operator must have only *real* eigenvalues.

(ii) When two eigenvalues are different their eigenfunctions must be orthogonal.

(4.34)

Note that Case (ii) in (4.34) doesn't say anything at all when two eigenvalues are equal (i.e. the state is degenerate). But if $(\Psi_2|\Psi_1) = S \neq 0$ the two functions can always be *made* orthogonal by adding a bit of one to the other, which does not change the eigenvalue. Thus, leaving Ψ_1 unchanged, but re-defining the second function as $\Psi'_2 = \Psi_2 + c\Psi_1$, you can always choose c so as to make $(\Psi'_2|\Psi_1) = 0$.

(Try this as an exercise, supposing Ψ_1, Ψ_2 are normalized functions with a given value of $S = (\Psi_2 | \Psi_1)$) This procedure is called "orthogonalization" and can always be used if you know the 'overlap' (S) between the two functions. It can be applied to any number of degenerate functions and means that, in general arguments, a set of basis functions can always be *supposed* orthogonal.

4.8 But what happens to a state as time passes?

So far none of our equations contains a **time variable** t. We've been talking about 'wave functions', obtained as solutions of Schrödinger's **wave equation**, but we all know that waves don't just stand still! They usually move through space in one direction or another, like the waves on the sea that break as they come in towards the shore, or the ripples on a pond that move outward in circles when you drop a pebble into it. The trouble with Schrödinger's equation, as we've written it so far, is that it doesn't contain the *time* and can therefore lead only to **stationary states** which don't change in any way as time passes. Even though Ψ is not in any sense *real*, and only $|\Psi|^2$ can give any information that can be tested experimentally, it must allow us to describe how a system can develop in time. So something is missing – and to put it in we'll go back to the beginnings of quantum theory.

In the early part of the last century there was a great argument about the difference between material particles, which could carry mass and electric charge, and waves – which seemed to have no mass and carried only **energy**. On the one hand, although it was accepted that radiation consisted of 'waves' in the electromagnetic field (see Book 10, Chapter 6), there was experimental evidence that it was carried by zero-mass 'particles' which were given the name **"photons"**: the photon had an energy $\epsilon = h\nu$, depending on the radiation *frequency* ν , which could be absorbed by material objects only in amounts of this size (i.e. one 'quantum' at a time). But on the other hand the wave nature of radiation was well known from optical experiments: for example, light passing through a tiny hole didn't make just one spot on a photographic plate but rather a central spot surrounded by **diffraction rings** – like the ripples on a pond that surround the pebble you dropped into it. The photon showed both wave and particle properties.

Similarly a 'real' particle of matter – such as an electron – sometimes behaved as if it had wavelike properties: **electron diffraction**, when a beam of electrons went through a small hole and fell on a photographic plate, was also experimentally well known.

So was an electron a particle or a wave? And was a photon a wave or a particle? The conclusion can only be that both are **quantum particles**. The co-existence of wave and particle properties in the same 'object' is called "**duality**". The main task in the early days of quantum theory was to understand this duality and Schrödinger's wave function Ψ was to play a central role. The breakthrough came when **time** was included in the wave function.

Chapter 5

The bridge between 'classical' and 'quantum' physics

5.1 How can we bring in the time?

In classical physics we started from ideas about **motion** and motion depends on the idea of **time**: but in developing quantum mechanics we started from the measurement of observable quantities, such as particle positions and velocities, supposing only that they could be assigned observed values. We've come a long way in the belief that the quantities we met in classical physics, such as momentum, kinetic energy, electric field strength, and so on, could be defined and used in similar ways – even for systems too small to be observed directly in the laboratory. A new calculus had to be invented – quantum mechanics – which could be used when classical methods no longer applied.

But we expect some kind of 'smooth transition' between the 'classical' and 'quantum' regimes – not a sudden break where all our ideas have to be thrown away and replaced by new ones. We need a **correspondence principle** to provide a bridge between the two; and to find it we must first introduce the time.

If we want to have a 'picture' of a moving $\Psi(x)$ how do we get one? We can plot the values of the function (whatever they may stand for) at points on the x-axis, if that's the direction the wave is going in. You've seen pictures like that before, in Book 10, and you may want to look there again at Figure 39 in Chapter 6 on waves. That Figure showed a 'wave profile' in two positions: first, where it started from; and, second, where it got to at time t.

Here we'll show a similar picture (in Figure 9, below), where each wave profile has been put in its own 'picture frame'. The profile shown in Frame 1 shows what the wave looked like when it set off at t = 0. The one shown in Frame 2 shows the wave profile at time t. It is exactly the same function of x', we are supposing, where the distance x' is measured from the origin O' at the bottom left-hand corner of Frame 2; so $\phi = f(x')$. But x' is related to x measured from the origin O, fixed in space, by x' = x - ut as you can see from Fig.9. It follows that for the wave moving with velocity u

$$\Psi = f(x - ut),\tag{5.1}$$

which is referred to the fixed origin and now includes the time. All you had to do was to replace the variable x by x - ut, where u is the velocity with which the wave moves to the right and ut is the distance gone; and if you change u to -u you'll get a wave profile moving to the left, along the negative x-axis. In both cases the profile of Ψ values moves without change of shape.



Figure 9. Travelling wave profile.

Before moving on, we should be sure what we *mean* when we talk about a "travelling" wave: what is actually moving? The 'pattern' of Ψ values in Fig.9 is surely going to the right, but if you want to be more precise: the velocity u is really the **phase velocity** and is measured by the rate at which the 'argument' of the wave function, for example the variable x - ut in (5.1) is changing with time. This will be clearer when we look at an example we've already studied – that of a wave function representing a state of definite momentum, given in (4.5) as $\Psi(x) = A \exp ikx$. This is still a *stationary* state because it was obtained by solving an eigenvalue equation which didn't contain a time variable t: nothing is changing as time passes.

But now we know how to make a wave *move* we can easily introduce the time, simply by changing the variable x to x - ut, to make it move to the right, or to x + ut, to make it go to the left. So let's start from the momentum eigenstate of definite $p_x = k\hbar$ with wave function (4.5) and make it move to the right with velocity u. To do that, we simply replace x by x - ut and find the function of the two variables, x and t:

$$\Psi(x,t) = A \exp ik(x-ut), \tag{5.2}$$

where k is the wave number $k = p_x/\hbar$. When x increases by $2\pi/k$, for any value of t, the form of Ψ is repeated: $\lambda = 2\pi/k$ is the wavelength. And, as you know, such a wave function can also be expressed in terms of sine and cosine functions (since $\exp i\theta = \cos \theta + i \sin \theta$ for any 'angle' θ (look back at Chapter 4 if you need to). Here it's convenient to use the real form instead of (5.2):

$$\Psi(x,t) = A\cos k(x-ut). \tag{5.3}$$

It's also convenient to introduce two more time-dependent concepts – the **period** and the **frequency** – which are important whenever we deal with oscillations of any kind. You may have met them in Book 10 (Chapter 6). Just as the wavelength λ is the *length* of one complete up-down 'wiggle', the period τ is the amount of *time* that each wiggle takes to pass by. If you think of the waves as coming out from a starting point at a rate of ν per second, each one with a wavelength λ , then you'll see that the distance covered by the first wavefront in one second must be frequency × wavelength – and this will be the **velocity of propagation**, or the "phase velocity" of the wave: in symbols

$$\nu \times \lambda = u. \tag{5.4}$$

This important result allows us to write (5.3) in various forms – all saying exactly the same thing! The simplest follows from (5.3) on noting that the argument k(x - ut) may be rewritten as

$$k(x - ut) = kx - ku = kx - (2\pi/\lambda)(ut) = kx - (2\pi/\lambda)(\lambda/\nu) = (kx - \omega t),$$

where $\omega = 2\pi\nu$ is the **angular frequency**. The simplest form is thus

$$\Psi(x,t) = A\cos(kx - \omega t) \tag{5.5}$$

and is the form we'll normally use. But another form, in terms of wavelength λ and period τ , is

$$\Psi(x,t) = A\cos 2\pi \left(\frac{x}{\lambda} - \frac{t}{\tau}\right).$$
(5.6)

Here, the distance travelled is measured in wavelengths, while the time taken is measured in periods. (Note that the the wavelength λ is $2\pi/k$; but the period τ is $1/\nu$ (without the 2π . Also, in both cases, the cosine forms may be replaced by complex exponentials e.g. (5.5) is the real part of $A \exp i(kx - \omega t)$.

Combining waves of different wavelength and frequency

In Chapter 4 (Example 4.2) we found **standing wave** solutions of Schrödinger's equation for a particle in a box. Although these had a wave-like ('sinusoidal') form they were certainly not *moving*: they were real functions of position in space and they didn't depend on time. But they could be written as combinations of different momentum eigenfunctions, corresponding to equal and opposite values of the linear momentum. For a particle moving along the x-axis, a state of definite energy $p_x^2/2m$ was found to be a mixture of different eigenstates of momentum $\pm p_x$; and this was interpreted pictorially as a state in which the particle was bouncing backwards and forwards, with equal probability, between the walls of the container. The possibility of combining different eigenfunctions, to obtain new wave functions with other properties, arises from the **linearity** of the eigenvalue equation (if Ψ_1 and Ψ_2 are solutions of the equation, then so is $c_1\Psi_1 + c_2\Psi_2$, where the coefficients are arbitrary consstants). The following two Examples show how this can help us to understand travelling waves and time-dependent wave functions – and even the co-existence of wave and particle properties in the same system!

Example 5.1 Combining two waves

Let's combine, or 'superimpose', two waves with the same amplitude but slightly different wavelength and frequency, namely

$$\Psi_1 = A\cos(k_1x - \omega_1 t), \qquad \Psi_2 = A\cos(k_2x - \omega_2 t).$$

Addition gives a resultant which can be written (using the formula $\cos a + \cos b = \cos \frac{1}{2}(a+b) \times \cos \frac{1}{2}(a-b)$, which is easily proved along the lines indicated at the end of Chapter 4, Book 2) as

$$\Psi = \psi_1 + \Psi_2 = 2A\cos\left(\frac{k_1 + k_2}{2}x - \frac{\omega_1 + \omega_2}{2}t\right) \times \cos\left(\frac{k_1 - k_2}{2}x - \frac{\omega_1 - \omega_2}{2}t\right).$$

Each of the two factors has a wave-like form, but the first one has an *average* wave number $\frac{1}{2}(k_1 + k_2)$ and an average frequency $\frac{1}{2}(\omega_1 + \omega_2)$; while the second contains corresponding *differences*, $\frac{1}{2}(k_1 - k_2)$ and $\frac{1}{2}(\omega_1 - \omega_2)$. When the original waves have nearly the same wavelength and frequency the first factor in the resultant is not much different; but the second factor varies much more slowly and its main effect is just to 'modulate' the wave by changing its amplitude in a slowly varying way.

What happens when you combine two waves in this way is shown pictorially in Figure 10 (below).



Figure 10. Travelling profile: a series of 'pulses'

The resultant wave has a profile, containing a series of 'pulses', which moves much more slowly than the rapidly oscillating waves within each pulse. From (5.4) the profile will move with a speed (frequency×wavelength) $\nu \lambda = 2\pi\nu/k$, with $\nu = \nu_1 - \nu_2$ and $k = k_1 - k_2$. This is called the **group velocity** and is the speed with which each 'group', or 'packet', of waves inside the profile travels through space. Here there are many packets, not just one – which, by itself, could be described as fairly 'localized'. That's because we combined only two wave trains. But we have an infinite number of eigenfunctions to choose from; and they can be put together with any coefficients we wish. So let's choose one 'main component', with wave number k_0 , and combine it with all the other wave trains whose wave numbers k are close to k_0 , hoping to get a single sharply defined packet.
5.2 The Gaussian wave packet

In Example 5.1 the mixture of wave trains had the form

$$\Psi(x,t) = \sum_{k} A_k \exp i(kx - \omega t),$$

with only two terms in the sum. But now let's consider an *infinite* sum, where A_k is replaced by a continuous function a(k) and the summation turns into an integration over all values of k in the interval $(-\infty, +\infty)$.

In the next Example we'll choose an amplitude function of the 'Gaussian' form that you first met in Book 5, namely

$$a(k) = A \exp -\sigma (k - k_0)^2, \tag{5.7}$$

and define a **Gaussian wave packet** based on (5.7). Using the complex exponential form,

$$\Psi(x,t) = \int_{-\infty}^{+\infty} A \exp -\sigma (k-k_0)^2 \exp i(kx - \omega t) \,\mathrm{d}k.$$
(5.8)

where k_0 is the wave number defining the 'centre' of the packet and the integration runs over all other values.

Example 5.2 The Gaussian packet

The integral (5.8) defining the packet is difficult to evaluate, but in any list of standard definite integrals you can find something quite similar, namely

$$\int_{-\infty}^{+\infty} \exp(-ps^2 - qs) \, \mathrm{d}s = \sqrt{\frac{\pi}{p}} \exp\left(\frac{q^2}{4p}\right),$$

and this is valid for all values of the constants p and q, real or complex – as long as p has a positive real part. (Note that the integration variable has been called s, instead of the usual x, as we already have an x meaning the x-coordinate in the wave function.)

Suppose the packet sets off at time t = 0. Then the zero-time form of the wave function will be

$$\Psi(x,0) = \int_{-\infty}^{+\infty} [A \exp -\sigma(k-k_0)^2 \times \exp ikx] \, \mathrm{d}k$$

If we set $k - k_0 = s$, a new integration variable, we can put $k = k_0 + s$ and write the expression for $\Psi(x, 0)$ as

$$\Psi(x,0) = \int_{-\infty}^{+\infty} A \exp[-\sigma s^2 + xi(k_0 + s)] \, \mathrm{d}s.$$

Here we've used the fact that $e^a \times e^b = e^{a+b}$ and that dk = ds. (If you need help on 'changing the variable' you can get it in Book 3, towards the end of Chapter 3.)

Now the integrand contains a factor $\exp ixk_0$, which is a constant – not depending on the integration variable s. So let's take it outside the integration, leaving us with $\Psi(x,0) = A \exp ixk_0 \times I$, where the integral I is given by

$$I = \int_{-\infty}^{+\infty} \exp[-\sigma s^2 + ixs] \,\mathrm{d}s.$$

But if you put $\sigma = p$ and ix = q this becomes the standard integral we started from, with the value

$$I = \sqrt{\frac{\pi}{p}} \exp\left(\frac{q^2}{4p}\right) = \sqrt{\frac{\pi}{\sigma}} \exp\left(\frac{-x^2}{\sigma}\right).$$

That's the end of the story for the wave packet when it sets off: at t = 0 it has the form

$$\Psi(x,0) = A\sqrt{\pi/\sigma} \exp ik_0 x \exp \frac{-x^2}{4\sigma}.$$

Example 5.2 has shown us that a *single* and quite strongly *localized* wave packet can be set up by combining an infinite number of wave trains whose k-vectors are all close to the one with $k = k_0$, provided their amplitudes fall rapidly towards zero as you go away from $k = k_0$. In fact we chose a Gaussian amplitude function (5.7) and found a wave packet

$$\Psi(x,0) = A\sqrt{\pi/\sigma} \exp(ik_0 x) \exp(-(x^2/4\sigma)).$$
(5.9)

Before it begins to move, this function has the form shown on the left in Figure 11, centred around the point x = 0.



Figure 11. A Gaussian wave packet, moving from x = 0 to $x = u_g t$.

In the Figure, the packet at x = 0 has been shown as a grey 'shadow' (with a broken line indicating the Gaussian profile) – because that's only where it *started* from at t = 0. At time t it's no longer there, having moved a distance $x = u_g t$ where u_g is the **group velocity**. But now there's only *one* packet and we don't yet know how fast it travels, though Example 5.1 suggests that u_g may be very different from the phase velocity u. How can we find the group velocity?

Any kind of wave depends on the **frequency** with which something goes up and down or 'oscillates' – whether it's the level of the water in a pond, or the strength of an electric field in a radio wave – and it's the frequency that brings in the idea of time and relates it to the the velocity with which something is changing. Now, we're talking about waves $\Psi(x,t)$ which don't give us anything to measure with our instruments; but we already know that wavelength and frequency are useful and important ideas when we're thinking about the motion of particles in quantum mechanics:

- For a particle moving with given energy E and momentum p along the x-axis, the wavelength of the associated Ψ is related to the momentum by $\lambda = h/p$, where h is Planck's constant, or by $\lambda = 2\pi\hbar/p$ in terms of the 'rationalized' Planck constant
- A particle oscillating at frequency ν can exchange energy with a field of radiation, of the same frequency ν , but only in integral multiples of a 'quantized' amount $\epsilon = h\nu$.

The proportionality constant h has the same observed value in each case! Could it be that these two experimentally based observations hold more generally?

Let's suppose they do and that we can apply them to a particle moving along the x-axis in a quantum state of definite energy E and definite momentum p (dropping the subscript 'x' when it's not needed). As we discovered in Example 5.1, p and E are then related just as in classical physics: $E = (1/2m)p^2$. So we should be able to use them in the wave function Ψ , taking the wavelength to be $\lambda = 2\pi\hbar/p$ and the 'frequency' (whatever it may mean!) to be $\nu = E/2\pi\hbar$. That way it seems likely that the phase velocity will be

$$u = \lambda \nu = \frac{2\pi h}{p} \frac{E}{2\pi h} = \frac{E}{\sqrt{2m(E-V)}},$$
 (5.10)

where E is in general the particle energy $E = (1/2m)p^2 + V$, while p arises from the kinetic energy term alone (which gives $p^2/2m = E - V$. When the particle moves into a region where its potential energy is varying, the formula for the phase velocity of the associated wave Ψ will still be (5.10). But this velocity will now depend also on particle position (through the PE function V and the resultant variation of λ).

If the wave packet contained only one component, with $k = k_0$ say, and frequency $\nu = \nu_0$, Ψ would have the form

$$\Psi(x,t) = A \exp[i(k_0 x - 2\pi\nu_0 t)],$$

which follows from the zero-time function $\Psi(x, 0)$ on replacing x by $x - u_0 t$ and remembering that $k_0 u_0 = \nu_0$. This would give the same function, but moved to the right through a distance $u_0 t$ without change of shape.

In Example 5.2, however, we replaced the amplitude A, by a Gaussian distribution of k-dependent amplitude factors, clustering around the one with $k = k_0$, and integrated over all k-values to get the packet (5.8). But now we have to do the same thing when t is no longer zero; and this is much more difficult because the frequency and the phase velocity will now vary as we go from one component (k) to the next. This phenomenon, in which the phase velocity of different wave trains in the packet depends on frequency, is called **dispersion**. It determines how the wave packet travels – both its speed and its shape. Here we'll just indicate the argument to show how it goes – you don't need to follow the details, but they will be there if ever you want to come back!

In Book 3 you learnt about Taylor's theorem (Section 5.3) and saw how you could expand any (wellbehaved) function f(x) around any chosen point $x = x_0$. Thus,

$$f(x) = f(x_0) + \alpha(x - x_0) + \frac{1}{2}\beta(x - x_0)^2 + \dots$$

where the coefficients are first, second, and higher derivatives of the function, all evaluated at the point $x = x_0$. This is just what is needed: we can write

$$\nu = \nu_0 + \alpha (k - k_0) + \frac{1}{2}\beta (k - k_0)^2 + \dots,$$

where

$$\nu_0 = \nu(k = k_0), \quad \alpha = (\mathrm{d}\nu/\mathrm{d}k)_0, \quad \beta = (\mathrm{d}^2\nu/\mathrm{d}k^2)_0, \dots$$

and then go ahead to evaluate that difficult integral, which now includes the time.

The calculation follows the same lines as in Example 5.2, using the same standard definite integral. On including only the first two terms in the Taylor expansion, the result is

$$\Psi(x,t) = A\sqrt{\frac{\pi}{\sigma}} \exp\left(\frac{-\pi^2(x-\alpha t)^2}{4\sigma}\right) \exp[i(k_0x - 2\pi\nu_0 t)].$$

Note that σ controls the width of the packet (the smaller it is, the 'tighter' the packet), while α plays the part of a velocity (remembering that changing x to x - vt displaces any function through a distance vt along the positive x-axis. The 'central' wave train moves with the phase velocity $u_0 = \nu_0/k_0$, but the packet itself moves as a whole, with the **group velocity** $u_g = \alpha$.

To find u_g we need only remember the Taylor expansion of frequency (ν) as a function of wave number (k), in which $\alpha = d\nu/dk$. The first derivative of $\nu = E/h$ is now a function of wave number because

$$E = (KE + PE) = (1/2m)p^2 + V = (1/2m)\hbar^2k^2 + V.$$

Consequently, on evaluating $u_g = \alpha$, we find

$$u_g = 2\pi \frac{\mathrm{d}\nu}{\mathrm{d}k} = \frac{2\pi}{h} \frac{\mathrm{d}E}{\mathrm{d}k} = \frac{1}{\hbar} \frac{\mathrm{d}}{\mathrm{d}k} \left(\frac{\hbar^2 k^2}{2m} + V\right) = \frac{\hbar k}{m} = v, \qquad (5.11)$$

since V does not depend on particle velocity (v), while $\hbar k = p = mv$.

The remarkable result is that the group velocity of the wave packet is the same as the velocity of the moving particle it describes.

This is just what you'd expect on grounds of 'common sense': it would be nonsense if the particle went one way while the wave packet describing it went another; and if their velocities didn't match exactly they would eventually find themselves an infinite distance apart, not connected in any way!

5.3 Schrödinger's equation including the time

The time variable t was first mentioned in Section 5.1, long after the introduction of the wave function Ψ . That was because we'd studied only the properties of **stationary states**, which were eigenfunctions of operators that were *time independent*, involving only dynamical variables like components of particle position and momentum. We were looking for states in which the energy E had a definite and unchanging value; and in Schrödinger's representation of quantum mechanics (Chapter 4) these were solutions of a partial differential equation which didn't include the time. In Example 4.1 we found an energy eigenfunction $\Psi = \exp ikx$ for a particle free to move along the x-axis – a solution of the "Schrödinger equation without the time", namely

$$\mathbf{H}\Psi = E\Psi$$
, with $\mathbf{H} = -(\hbar^2/2m)(\mathrm{d}^2\Psi/\mathrm{d}x^2)$.

However, the same Ψ was also an eigenfunction of the momentum operator \mathbf{p}_x , with eigenvalue $k\hbar$: $\mathbf{p}_x\Psi = k\hbar\Psi$ where k can take any real value, positive or negative. The energy is $E = p_x^2/2m$ and clearly doesn't depend on the ±sign of k; but the value of $p_x \ does -$ if k is positive, the momentum component corresponds (in classical language) to the particle moving to the right, with a velocity $v_x = p_x/m = k\hbar/m$. Of course, classically, v_x is a time derivative $v_x = dx/dt$. So the question is how to get the time into the equations of quantum mechanics.

At the beginning of the present Chapter 5, we found a way of getting time into the wave function Ψ by using what we know about waves in general: a wave with some 'profile' $\Psi(x)$ can be displaced through distance d along the x-axis simply by changing x to x - d. And if u is the rate of displacement, then d = ut is what you get after time t. A sinusoidal wave, of amplitude A and with profile $\Psi(x) = A \exp ikx$, thus gives the time-dependent wave function (5.2). With the definitions of frequency, wavelength and phase velocity, given in Section 5.1, we get the starting point we need. The wave function for a free particle moving along the x-axis with momentum $k\hbar$ is

$$\Psi(x,t) = A \exp i(kx - 2\pi\nu t) \tag{5.12}$$

- a function of both position x and time t.

Now the frequency ν has been related to the particle energy E through Planck's rule $E = h\nu$ and if the exponential in (5.13) is written as a product of two factors it becomes $\Psi(x,t) = \Psi(x) \times f(t)$, where $\Psi(x)$ is a time independent eigenfunction of the Hamiltonian operator H, with eigenvalue E, while

$$f(t) = \exp -i(2\pi E/h)t = \exp -i(E/\hbar)t$$
(5.13)

and is a function of time alone. It is simply a **phase factor**, a complex number of unit modulus. All physically measurable quantities, such as the probability density $|\Psi|^2 = \Psi^* \Psi$, contain two Ψ -factors, $\Psi(x,t)$ and $\Psi^*(x,t)$ and are therefore real and time-independent, since $f^*(t)f(t) = 1$.

So why do we need to include the time at all? The answer should be clear by now: without it there would be no possibility of finding travelling waves, or of setting up moving wave packets, which show how wave and particle properties can co-exist in the same system; or of dealing with non-stationary states, which change as time goes on.

The next step is to find a time-dependent form of Schrödinger's equation. Let's use again the example of the momentum eigenfunction $\Psi(x,t) = \Psi(x) \times f(t)$, which is also an eigenfunction of the Hamiltonian: $H\Psi(x,t) = E\Psi(x,t)$, since f(t) is just a multiplicative factor. On the other hand, differentiating f(t) with respect to t simply multiplies it by $-i(E/\hbar)$. The time-dependent wave function must therefore satisfy the two equations

$$\mathsf{H}\Psi(x,t) = E\Psi(x,t), \qquad \frac{\partial}{\partial t}\Psi(x,t) = -i\frac{E}{\hbar}\Psi(x,t).$$

On substituting $E\Psi(x,t) = \mathsf{H}\Psi(x,t)$ from the first equation into the second, and multiplying the result by $i\hbar$, there follows Schrödinger's equation including the time:

$$i\hbar\frac{\partial\Psi}{\partial t} = \mathsf{H}\Psi$$
(5.14)

Of course nothing has been *proved*, because we considered only a single 'point' particle, moving along the x-axis with a definite momentum and energy. But the equation deserves a 'box' because it does in fact hold generally, for any kind of quantum system as long as we can set up a Hamiltonian to describe it. It's just as important in quantum physics as Newton's laws of motion were in classical physics; and the rest of this book, together with the applications to many-electron atoms and molecules in Book 12, will give a small taste of what we can get from it. First, however, we need to look for the quantities that will *correspond* in some way to the things like linear momentum and force – which were supposed to have definite measurable values in classical theory. (If you need to be reminded of classical physics go back to Book 4.)

5.4 What makes a wave packet move?

Classically, **force** is what make things move i.e. change position (or state of motion). If you look at the wave packet in Figure 11 there is no precise value of x that will indicate its position: the best you can do is to give the **expectation value** $\langle x \rangle = \langle \Psi | \mathbf{x} \Psi \rangle$, which may give a good approximation to the packet's position when it is confined within a very small region. This quantity is a function of time and moves with the group velocity (5.11). Similarly, the particle momentum will have an expectation value

$$\langle p_x \rangle = \langle \Psi | \mathsf{p}_x \Psi \rangle. \tag{5.15}$$

Lastly, the corresponding force component F_x is defined classically as the rate of decrease of the particle's potential energy when $x \to x + dx$. Thus,

$$F_x = -\frac{\partial V}{\partial x}$$

where the PE function V = V(x, y, z) depends on the system considered (Examples were given in Chapter 4). The corresponding quantum mechanical operator for a force component is obtained in the usual way from Schrödinger's recipe (3.16) and the expectation value of the component follows as

$$\langle F_x \rangle = \langle \Psi | \mathsf{F}_x \Psi \rangle. \tag{5.16}$$

According to Newton's second law, a particle of mass m moves under the action of a force F with acceleration a, given by ma = F. Since a is the rate of increase of velocity v and

 $m\mathbf{v} = \mathbf{p}$ this is equivalent to saying

$$(\mathrm{d}p_x/\mathrm{d}t) = F_x$$

and similarly for the y- and z-components of the vectors. (We'll usually consider only the x-component.)

What we'd like to show now is that, in the 'quantum limit' where classical physics breaks down, there will be a similar relationship between the *expectation values* of the quantities involved. In other words, we want to show that the expectation value $\langle F_x \rangle$ of the force component F_x will give the rate of change of the corresponding momentum expectation value:

$$\langle F_x \rangle = \langle \Psi | \mathsf{F}_x \Psi \rangle = \frac{\mathrm{d}}{\mathrm{d}t} \langle \Psi | \mathsf{p}_x \Psi \rangle.$$
 (5.17)

In order to deal with rates of change of expectation values and operators we need a basic result, which follows from the time-dependent Schrödinger equation (5.14). For a general observable A, with operator A, $\langle A \rangle = \langle \Psi | A \Psi \rangle$, which is a scalar product with Ψ^* on the left and $A\Psi$ on the right. Very often the three factors are separated more clearly by putting in an extra vertical bar to give

$$\langle A \rangle = \langle \Psi | \mathsf{A} | \Psi \rangle$$

and from now on we'll use this notation. The rate of change can now be written

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle A\rangle = \langle \partial \Psi / \partial t | \mathsf{A} | \Psi \rangle + \langle \Psi | \mathsf{A} | \partial \Psi / \partial t \rangle + \langle \Psi | \partial \mathsf{A} / \partial t | \Psi \rangle,$$

where you'll remember that the Ψ on the left in any scalar product is the complex conjugate of the wave function Ψ . The last term is included because the general observable Amay itself depend on the time.

From (5.14) we get

$$\frac{\partial\Psi}{\partial t} = (1/i\hbar)\mathsf{H}\Psi, \qquad \frac{\partial\Psi^*}{\partial t} = -(1/i\hbar)\mathsf{H}\Psi^*$$

and substitution in the last equation then gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle A\rangle = -\frac{1}{i\hbar}\langle \Psi |\mathsf{H}\mathsf{A}|\Psi\rangle + \frac{1}{i\hbar}\langle \Psi |\mathsf{A}\mathsf{H}|\Psi\rangle + \langle |\Psi\partial\mathsf{A}/\partial t|\Psi\rangle.$$

In terms of the commutator [A, H] = AH - HA, this becomes

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \langle A \rangle = \langle \Psi | [\mathsf{A}, \mathsf{H}] | \Psi \rangle + i\hbar \langle \Psi | \partial \mathsf{A} / \partial t | \Psi \rangle.$$
(5.18)

Before we can use this result we need two special cases: the rates of change of expectation values $\langle x \rangle$ and $\langle p_x \rangle$ of the basic position and momentum operators.

The operators have the basic commutation property $px - xp = (\hbar/i)I$, where the subscript on p has been dropped to make things clearer. (Note also that $\hbar/i = -i\hbar$, so you can write instead $px - xp = -i\hbar$.) We're going to need the special cases of (5.18) for A = x and $A = p_x$. Both operators are time independent, so the last term vanishes; but the expectation value of the commutator does not – and it contains H, which is generally a function (e.g. a polynomial) of x, p. Let's take the case A = x first and look for a commutator of p and x^n – to use in a polynomial.

From $px - xp = -i\hbar l$ we can multiply from the right by x, to get $px^2 - xpx - \hbar x l = 0$ and from the left to get $xpx - x^2p + i\hbar x l = 0$. Addition then gives $px^2 - x^2p + 2i\hbar x l = 0$. So we have *two* results, which can be written both together as

$$\mathbf{p}\mathbf{x}^n - \mathbf{x}^n \mathbf{p} + in\hbar \mathbf{x}^{n-1} = 0, \quad (A)$$

with n = 1 and n = 2, respectively. Could it be that the general result (A) is good for any value of the integer n?

Let's suppose that it's true for one particular value of n and then ask if it remains true when we change n to n + 1. If we can show that it does, then the result will be true for *all* integers. This way of proving something is called the "method of **mathematical induction**".

Write out (A), multiplying first by \times from the right to get

$$px^{n}x - x^{n}px + in\hbar x^{n} = 0$$

and then from the left, getting

$$\mathsf{x}\mathsf{p}\mathsf{x}^n - \mathsf{x}^{n+1}\mathsf{p}\mathsf{x} + in\hbar\mathsf{x}^n = 0.$$

By adding the two, we get

$$px^{n+1} - x^{n+1}px + xpx^n - x^n px + 2in\hbar x^n = 0.$$

This doesn't look very promising, but remember the commutation rule $px - xp = -i\hbar I$, which allows us to change the order of the operators by writing $xp = px + i\hbar I$ and $px = xp - i\hbar I$. On using these substitutions in the last equation (and leaving out the identity I as it's not really needed) we get

$$\mathbf{p}\mathbf{x}^{n+1} - \mathbf{x}^{n+1}\mathbf{p} + (\mathbf{p}\mathbf{x} - \hbar)\mathbf{x}^n - \mathbf{x}^n(\mathbf{x}\mathbf{p} + \hbar) + 2in\hbar\mathbf{x}^n = 0.$$

And when you expand the terms containing the parentheses and rearrange what's left (do it!) you'll get the final result

$$2 \times [px^{n+1} - x^{n+1}p] + 2i\hbar(n+1)x^n = 0.$$

On taking out the factor of 2, it follows that

$$px^{n+1} - x^{n+1}p = -i\hbar(n+1)x^n$$
. (B)

We're through! Equation (B) is just (A), but with n replaced by n + 1.

Example 5.3 has shown that the basic commutation law for the position and momentum operators, $\mathbf{p}_x \mathbf{x} - \mathbf{x} \mathbf{p}_x = (\hbar/i)\mathbf{I}$, leads to the general commutator (putting the x-subscript back on the \mathbf{p})

$$\mathbf{p}_x \mathbf{x}^n - \mathbf{x}^n \mathbf{p}_x = n(\hbar/i) \mathbf{x}^{n-1}, \tag{5.19}$$

valid for all positive integer values of n. The commutators for the other conjugate pairs of operators follow simply by changing the subscript x (labelling the axis) to y or z.

Now (5.19) can be written, with the usual notation $\mathbf{p}_x \mathbf{x}^n - \mathbf{x}^n \mathbf{p}_x = [\mathbf{p}_x, \mathbf{x}^n]$, as

$$[\mathbf{p}_x, \mathbf{x}^n] = n(\hbar/i)\mathbf{x}^{n-1} = (\hbar/i)(\partial/\partial \mathbf{x})\mathbf{x}^n.$$

In other words, taking the commutator of p_x with x^n is the same as differentiating with respect to x and multiplying by \hbar . This will be true, term by term, for any polynomial in x (the operators being linear); and for any convergent function f(x) it follows that

$$[\mathbf{p}_x, f(\mathbf{x})] = (\hbar/i)f'(\mathbf{x}) \tag{5.20}$$

– where the prime indicates, as usual, the derivative of the function with respect to the variable x.

A second result follows without any extra work!

(When Schrödinger's representation of the operators was first introduced, in Section 4.1, it was noted that a different representation of the abstract operators x and p_x could be obtained by making the association

$$\mathbf{p}_x \to p_x, \qquad \mathbf{x} \to -(\hbar/i)(\partial/\partial p_x),$$

with the operators working on a function $\Phi(p_x, p_y, p_z)$ of the *momentum* components. This amounts to interchanging the operators x, p_x in the original commutator and changing the sign.)

Here we'll just note the result, which is a 'partner' to (5.20): it is

$$[\mathsf{x}, f(\mathsf{p}_x)] = -(\hbar/i)f'(\mathsf{p}_x) \tag{5.21}$$

– with similar results for the y- and z-components.

We're now ready to go back to the proof of (5.17), starting from the general formula (5.18) with A = x and $A = p_x$ in turn. Both are time-independent operators, so

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \langle \mathbf{x} \rangle = \langle [\mathbf{x}, \mathbf{H}] \rangle \qquad (A)$$

and

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \langle \mathbf{p}_x \rangle = \langle [\mathbf{p}_x, \mathbf{H}] \rangle \qquad (B)$$

Now let's use (5.20) and (5.21), supposing that $H = f(x, p_x)$, a function of both position and momentum operators.

From (5.21) we get

$$[\mathbf{x}, \mathbf{H}] = i\hbar(\partial \mathbf{H}/\partial \mathbf{p}_x) \quad (a),$$

while from (5.20) it follows that

$$[\mathbf{p}_x, \mathbf{H}] = -i\hbar(\partial \mathbf{H}/\partial \mathbf{x}) \quad (b).$$

On putting (a) into (A), the rate of change of $\langle x \rangle$ becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathbf{x}\rangle = \left\langle \frac{\partial \mathsf{H}}{\partial \mathsf{p}_x} \right\rangle \tag{5.22}$$

while putting (b) into (B) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathbf{p}_x \rangle = -\left\langle \frac{\partial \mathsf{H}}{\partial \mathsf{x}} \right\rangle. \tag{5.23}$$

That was hard –but the rest is easy. The usual 1-particle Hamiltonian will be

$$\mathsf{H} = (\mathsf{p}_x^2 + \mathsf{p}_y^2 + \mathsf{p}_z^2)/2m + V(x, y, z),$$

so the partial derivatives are simply

$$\frac{\partial \mathsf{H}}{\partial \mathsf{p}_x} = \frac{\mathsf{p}_x}{m} = v_x,$$
$$\frac{\partial \mathsf{H}}{\partial \mathsf{x}} = \frac{\partial V}{\partial x} = -F_x.$$

If we substitute these results in (5.22) and (5.23) we obtain

$$m \frac{\mathrm{d}}{\mathrm{d}t} \langle \mathsf{x} \rangle = \langle \mathsf{p}_x \rangle$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \mathsf{p}_x \rangle = \langle \mathsf{F}_x \rangle$$

In other words

- The rate of change of the 'average' x-coordinate of the particle, multiplied by its mass, is equal to the 'average' x-component of the momentum.
- And the rate of change of the 'average' x-component of momentum is equal to the 'average' x-component of the force acting on the particle.
- The 'average' here means the 'quantum mechanical expectation value' and the force components are defined from the classical formula, as rates of change of the potential energy function ($F_x = -(\partial V/\partial x)$, etc.).

These important results were first obtained by the Austrian physicist Paul Ehrenfest in 1931 and are usually called "Ehrenfest's theorem". They are often derived using Schrödinger's formulation of quantum mechanics (but in that case you have to do a lot of calculus to get them). Their meaning is clear if you look back at Figure 11, where the moving particle is described by a **wave packet**, travelling through space with a **group velocity**. In classical physics the particle is supposed to be located at a single point, with precisely known coordinates, moving with precisely known momentum components. But in quantum mechanics those 'known' values have to be replaced by **expectation values**, obtained by averaging over the whole region in which the particle is likely to be found. The predictions of classical mechanics are in close agreement with those of quantum mechanics only for wave packets that are very strongly **localized**. In fact, this is true only for **heavy particles**. In deriving the formula (5.11) for a Gaussian wave packet we combined an infinite number of wave trains, travelling with varying frequency and phase velocity: this 'dispersion' was taken into account by expanding the frequency as a power series in terms of wave number k and including only the first two terms and that was enough to give us a wave packet moving without change of shape i.e. with constant width (σ). By including the next term, however, you find the width of the packet increases as it travels: it gradually 'spreads' and, after a long enough time, you completely lose track of the particle!

All this is true only for particles of 'atomic' dimensions; as particle mass increases the behaviour of the particle follows more and more closely the classical 'Newtonian' laws. For an electron you *must* use quantum mechanics, but for a proton (nearly 2000 times heavier) (and even more so for a heavier atomic nucleus) classical physics often gives an acceptable approximation.

We're still talking about only a single particle! You may be thinking there's not much to show for all that hard work. But once you've really understood how to deal with *one* particle you're well on the way to dealing with systems of *many* particles. In Book 4 we started with a single particle (even when it was as big as a planet moving round the Sun), but then we went on to rigid bodies, thinking of them as systems of very many particles – all interacting with each other. And the new principles that came out weren't much more difficult than those that went in.

So here we'll go ahead in much the same way; though there's still quite a lot to do before you can start on 'real' *many-electron* systems like atoms and molecules.

5.5 Commutators and 'constants of the motion'. Angular momentum

By now you've understood how important commutators can be: when the operators associated with two quantities (A and B, say) commute then those quantities can be found with simultaneously definite values – measurement of one doesn't disturb the other. And now the *time* t has been introduced it follows from (5.18) that, for a system whose Hamiltonian is time-independent, the expectation value of any quantity A with operator A – not depending explicitly on the time and commuting with H – the expectation value $\langle A \rangle$ will also be time-independent : it will be a **constant of the motion**. Expressed in symbols, when $(\partial/\partial t)A = 0$ and [A, H] = 0, then $\langle A \rangle$ = constant.

Two examples will remind you of what this means. When A is the Hamiltonian itself (A = H), equation (5.18) tells us that the rate of change of the energy $E = \langle H \rangle$ will also be zero: the value of E will never change. As a second example, look back at the free particle with energy E and momentum components p_x, p_y, p_z (Chapter 4). Here we can take A to be any of the momentum operators (e.g. $A = p_x$) and in this case the eigenvalue $p_x = \langle \mathbf{p}_x \rangle$ will never change in value – it will be a true constant of the motion. The same will be true of the y- and z-components, whose operators also commute with H, and this means the momentum vector will have components that never change from their initial values – the direction of the particle's motion will be fixed.

Let's now turn to something new, *angular* momentum, which we studied in Book 4 (Chapter 5) according to classical physics. For planets moving around the Sun we discovered other constants of the motion: having defined the angular momentum (around the origin) of a particle at point r as the **moment of (linear) momentum p** (i.e. as the vector product $r \times p$), it turned out that the magnitude of the angular momentum was constant in time, along with its three components. These were essentially the laws of Kepler: the particle moves in a plane and the plane of its orbit is also constant in time.

The correspondence principle suggests that something similar may be true at the 'atomic' level. To see if it is, we start from the classical quantities

$$\lambda_x = yp_z - zp_y, \quad \lambda_y = zp_x - xp_z, \quad \lambda_z = xp_y - yp_x,$$

which are the Cartesian components of the angular momentum vector, and set up corresponding **angular momentum operators** in the usual way. The classical components $(\lambda_x, \lambda_y, \lambda_z)$ will be replaced by the *operators* $yp_z - zp_y$ etc., but the quantum mechanical operators are more conveniently defined as

$$\hbar \mathsf{L}_x = \mathsf{y}\mathsf{p}_z - \mathsf{z}\mathsf{p}_y, \quad \hbar \mathsf{L}_y = \mathsf{z}\mathsf{p}_x - \mathsf{x}\mathsf{p}_z, \quad \hbar \mathsf{L}_z = \mathsf{x}\mathsf{p}_y - \mathsf{y}\mathsf{p}_z.$$
 (5.24)

This is in line with the definitions of 'spin' angular momentum, used in Chapter 1 (Section 1.3), where it was noted that the constant \hbar has the physical dimensions of angular momentum, so that L_x, L_y, L_z are now **dimensionless operators**.

The basic commutation properties for position and momentum variables, namely

$$[\mathbf{x}, \mathbf{p}_x] = i\hbar \mathbf{I}, \qquad [\mathbf{y}, \mathbf{p}_y] = i\hbar \mathbf{I}, \qquad [\mathbf{z}, \mathbf{p}_z] = i\hbar \mathbf{I}, \qquad (5.25)$$

must determine the properties of all operators constructed from them. The next Example shows how the argument goes.

Example 5.4 Commutation rules for angular momentum

To evaluate the first commutator $[L_x, L_y] = L_x L_y - L_y L_x$, let's write out the two products of the operators, $\hbar L_x$ and $\hbar L_y$, using (5.24) and noting that position and momentum operators commute when their subscripts differ: thus

$$\hbar^{2}\mathsf{L}_{x}\mathsf{L}_{y} = \mathsf{y}\mathsf{p}_{x}\mathsf{p}_{z}\mathsf{z} - \mathsf{y}\mathsf{x}\mathsf{p}_{z}^{2} - \mathsf{z}^{2}\mathsf{p}_{y}\mathsf{p}_{x} + \mathsf{x}\mathsf{p}_{y}\mathsf{z}\mathsf{p}_{z},$$

$$\hbar^{2}\mathsf{L}_{y}\mathsf{L}_{x} = \mathsf{y}\mathsf{p}_{y}\mathsf{p}_{z}\mathsf{z} - \mathsf{x}\mathsf{y}\mathsf{p}_{z}^{2} - \mathsf{z}^{2}\mathsf{p}_{x}\mathsf{p}_{y} + \mathsf{y}\mathsf{p}_{x}\mathsf{z}\mathsf{p}_{z}.$$

On taking the difference we find

$$\hbar^{2}(\mathsf{L}_{x}\mathsf{L}_{y}-\mathsf{L}_{y}\mathsf{L}_{x})=\mathsf{p}_{z}\mathsf{z}(\mathsf{y}\mathsf{p}_{x}-\mathsf{x}\mathsf{p}_{y})-(\mathsf{y}\mathsf{p}_{x}-\mathsf{x}\mathsf{p}_{y})\mathsf{z}\mathsf{p}_{z}$$

and on using the last commutator in (5.25) this becomes

$$\hbar^2(\mathsf{L}_x\mathsf{L}_y-\mathsf{L}_y\mathsf{L}_x)=i\hbar(\mathsf{x}\mathsf{p}_y-\mathsf{y}\mathsf{p}_x)=i\hbar^2\mathsf{L}_z.$$

Thus, the first commutator becomes $[L_x, L_y] = iL_z$, and two more follow on making the cyclic permutations $x, y, z \to y, z, x \to z, x, y$.

From Example 5.4 we have the commutation relations for angular momentum,

$$\begin{aligned} [\mathsf{L}_x, \mathsf{L}_y] &= (\mathsf{L}_x \mathsf{L}_y - \mathsf{L}_y \mathsf{L}_x) = i\mathsf{L}_z, \\ [\mathsf{L}_y, \mathsf{L}_z] &= (\mathsf{L}_y \mathsf{L}_z - \mathsf{L}_z \mathsf{L}_y) = i\mathsf{L}_x, \\ [\mathsf{L}_z, \mathsf{L}_x] &= (\mathsf{L}_z \mathsf{L}_x - \mathsf{L}_x \mathsf{L}_z) = i\mathsf{L}_z, \end{aligned}$$

$$(5.26)$$

which apply to all kinds of angular momentum (spin included, as you see from Chapter 1). They will be used again and again. Clearly the operators for different components do not commute with each other; so if we know one of them has a definite value then we can't assign values to the others – they are examples of **incompatible observables**. On the other hand it may be that the operators commute with, say H or p^2 – the operators for the energy of the particle or the square of its total angular momentum around the origin. We'll look at these possibilities, both suggested by classical physics (See Book 4, Chapter 5), in the two Examples that follow. the vector components $L_x^2 + L_y^2 + L_z^2$ and the associated operator will be denoted by

$$\mathsf{L}^{2} = \mathsf{L}_{x}^{2} + \mathsf{L}_{y}^{2} + \mathsf{L}_{z}^{2}. \tag{5.27}$$

Note that the squared magnitude is a scalar quantity and its operator carries no subscripts. Its commutation properties follow in the next Example.

Example 5.5 Squared magnitude of the angular momentum

To find the properties of L^2 it's useful to have a simple rule for getting a commutator involving three operators, namely [A, BC]. The commutator, written out in full, is ABC – BCA and is not anything we know about. But if we subtract BAC from the first term and add the same product to the second (which makes no difference to the whole thing) we get

$$[A, BC] = ABC - BAC + BAC - BCA$$

and this can just as well be written (associative and distributive laws, which you've been using ever since Book 1)

$$[\mathsf{A},\mathsf{B}\mathsf{C}] = [\mathsf{A},\mathsf{B}]\mathsf{C} + \mathsf{B}[\mathsf{A},\mathsf{C}].$$

Now look at (5.27) and ask whether the three operator squares commute with L_x , first, and then with L_y and L_z . Of course L_x and L_x^2 commute; but what about L_x and L_y^2 ? Using the 'three-operator rule' (above), choosing $A = L_x$ and $BC = L_y L_y$, you find

$$[\mathsf{L}_x,\mathsf{L}_y^2] = [\mathsf{L}_x,\mathsf{L}_y]\mathsf{L}_y + \mathsf{L}_y[\mathsf{L}_x,\mathsf{L}_y] = i(\mathsf{L}_z\mathsf{L}_y + \mathsf{L}_y\mathsf{L}_z).$$

The commutator of L_x with L_y^2 is *not* zero. But now use the same rule after choosing $A = L_x$ and $BC = L_z L_z$ – and you will find (do it yourself!)

$$[\mathsf{L}_x,\mathsf{L}_z^2] = [\mathsf{L}_x,\mathsf{L}_z]\mathsf{L}_z + \mathsf{L}_z[\mathsf{L}_x,\mathsf{L}_z] = -i(\mathsf{L}_y\mathsf{L}_z + \mathsf{L}_z\mathsf{L}_y).$$

If you add together the two results, getting zero, you'll see that even though L_x doesn't commute with either L_y^2 or L_z^2 separately it does commute with their sum. And since you know it commutes with L_x^2 you can say L_x commutes with L^2 . Moreover, what's good for the x-direction is also good for the y- and z-directions. Similarly, L_x, L_y, L_z all commute with L^2 , though not with each other.

The Example has shown that the square of the angular momentum and *any one* of its components can take simultaneously definite values, since their associated operators commute. We'll suppose this 'special' component defines the z-axis; so L^2 and L_z commute and the state of motion can be labelled by eigenvalues L^2 and L_z . But will it stay that way as time passes? We know from (5.18) that any operator A, not itself depending on t, will be constant it time along with its eigenvalues, provided it commutes with the *Hamiltonian* of the system considered.

For a particle moving in a central field the Hamiltonian used in (4.2) may be written in the general form

$$\mathbf{H} = -(1/2m)\mathbf{p}^2 + V(\mathbf{r}^2), \tag{5.28}$$

since V(x, y, z) for the PE of a particle at point x, y, z depends in this case only on distance r from the origin – and $r^2 = x^2 + y^2 + z^2$.

We now want to know if the operators L^2 and L_z commute with this Hamiltonian, in which case the energy, the squared angular momentum and its z-component can all be constants of the motion.

Example 5.6 Commutation with the kinetic energy term in H

We can again use the 'three operator rule' of Example 5.5, first to see if L_z commutes with the terms in $p^2 = p_x^2 + p_y^2 + p_z^2$. From the definitions (5.24) and the commutation rules (5.25), it follows easily that

$$\begin{aligned} [\mathsf{L}_{z},\mathsf{p}_{x}] &= \mathsf{p}_{y}/i, \\ [\mathsf{L}_{z},\mathsf{p}_{y}] &= -\mathsf{p}_{x}/i, \\ [\mathsf{L}_{z},\mathsf{p}_{z}^{2}] &= 0. \end{aligned}$$

Then by using [A, BC] = [A, B]C + B[A, C], with $BC = L_x L_x$ and $BC = L_y L_y$, in turn, we get

$$\begin{aligned} [\mathsf{L}_z,\mathsf{p}_x^2] &= (\mathsf{p}_x\mathsf{p}_y+\mathsf{p}_y\mathsf{p}_x)/i, \\ [\mathsf{L}_z,\mathsf{p}_y^2] &= -(\mathsf{p}_x\mathsf{p}_y+\mathsf{p}_y\mathsf{p}_x)/i, \\ [\mathsf{L}_z,\mathsf{p}_z^2] &= 0. \end{aligned}$$

Thus, L_z commutes with the operator $sum p_x^2 + p_y^2 + p_z^2 = p^2$. The same must be true for L_x and L_y Finally, if the operators L_x, L_y, L_z all commute with p^2 then so must their squares; so $L_x^2 + L_y + L_z^2 = L^2$ must commute with p^2 .

We now know that both L_z and L^2 commute with p^2 , and therefore with the KE term in the Hamiltonian (5.28). You should try to use similar arguments to show that they also commute with any function of the operator r^2 i.e. with the *potential* energy part of the Hamiltonian. But there's also a nicer way of doing it: just before equation (5.21) it was noted that swapping the operators that refer to corresponding components of position and momentum left the basic commutation rules unchanged except for a sign reversal. Knowing that $[L_z, p^2] = 0$ you can say that $[L_z, r^2] = 0$ (momentum operators p_x^2, p_y^2, p_z^2 being replaced by position operators x^2, y^2, z^2), so $p^2 \rightarrow r^2$. It then follows that $[H, L_z] = 0$ and $[H, p^2] = 0$.

From the last Examples it appears that in any central field system the three operators H, L_z, L^2 all commute with each other; that the energy E, the squared magnitude of the angular momentum vector, and one of its components (which we've called L_z) can all take

simultaneously definite values. And these values will be constants of the motion. This is quite different from the situation according to classical mechanics. The classical and quantum predictions are compared in Figure 12, below.

In the classical picture (a) the particle moves in a precise orbit (roughly circular) around the centre to which it is attracted, just like the Earth moving around the Sun. The orbit lies in a plane (here shown as the xy-plane, with the x-axis pointing towards you) and the angular momentum, indicated by the L-vector, is a constant of the motion, like the total energy. But the three components (L_x, L_y, L_z) are all constants of the motion and, as they determine the direction of the rotation axis – which is normal to the plane of the orbit, this means the orbit never moves out of its plane. All this is just what Kepler and Newton told us, hundreds of years ago.



Figure 12. Motion in a central field

According to quantum mechanics, on the other hand, (look at picture (b)) there can be no precise orbit when you go down to 'atomic' dimensions: instead, all you can get is a wave function (in 'Schrödinger language) with $|\Psi|^2$ showing where the particle is most likely to be found. But some of the concepts of classical physics still show up: the particle can have an angular momentum around the attracting centre (even if you can't tell exactly where it is!); and it turns out that the length² of the L-vector is observable, along with one of its components – which we've called L_z . The other two components you can never find – because if you try to measure them you'll disturb the value you found for L_z . Only their average values (which are zero) can be observed: $\langle L_x \rangle = \langle L_y \rangle = 0$ for the components transverse to the z-axis. If you want a semi-classical picture the best you can get is the one indicated in Fig.12(b), where the L-vector doesn't point along the measured L_z axis but **precesses** around it, lying always on the surface of a cone, with its transverse components bounded by definite limits.

In the present Chapter 5 we've constructed a solid 'bridge' between 'classical' and 'quantum' concepts. Now we're ready to cross into the quantum world.

Chapter 6

Hydrogen – the simplest atom of all

6.1 An electron in the central field of a nucleus

We're now all set to study the 'classic' central-field problem of a single (negatively charged) electron moving in the field of a positively charged nucleus. What are the allowed **energy levels** of the system? How can we find them and how can we begin to understand **atomic structure**? As you know from Book 5, these questions lie at the root of chemistry. They go far beyond – for molecules and all kinds of matter are built up from atoms – even the enormous molecules of life (which you already met in Book 9). By starting with the simplest system of all, the single hydrogen atom, we can uncover many new ideas and methods that apply very generally.

In Chapter 5 you studied the motion of a single particle in a central field, where the Hamiltonian operator had the form $H = -(1/2m)p^2 + V(r)$, p being the momentum operator and r the radial distance of the particle from the origin. In Schrödinger language this takes the form

$$\mathbf{H} = -\frac{\hbar^2}{2m} \nabla^2 - Z \frac{e^2}{\kappa_0 r},\tag{6.1}$$

as used in Section 4.5, ∇^2 being the 'del squared' operator defined in (4.3). The **atomic number** Z has been inserted (this being the number of electrons in the uncharged atom) so that we can also study the positively charged **1-electron ions** that remain when one or more electrons have been taken away from a heavier atom. (For example, the Lithium atom (Li) has three electrons, but on taking two of them away you get the Li²⁺ ion – another 1-electron system.) With Hamiltonian (6.1) the energy eigenstates of the system follow on solving the Schrödinger equation

$$\mathsf{H}\Psi = E\Psi,\tag{6.2}$$

which determines the allowed values of the energy E in the possible stationary states of the system.

By the end of the last Section it was clear that in any central field system the three operators H, L_z, L^2 would all commute with each other; that the energy E, the squared magnitude of the angular momentum vector, and one of its components (which we've

called L_z) could all take simultaneously definite values. And that these values would be **constants of the motion**. Before moving on we need to fill in the details. How can we find the eigenstates? – and what will be the corresponding eigenvalues of the angular momentum operators? The next two Examples show how we can get them.

Example 6.1 Eigenstates of angular momentum

We start from the commutation rules for L_x, L_y, L_z . The first one is

$$[\mathsf{L}_x,\mathsf{L}_y] = (\mathsf{L}_x\mathsf{L}_y - \mathsf{L}_y\mathsf{L}_x) = i\mathsf{L}_z,$$

and there are two other equations, obtained by making a cyclic permutation of the labels x, y, z. We also need step-up and step-down operators similar to those used in Section 2.2, where they were defined for spin as

$$S^+ = S_x + iS_y, \qquad S^- = S_x - iS_y.$$

Here we'll have L instead of S and the properties we need will be

$$\mathsf{L}^{-}\Psi_{L,M} = c_{M}^{-}\Psi_{L,M-1}, \qquad \mathsf{L}^{+}\Psi_{L,M} = c_{M}^{+}\Psi_{L,M+1},$$

where the constants $c_M^+ c_M^-$ are put in because we don't yet know if the operators leave the states normalized – as we would wish. (In Example 6.3 we'll find their values.)

In the case of spin, we were able to find alternative expressions for the *total* spin operator $S^2 = S_x^2 + S_y^2 + S_z^2$, namely

$$S^{2} = S^{-}S^{+} + S_{z} + S_{z}^{2} = S^{+}S^{-} - S_{z} + S_{z}^{2}$$
.

And, as we already know, you can get corresponding relations for other kinds of angular momentum just by changing the names (e.g. using "L" in place of "S"). The operator for the squared magnitude of the *orbital* angular momentum will thus be, using the first form of the equation for S^2 but with "S" changed to "L",

$$\mathsf{L}^2 = \mathsf{L}^+ \mathsf{L}^- - \mathsf{L}_z + \mathsf{L}_z^2.$$

This last result lets us express L^+L^- in terms of operators whose eigenvalues we know. In particular, $L^+L^- = L^2 - L_z^2 + L_z$ and if we operate with this on any eigenstate $\Psi_{X,M}$, where X and M are eigenvalues of L^2 and L_z , the result will be (do it!)

$$L^+L^-\Psi_{X,M} = (X - M^2 + M)\Psi_{X,M}.$$

Example 6.1 has given the following important results:

$$L^{+}L^{-} = L^{2} - L_{z}^{2} + L_{z}, \qquad L^{-}L^{+} = L^{2} - L_{z}^{2} - L_{z}, \qquad (6.3)$$

and when acting on eigenstates of L^2 and L_z , with eigenvalues X and M, these results show that

$$L^{+}L^{-}\Psi_{X,M} = (X - M^{2} + M)\Psi_{X,M},$$

$$L^{-}L^{+}\Psi_{X,M} = (X - M^{2} - M)\Psi_{X,M},$$
(6.4)

The next Example shows how you can get the values of M and X.

Supposing any given eigenstate is normalized, the length² of the vector $\Psi_{X,M}$ will be the scalar product $(\Psi_{X,M}|\Psi_{X,M}) = 1$. On taking the scalar product from the left of the last equation in (6.4), with $\Psi_{X,M}$, the result will be

$$(\Psi_{X,M}|\mathsf{L}^+\mathsf{L}^-\Psi_{X,M}) = (X - M^2 + M).$$

Now in Chapter 2, Example 2.4 (which you may want to read again), it was noted that you could move any operator C from one side of a scalar product to the other (i.e. across the vertical 'bar') provided you changed it into its **adjoint operator** C^{\dagger} . If you do this with L⁺ in the last equation you get

$$(X - M^2 + M) = (\mathsf{L}^- \Psi_{X,M} | \mathsf{L}^- \Psi_{X,M})$$

- because the adjoint of L⁺ (which is $L_x + iL_y$) is L⁻. (Check it out, remembering that L_x, L_y are *self*-adjoint, but that you have to change the sign of *i*.) But this scalar product is the squared length of the vector $L^-\Psi_{X,M}$ and can only be zero or positive: so $(X - M^2 + M) \ge 0$.

If you repeat all that, starting with L^-L^+ in place of L^+L^- , you find a second inequality $(X - M^2 - M) \ge 0$ and on adding them both together you get $X \ge M^2$:

> The eigenvalue M of the angular momentum z-component cannot exceed X, the eigenvalue of the operator L^2 for the square of the total angular momentum.

Finally, let the highest and lowest values of M be M_+ and M_- This means that the 'top' state vector can't be stepped up by applying L⁺; or in other words L⁺ $\Psi_{X,M_+} = 0$ – for otherwise there would be a state with M-value higher than the highest!

Also, taking L^2 in the alternative form $L^2 = L^-L^+ + L_z + L_z^2$, and putting $M = M_+$, it follows that

$$\mathsf{L}^{2}\Psi_{X,M_{+}} = (M_{+}^{2} + M_{+})\Psi_{X,M_{+}}$$

– since the multiplier on the right is the *eigenvalue* of the operator on the left, which we called X. This means $X = M_+(M_+ + 1)$ in terms of the largest possible value of the angular momentum z-component. Using the same argument on the 'bottom' state Ψ_{X,M_-} you'll find $X = M_-(M_- - 1)$ and equating the two equal results gives

$$M_{+}^{2} + M_{+} = M_{-}^{2} - M_{-}.$$

This may be re-written as (check it!)

$$(M_{+} + M_{-})(M_{+} - M_{-} + 1) = 0,$$

with the only acceptable solution (say why) when the first factor is zero.

If the highest M-value a component can take is denoted by $L (= M_+)$, this fixes the eigenvalue of L^2 as X = L(L+1) and tells us that L_z will have eigenvalues going in unit steps from -L to +L, the integer L being the **total angular momentum quantum number**.

From Example 6.2, we can write the eigenvalue equations for orbital angular momentum in the standard forms

$$\mathsf{L}^{2}\Psi_{L,M} = L(L+1)\Psi_{L,M} \tag{6.5}$$

and

$$\mathsf{L}_{z}\Psi_{L,M} = M\Psi_{L,M} \quad (M = L, L - 1, \dots - L), \tag{6.6}$$

where L and M are used as the standard labels for the quantized states. It remains only to find values of the coefficients c_M^+ and c_M^- used in Example 6.1, choosing them so that $\Psi_{L,M}$ will stay normalized when the value of M is stepped up or down

Example 6.3 Values of the constants c_M^{\pm}

In moving from one eigenstate $\Psi_{L,M}$ to another, by using the step-up operator L^+ to get $\Psi_{L,M+1} = c_M^+ \mathsf{L}^+ \Psi_{L,M}$, we want to be sure the new eigenstate will still be normalized. We therefore require

$$(\Psi_{L,M+1}|\Psi_{L,M+1}) = |c_M^+|^2 (\mathsf{L}^+ \Psi_{L,M}|\mathsf{L}^+ \Psi_{L,M}) = 1,$$

a vector of unit length for all states.

Now the scalar product on the right can be written (moving the first operator L^+ over the vertical bar and changing it to L^-) as $(\Psi_{L,M}|L^-L^+|\Psi_{L,M})$ and the resultant operator product L^-L^+ may be replaced by $L^-L^+ = L^2 - L_z^2 - L_z$.

It follows that, since $\Psi_{L,M}$ was supposed to be an eigenstate of L^2 and L_z with eigenvalues L(L+1) and M,

$$|c_M^+|^2(\Psi_{L,M+1}|\Psi_{L,M+1}) = (\mathsf{L}^+\Psi_{L,M}|\mathsf{L}^+\Psi_{L,M}) = (\Psi_{L,M}|[L(L+1) - M^2 - M]\Psi_{L,M}).$$

In other words the squared length of the 'shifted' eigenvector is $[L(L+1) - M^2 - M]$ times that of the original; which identifies the constant c_M^+ as $c_M^+ = \sqrt{[L(L+1) - M^2 - M]}$. A similar calculation for the step-down operator (do it!) shows that $c_M^- = \sqrt{[L(L+1) - M^2 + M]}$. (Note, however, that only the square modulus was fixed and the coefficients may be multipled by ± 1 or by any unimodular complex number $e^{i\theta}$ without changing anything: the coefficients shown therefore indicate a particular choice of the 'phase factor', with $\theta = 0$. This is called a "phase convention".)

The results from Example 6.3 are usually written in a slightly different form by noting that $L(L+1) - M(M \pm 1) = (L \mp M)(L \pm M + 1)$, from which there follows

$$\mathsf{L}^{+}\Psi_{L,M} = \sqrt{(L-M)(L+M+1)}\Psi_{L,M+1}, \tag{6.7}$$

$$\mathsf{L}^{-}\Psi_{L,M} = \sqrt{(L+M)(L-M+1)}\Psi_{L,M-1}.$$
(6.8)

The equations found above are very general and may be used for all kinds of angular momentum, depending only on the commutation relations satisfied by the operators. We'll meet them again and again.

But so far nothing has been said about the **energy eigenstates**, which can exist for all values of the angular momentum and its components. To find them we have to return to equation (6.1).

6.2 Finding the energy levels for a 1-electron atom

It's quite difficult to find general solutions of the eigenvalue equation (6.1), but in Section 4.5 we did find the lowest-energy eigenstate of the hydrogen atom – both the energy value and the corresponding wave function. (You should read that Section again before going on.)

Here we'll look for particular solutions in which the wave function has **spherical symmetry**, depending only on the distance r of the electron from the nucleus. Such functions are called **s-orbitals** and a whole series of solutions of that type can be found without much trouble. To stress that they are *radial* functions, depending on the single variable r, they will be denoted by $\Psi = R(r)$.

We start from (6.1) written in the re-arranged form

$$\left(\nabla^2 + \frac{2mZe^2}{\kappa_0\hbar^2}\frac{1}{r} + \frac{2m}{\hbar^2}E\right)R = 0.$$

This looks a bit messy with all those constants, but it can be simplified by using the units we first met in Chapter 4 (Section 4.5). Remember that distances and energies on an atomic scale are often measured in terms of the "Bohr radius" a_0 and the "Hartree" energy unit $e_{\rm H}$, defined as

$$a_0 = \kappa_0 \hbar^2 / m e^2$$
, $e_{\rm H} = m e^4 / \kappa_0^2 \hbar^2$.

Thus, $r = \rho a_0$ and $E = \epsilon e_{\rm H}$ are the quantities r and E, expressed as multiples of a_0 and $e_{\rm H}$, respectively, while ρ and ϵ (the corresponding Greek letters) are the numerical measures of the quantities. Sometimes we use the quantities as if they were just numbers – not thinking about the units – but we must be careful! When we come across "exp x", for example, the x must be a number, because $e^x = 1 + x + \frac{1}{2}x^2 + \ldots$ and if x measured a quantity its value would change when the units were changed – giving a completely new series!

So let's write the equation we want to solve in the form

$$\left(\nabla^2 + \frac{2mZe^2}{\kappa_0\hbar^2}\frac{1}{a_0}\frac{1}{\rho} + \frac{2m}{\hbar^2}\mathbf{e}_{\mathrm{H}}\epsilon\right)R = 0,$$

where the units are shown explicitly. On noting that $e_{\rm H} = e^2/\kappa_0 a_0$, this becomes (check it out!)

$$\left(\nabla^2 + \frac{2m\mathbf{e}_{\mathrm{H}}}{\hbar^2} \left[\frac{Z}{\rho} + \epsilon\right]\right) R = 0.$$

Here the physical constants are neatly put together in the coefficient $(2m_{\rm H}/\hbar^2)$, which has zero dimensions and is thus a pure number, not depending on how the units are chosen. To prove it, note that energy has dimensions $[E]=ML^2T^{-2}$, while $[\hbar] = MLT^{-1}$. (Look back at Book 4, for example, if you've forgotten what is meant by [E] – "the physical dimensions of energy".)

Even better, we can make the constants disappear from our equations by using the full set of 'atomic units' defined in Section 4.5, in which $e, m, \kappa_0, \hbar, a_0$ and e_H all take unit values!

Now we can start looking for s-type solutions of the eigenvalue problem. We know from Example 4.7 that

$$\nabla^2 \equiv \frac{\mathrm{d}^2}{\mathrm{d}\rho^2} + \frac{2}{\rho} \frac{\mathrm{d}}{\mathrm{d}\rho},$$

so we obtain an ordinary differential equation

$$\frac{\mathrm{d}^2 R}{\mathrm{d}\rho^2} + \frac{2}{\rho} \frac{\mathrm{d}R}{\mathrm{d}\rho} + \frac{2Z}{\rho} R + 2\epsilon R = 0, \tag{6.9}$$

which is to be solved, subject to the usual conditions (R finite, continuous, etc.). This is done in the next Example.

Example 6.4 Solving the differential equation

Here we'll use a common method of getting a solution to equations like (6.9). First we look for an 'asymptotic' solution, valid in the range where r (and hence ρ) takes very large values. On letting $\rho \to \infty$ the terms with ρ in the denominator tend to zero and may be dropped leaving

$$\frac{\mathrm{d}^2 R}{\mathrm{d}\rho^2} + 2\epsilon R = 0,$$

which has the solution $R = \exp \pm \sqrt{2\epsilon}\rho$ (differentiating R twice simply multiplies it by 2ϵ). (Remember that in dealing with the ground state we called the exponential factor e^{-br} ; and we'll do the same here by putting $\sqrt{2\epsilon} = b$.) The positive sign in the exponent is no good because $R = \exp -b\rho$ would then go off to infinity for $\rho \to \infty$; but the negative sign gives a well-behaved function when $\rho \to \infty$. On the other hand there is a *singularity* in the potential energy function Z/ρ , which becomes infinite when $\rho \to 0$. To make sure the solution doesn't 'blow up' at that point we can include in the 'trial function' a factor such as ρ – which goes to zero. In fact, we'll look for the equation to be satisfied by a new function $P = \rho R$, which remains zero as $\rho \to 0$

To find the equation satisfied by P we first get the derivatives

$$\frac{\mathrm{d}P}{\mathrm{d}\rho} = R + \rho \frac{\mathrm{d}R}{\mathrm{d}\rho}, \qquad \frac{\mathrm{d}^2 P}{\mathrm{d}\rho^2} = 2\frac{\mathrm{d}R}{\mathrm{d}\rho} + \rho \frac{\mathrm{d}^2 R}{\mathrm{d}\rho^2};$$

and then note that $d^2 P/d\rho^2$ includes *both* the derivatives in (6.9). The differential equation can then be rewritten, after removing a common factor ρ^{-1} , as

$$\frac{\mathrm{d}^2 P}{\mathrm{d}\rho^2} + \left(\frac{2Z}{\rho} + 2\epsilon\right)P = 0$$

– which looks much nicer than (6.9).

Again, there is an asymptotic solution for large values of ρ , where $P \approx \exp -b\rho$. At the other limit, where $\rho \rightarrow 0$, the factor ρ should be included, and between the two limits it seems reasonable to represent the wave function by writing

$$P = A\rho \exp{-b\rho}, \qquad A = a_0 + a_1\rho + a_2\rho^2 \dots a_k\rho^k; \dots,$$

where the constants $a_0, a_1, ..., a_k$... are to be determined. (Don't be confused by the ' a_0 '! It's just the coefficient of ρ^0 (= 1) – not the 'Bohr radius'.)

To get the values of the constants, for any particular solution, you have to put the function P into the differential equation and then insist that the resultant polynomial be zero: if the polynomial is *finite*, then the series for A will finish after, say n terms. In that case you can get a finite solution of the nth degree in ρ . But you have to work for it by showing that the coefficients $a_0, a_1, ..., a_n$ must then be related: in fact there is a **recurrence relation**

$$k(k+1)a_k + 2a_{k-1}(Z - kb) = 0$$

from which, starting from $a_0 = 1$ (k = 0), it follows that $2a_1 = 2(Z - b)a_0$ (k = 1); $6a_2 = 2(Z - 2b)a_1$ (k = 2); etc.

If the solution you're looking for contains only one term, $a_0\rho^0$, then the series for A must end with $a_0 = 1$, giving $a_1 = (Z - b)a_0 = Z - b = 0$. This represents the ground state; and since $b = \sqrt{-2\epsilon}$ it determines the energy of the state, namely $-2\epsilon = b^2 = Z^2$. Thus (in atomic units), $E_{1s} = -\frac{1}{2}e_{\rm H}$. In the same way, but with two terms, you'll find

$$a_0 = 1$$
, $a_1 = (Z - b)a_0$, $a_2 = (1/3)(Z - 2b)a_1$

and if this third term has to be zero then (Z - 2b) = 0. The 2-term series thus corresponds to energy $-\epsilon = -\frac{1}{2}(Z/2)^2 = -\frac{1}{2}(Z^2/4)e_{\rm H}$. And so it goes on. The general result, as you can guess, is simply $E_{ns} = -\frac{1}{2}(Z^2/n^2)e_{\rm H}$.

From now on we'll use atomic units throughout, so when you see r it will mean "r in units of a_0 " and an energy E will mean "E in units of e_H ".

The results obtained so far can now be summarized as follows:

• For one electron in the central field of a nucleus of charge Z, the Schrödinger equation has well-behaved spherically symmetrical solutions of the form $P_n(r) = rR_n(r)$, with corresponding energy values

$$E_1 = -\frac{1}{2}\frac{Z^2}{1}, \quad E_2 = -\frac{1}{2}\frac{Z^2}{2^2}, \quad E_3 = -\frac{1}{2}\frac{Z^2}{3^2}, \dots \quad E_n = -\frac{1}{2}\frac{Z^2}{n^2}, \dots$$
(6.10)

These solutions define the **atomic orbitals (AOs)** of 's type' and form an infinite sequence, converging to an upper limit E = 0; in other words they define states of *negative energy*, in which the electron is *bound* to the nucleus. Above these discrete levels there will be *positive-energy* solutions, corresponding to the electron moving freely through space (cf. Section 4.xx). We haven't yet studied such solutions; but they describe an electron 'scattered' by the presence of the proton, without being 'captured' by it. These energy levels form a *continuum*, being infinitely close together.

• For the hydrogen atom (Z = 1) in its ground state, the **principal quantum number** has the value n = 1. To knock the electron out of the atom in this state, leaving a hydrogen *ion* H⁺ (i.e. a bare proton) and a **free electron**, a long way from the nucleus and with energy close to zero, you have to increase its energy by $\Delta E = 0 - (-\frac{1}{2})$, which means $\frac{1}{2}$ e_H. This **ionization energy** leads to the alternative interpretation of the atomic unit e_H as "twice the ionization energy of the hydrogen atom", as mentioned in Section 6.4. All this is clear from the energy-level diagram that follows (Figure 13).



Figure 13. Electronic energy levels, H atom

Bound states at $E_n = -\frac{1}{2}(1/n^2)e_{\rm H}$, up to E = 0Positive energy states form a continuum (grey) Bold up-arrow indicates ionization from E_1 The states referred to in Fig.13, whose wave functions have spherical symmetry, are 's states'. At this point we'll indicate their forms to show how they depend on the distance (r) of the electron from the nucleus. Note first however that the functions in the Table that follows are **normalized to unity** only when the normalizing factors N_{1s} , N_{2s} , etc. are suitably chosen. And remember what this means for a spherically symmetrical wave function: $|\Psi|^2$, when integrated over all space, must give unity (unit probability of finding the electron somewhere. In doing the integration, the volume element is $4\pi r^2 dr$ for a spherical shell of thickness dr, inside which $|\Psi(r)|^2$ has a constant value, namely $|R(r)|^2$. Integrating over all angles, around the nucleus, just gives a factor $4\pi r^2 dr$; and the remaining integration over radial distance r must therefore be multiplied by 4π to get the correct volume integral of $|\Psi|^2$. Consequently, even when the radial function R(r) is normalized for the integration over r, an extra factor of $1/\sqrt{4\pi} = 1/2\sqrt{\pi}$ is needed in normalizing the wave function Ψ .

The wave functions that follow apply to a system with positive charge Z on the nucleus (in units of e, the atomic unit of electric charge):

n	L	Wave function $\Psi_{n,L}(r)$
1	0	$\Psi_{1,0} = N_{1s} e^{-Zr}$
2	0	$\Psi_{2,0} = N_{2s}(1 - \frac{1}{2}Zr)e^{-Zr/2}$
3	0	$\Psi_{3,0} = N_{3s} [1 - (2/3)Zr + (2/27)Z^2r^2]e^{-Zr/3}$
4	0	$\Psi_{4,0} = N_{4s} [1 - (3/4)Zr + (1/8)Z^2r^2 - (1/192)Z^3r^3)e^{-Zr/4}$

Hydrogen-like s-type wave functions

The complete normalizing factors are listed below:

$$N_{1s} = \frac{1}{\sqrt{\pi}} Z^{3/2}, \ N_{2s} = \frac{1}{2\sqrt{\pi}} Z^{3/2}, \ N_{3s} = \frac{1}{3\sqrt{3\pi}} Z^{3/2}, \ N_{4s} = \frac{1}{8\sqrt{\pi}} Z^{3/2}.$$

The exponential factors in the Table above show that the wave functions and corresponding probability densities fall to zero more and more rapidly as Z increases. This means that the orbitals for atoms of the heavier elements shrink tightly around the nucleus; those for small values of the principal quantum number n thus give a highly concentrated probability density.

It is the *energies* of the hydrogen s states, given by the formula (6.10) with Z = 1, that are indicated in Figure 13: the *forms* of the orbitals are pictured in Figure 14 (roughly to scale).



Figure 14. Forms of the 1s, 2s and 3s orbitals (see text)

In the Figure the spherical regions where Ψ is positive are coloured pink; those where it is negative are shown in light blue. The solid lines show the **nodal surfaces**, on which $\Psi = 0$ and the sign is changing. (Of course the sign doesn't matter, as only $|\Psi|^2$ gives physical information, but the sign *changes* do.) The outer 'boundaries', indicated by broken lines, show where Ψ has fallen to a very small fraction of its maximum value; they give an idea of the relative sizes of the orbitals.

6.3 Eigenfunctions with angle dependence

Other sequences of bound states can be found in which the **angular momentum** variables L^2 , L_z) may also have simultaneously definite values, specified by the **quantum numbers** L, M found in Section 6.5. These correspond to 'p states' (with L^2 quantum number L = 1, and L_z , with M = +1, 0, -1); to 'd states' (with L^2 quantum number L = 2, and L_z , with M = 2, 1, 0, -1, -2); followed by 'f states', with L = 3, and M = 3, 2, 1, 0, -1, -2, -3), and so on. The first few levels of each series are set out in the diagram below (Figure 15).



Figure 15. Origin of the H-atom spectrum (see text)

The four energy-level diagrams refer to s, p, d and f states, with quantum numbers L = 0, 1, 2, 3, respectively. Note that, as long as the potential function V(r) has the inverse-distance form, the energy levels for the series with $L \neq 0$ exactly match those for the s series, given in (6.10) except that those with principal quantum number n < L are missing. Thus, the lowest energy p state (with L = 1) corresponds to n = 2 and is a "2p state", with the same energy as a "2s state"; in the same way, the d states (with L = 2) have a lowest energy coinciding with that of the 3s state.

The upward-sloping arrows in Figure 15 show how various series of **spectral lines** originate from **transitions** between the energy levels. For a central-field system with V(r) = Z/r (atomic units!), all states with the same value of the principal quantum number n have the same energy; so the energy *separation* of any two levels with quantum numbers n_1 and n_2 will be

$$\Delta E = -\frac{1}{2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right). \tag{6.11}$$

This celebrated formula, first obtained by Bohr, using his semi-classical model of the atom, is important because it relates so directly to experimental data. The energy change

when the system goes from State 1 to State 2 allows one to predict the frequency (ν) of a quantum of light absorbed or emitted, using only Planck's law: thus

$$\Delta E = -\frac{1}{2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) = \hbar \nu.$$
(6.12)

If 1 and 2 denote the 'initial' and 'final' states $\hbar\nu = \Delta E(1 \rightarrow 2)$ will come out positive if $n_2 > n_1$ (giving $E_2 > E_1$) and this transition will correspond to light **absorption**; but if $n_2 < n_1$ it will correspond to **emission**. It is such transitions that are often called "quantum jumps": when an atom increases its electronic energy through interaction with radiation of frequency ν (see Book 10) the energy absorbed is $\epsilon = \hbar\nu$ – the energy of one photon or 'light particle'; and when it loses a quantum of energy a photon of the same energy is emitted. In either case, there is no change in total energy of atom plus field – simply a transformation of one form into the other.

A note on energy units

The energy unit used in classical physics (Book 4) is much too big to be useful in atomic physics. You'll remember that the Joule (1 J) is the energy needed to move unit mass (1 kg) through unit distance (1 m) when unit force (1 N) acts on it. These units are all enormous by 'atomic' standards. So in atomic physics the most commonly used unit is the **electron volt** (1 eV), the energy needed to carry unit charge (1 e) through an electric potential difference of 1 volt: this is only about 1.60×10^{-19} Joules! On the other hand, the 'Hartree' 1 e_H is approximately 27.21 eV.

Another convenient unit is based on the Planck relationship $\epsilon = \hbar \nu$, which is the energy of one quantum of radiation of frequency ν . As you know from Book 10, $\nu = c/\lambda$ where cis the velocity of light in free space and λ is the wave length of the radiation: so $\epsilon = \hbar c/\lambda$ gives the energy of the quantum as $\epsilon = \hbar c \times k$ where k is the **wavenumber**, the number of waves per unit length i.e. the reciprocal of a length. On putting in the numerical values of the constants we find that

$$1 e_{\rm H} \equiv 27.21 \, eV \equiv 219.5 \times 10^3 \, cm^{-1}$$
,

which is a convenient relationship involving the most commonly used units.

The hydrogen atom absorption spectrum

Remember anything about radiation and the electromagnetic spectrum? If not, go back to Book 10 and read Chapter 6 again. Light is a form of electromagnetic radiation and 'white' light is a mixture of waves of all different frequencies, corresponding to wavenumbers between about 1.4×10^4 cm⁻¹, for red light, and 3.0×10^4 cm⁻¹ for violet light. The **spectroscope** (or spectrometer if it's used for *measuring*) is a device for separating the various frequencies. When a beam of white light goes in, it comes out as a 'rainbow' of all colours side-by-side – and that's the visible part of the spectrum. On either side of the band of colours the spectrum stretches out; on the low-frequency side (as far as the radiant heat from your electric radiator and the much lower-frequency radiation used in long-wave radio) and on the high-frequency side (as far as the ultra-violet radiation, that gives you sunburn, and way beyond to X-rays and the 'cosmic rays' that come from outer space) – all of them invisible to the human eye. A long time ago it was noticed that the spectrum of light from the sun was crossed by fine black lines at certain frequencies: these result from absorption of radiation by the atoms that get in its way, each black line arising from missing light of the frequency that causes electronic transitions. So the spectrometer gives us a powerful method of 'observing' the quantized energy levels in an atom. The hydrogen atom transitions indicated in Fig.15 fall into series that carry the names "Lyman" (for the big energy jumps that start from $n_1 = 1$ in the formula (6.12); "Balmer" (for those that start from $n_1 = 2$; "Paschen" (for those starting from $n_1 = 3$; and so on. Each series was named after its discoverer. Only the Balmer series falls in the visible part of the spectrum, the Lyman series being found in the ultra-violet (UV) region and the Paschen in the infra-red. These conclusions follow from the next Example.

Example 6.5 Frequencies of some absorption lines

On using (6.12) with $n_1 = 2$ the lowest energy transition in the Balmer series follows when $n_2 = 3$, as you can see from Fig.15, and this gives $\Delta E = -\frac{1}{2}[(1/4) - (1/9)]e_{\rm H} = 0.06945 e_{\rm H}$. From the equivalences given above, this translates into

 $\Delta E = 0.06945 \times 27.21 = 1.890 \,\mathrm{eV} \equiv 0.06945 \times 219.5 \times 10^3 \,\mathrm{cm}^{-1} = 15,244 \,\mathrm{cm}^{-1},$

which corresponds to absorption of light in the visible region (towards the red end of the spectrum).

A similar calculation for the lowest energy transition in the Lyman series (do it!) predicts absorption of radiation with wavenumber $82, 263 \,\mathrm{cm}^{-1}$. This much larger transition energy corresponds to absorption well inside the ultraviolet range.

By now you will be wondering what the orbitals 'look like' for wave functions $\Psi(r, \theta, \phi)$ which depend on two angles, θ and ϕ , as well as the radial distance r of the electron from the nucleus. In this case things become even more difficult! But it's possible to understand how the argument goes without following all the details and that's the path we'll follow in the next Section.

6.4 Angle-dependent wave functions

Remember that we're looking for wave functions with the general form, using **polar** coordinates,

$$\Psi = \Psi(r, \theta, \phi) = R(r)Y(\theta, \phi), \tag{6.13}$$

which is written in the **separated form** as a product of a radial factor R(r) – like the ones in the last Section – and a factor $Y(\theta, \phi)$ depending on two **rotation angles**.

When $Y(\theta, \phi) = 1$, the wave function $\Psi = R(r)$ has spherical symmetry, like the functions pictured in Fig.14, because it always looks the same whatever rotations you make. The radial factor is similar to the ones shown in Fig.14, R(r) alternating in sign as r increases – with positive and negative regions separated by spherical nodal surfaces. But more generally the angle dependence is quite complicated in terms of θ and ϕ ; it looks much simpler in terms of the Cartesian coordinates (x, y, z) so you need to know how the two coordinate systems are related. The next Figure shows the connection.



Figure 16. Spherical polar coordinates (r, θ, ϕ)

r = radial distance of electron from nucleus Position vector **r** lies in shaded plane, making angle ϕ with zx-plane, and θ with z-axis.

The angle dependence shows up very clearly in Cartesian coordinates. The eigenfunctions of L^2 with quantum number L = 1 (i.e. the p functions) have angle-dependent factors x, yz; and there are three of them, corresponding to the L_z quantum number $M = \pm 1, 0$. Similarly the d functions, with L = 2, are formed from products of the second degree, namely $x^2, y^2, z^2, xy, yz, zx$. And you might expect there to be six independent wave functions, one for each choice. But you'd be wrong, because there's a relationship among the first three: $x^2 + y^2 + z^2 = r^2$ and r^2 is not an angle-dependent function – the distance r of the electron from the nucleus doesn't change if you turn the system round. In fact there are only five independent d-type wave functions and these are usually taken to have the angle-dependent factors

$$\frac{x^2 - y^2}{r^2}, \quad \frac{3z^2 - r^2}{r^2}, \quad \frac{xy}{r^2}, \quad \frac{yz}{r^2}, \quad \frac{zx}{r^2}.$$

These factors are not in any way normalized; they are listed here only in order to explain the forms of the orbitals for higher values of the L quantum number. Some typical orbitals for L = 1 and L = 2 are indicated schematically in Figure 17:



Figure 17. Forms of some typical 2p and 3d orbitals (see text)

the first is a p function (2pz), pointing along the z-axis, while 2px and 2py (not shown) look exactly the same but point along the x- and y-axes; the next is a d function $(3d_{zx})$,

which has its four 'lobes' centred in the zx-plane and its normal pointing along the y-axis, and two more $(3d_{xy} \text{ and } 3d_{yz}, \text{ not shown})$ which are just like it but have their normals pointing along the z- and x-axes. There are two more 3d-functions: $3d_{x^2-y^2}$ looks like $3d_{xy}$, but rotated by 45° around the z-axis – so that its lobes are centred on the x- and y-axes, instead of falling between them. Finally there is one that looks quite different from the others, shown in the last picture of Fig.17. It has two big lobes on the z-axis, both positive, but a dough-nut shaped belt (called a "torus") around its middle, of opposite sign. However, you can get it simply by combining two functions, which you could call $(3d_{y^2-z^2})$ and $(3d_{z^2-x^2})$, formed by making the usual 'cyclic permutation' of the axis labels x,y,z. So the odd-looking $3d_{z^2}$ orbital is just a particular mixture of the 4-lobe functions you've already got – not really something different! We say "there are only five *linearly independent* 3d orbitals" and they all correspond to the same energy eigenvalue i.e. they are *degenerate*. In the same way, it turns out that there are seven degenerate 4f orbitals and in general 2L + 1 degenerate orbitals with angular momentum quantum number L. Later you'll understand why energy eigenfunctions are so often found in degenerate groups.

To end this Section, now that you know about the energy levels for an electron in a central field and how they relate to experimental spectroscopy, through the 'quantum jumps' that take the electron from one level to another, let's go back briefly to something that must have been worrying you –

What *makes* an electron jump?

So far we've been thinking always of an isolated system, which can stay in a definite energy eigenstate 'forever' – such states being *stationary*. To make the system change you must do something to it; you must disturb it and a small disturbance of this kind is called a **perturbation**.

The simplest way of disturbing the electron is to apply an electric field from outside the atom, in addition to that due to the nucleus (the central field considered so far). This will change the potential energy term in the Hamiltonian H, so that $V(x, y, z) \rightarrow V(x, y, z) + \delta V(x, y, z)$ and, if the 'external' field is uniform and is in the x-direction, the change will have the form $\delta V = Fex$. Here F is used for the field strength (so as not to mix it up with the energy E) and the arbitrary zero for δV is taken as the origin of coordinates (i.e. the nucleus). If F is constant its effect will be to produce a small **polarization** of the system by urging the electron in the (negative) direction of the field, since the electron carries a negative charge -e, and this means the probability function will 'lean' slightly to one side. This effect will be small: if the change in the wavefunction is neglected in a first approximation the change in expectation value of H will be $\delta E = \langle \delta H \rangle = \langle \delta V \rangle$ and this will be zero (can you say why?).

In order to produce a transition, between initial and final states with energies E_i and E_f corresponding to absorption of a quantum of energy $\hbar\nu = E_f - E_i$, it is necessary to 'shake' the system. This means the disturbance must be *time-dependent*, describing an *oscillating perturbation*. For example, the applied electric field could have the form $F = F_0 e^{i\omega t}$ with the 'angular frequency' $\omega = 2\pi\nu$. To deal with this kind of perturbation you have to use "time-dependent perturbation theory" – which you haven't yet met – but at least

you should be ready to understand how the argument will go. From the theory it also follows that there are **selection rules** governing the 'allowed' and 'forbidden' transitions in diagrams such as Figure 17. Everything has an explanation!

Now we're almost ready to move to the last big question in this Chapter: "What makes the central field so special? But first you'll have to do a bit more mathematics –

Chapter 7

More on symmetry

7.1 Operators and matrix representations

If you turn back to Chapter 1 you'll be reminded of just how much came out of one very simple idea – that all directions in space are equivalent, it doesn't matter what directions your x-, y-, and z-axes point in. If you turn your system round in space, all its observable properties are unchanged. We used this principle to find all the basic properties of the operators S_x, S_y, S_z associated with the x-, y-, and z-components of the spin angular momentum of an electron! – using nothing more than the simplest possible experimental observations (the Stern-Gerlach experiment).

The properties we discovered were that (i) there are two **spin states**, which we called α and β and pictured as orthogonal vectors; and (ii) the spin operators act on α and β as follows

$$S_{x}\alpha = \frac{1}{2}\beta, \qquad S_{y}\alpha = \frac{1}{2}i\beta, \qquad S_{z}\alpha = \frac{1}{2}\alpha, S_{x}\beta = \frac{1}{2}\alpha, \qquad S_{y}\beta = -\frac{1}{2}i\alpha, \qquad S_{z}\alpha = -\frac{1}{2}\alpha.$$
(7.1)

The last group of equations can be written alternatively as

$$S_{x}(\alpha \beta) = (\alpha \beta) \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \quad S_{y}(\alpha \beta) = (\alpha \beta) \begin{pmatrix} 0 & -\frac{1}{2}i \\ \frac{1}{2}i & 0 \end{pmatrix},$$

$$S_{z}(\alpha \beta) = (\alpha \beta) \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix},$$
(7.2)

where the square arrays of numbers are the **Pauli matrices**. From now on, we'll denote these matrices by S_x, S_y, S_z :

 $\mathbf{S}_{\mathbf{x}} = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \ \mathbf{S}_{\mathbf{y}} = \begin{pmatrix} 0 & -\frac{1}{2}i \\ \frac{1}{2}i & 0 \end{pmatrix}, \ \mathbf{S}_{\mathbf{z}} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}.$ (7.3)

The arrays each have two rows and two columns and are said to be '2× 2' ('two-by-two') matrices. The **basis vectors** α and β have been collected into a single row, which forms a 1-row, 2-column (or 1×2 matrix).

In order to use matrices in this way we have to invent a rule for combining them: if two matrices are denoted by bold symbols, \mathbf{L} and \mathbf{M} say, then the **matrix product LM** is the array whose pq-element, standing in the qth place of the pth row, is given by

$$(\mathbf{LM})_{pq} = \sum_{k} L_{pk} M_{kq}.$$
(7.4)

(Note the usual convention: in saying the 'pq-element', the row number (p) is put first.) In words, (7.4) simply says "multiply the kth element in Row p of **L** by the kth element in Column q of **M** and sum the products over all values of k". You can do this, of course, only if the number of elements in any row p of the first matrix matches the number in any column q of the second matrix. (In that case the matrices are said to be *conformable*: and otherwise you can't define their product.) Let's try all this out by showing that (7.1) and (7.2) say exactly the same thing:

Example 7.1 Using the Pauli spin matrices

The first equation in (7.2) has $S_x(\alpha \beta)$ on the left, which is interpreted as $(S_x \alpha S_x \beta)$, the S_x standing on the left of each element in the row. The right-hand side of the equation has α multiplied by the first element (0) of the first column in S_x plus β multiplied by the second element $(\frac{1}{2})$. So $S_x \alpha = \frac{1}{2}\beta$, just as it was in (7.1). If you go ahead in that way for all the equations in (7.1) (and you should do it, for practice!) you'll find there is complete agreement between (7.1) and (7.2).

You'll be wondering why anyone would ever want to make a simple statement like (7.1) look more complicated by inventing other ways of saying the same things. But the payoff will come when you go a bit further! The next Example will start you off.

Example 7.2 A matrix representation

On using the rule (7.4) for matrix multiplication, with $\mathbf{S}_{\mathbf{x}}$ and $\mathbf{S}_{\mathbf{y}}$ in place of \mathbf{L} and \mathbf{M} , you find

$$\begin{aligned} (\mathbf{S}_{\mathbf{x}}\mathbf{S}_{\mathbf{y}})_{11} &= (\mathbf{S}_{\mathbf{x}})_{11}(\mathbf{S}_{\mathbf{y}})_{11} + (\mathbf{S}_{\mathbf{x}})_{12}(\mathbf{S}_{\mathbf{y}})_{21} = 0 + (\frac{1}{2})(\frac{1}{2}i) = \frac{1}{4}i, \\ (\mathbf{S}_{\mathbf{x}}\mathbf{S}_{\mathbf{y}})_{12} &= (\mathbf{S}_{\mathbf{x}})_{11}(\mathbf{S}_{\mathbf{y}})_{12} + (\mathbf{S}_{\mathbf{x}})_{12}(\mathbf{S}_{\mathbf{y}})_{22} = 0 + 0 = 0, \\ (\mathbf{S}_{\mathbf{x}}\mathbf{S}_{\mathbf{y}})_{21} &= (\mathbf{S}_{\mathbf{x}})_{21}(\mathbf{S}_{\mathbf{y}})_{11} + (\mathbf{S}_{\mathbf{x}})_{22}(\mathbf{S}_{\mathbf{y}})_{21} = 0 + 0 = 0, \\ (\mathbf{S}_{\mathbf{x}}\mathbf{S}_{\mathbf{y}})_{22} &= (\mathbf{S}_{\mathbf{x}})_{21}(\mathbf{S}_{\mathbf{y}})_{12} + (\mathbf{S}_{\mathbf{x}})_{22}(\mathbf{S}_{\mathbf{y}})_{22} = (\frac{1}{2})(-\frac{1}{2}i) + 0 = -\frac{1}{4}i. \end{aligned}$$

And if you now do the same calculation after changing the order of $\mathbf{S_x}$ and $\mathbf{S_y}$ (do it!) you'll find

$$(\mathbf{S}_{\mathbf{y}}\mathbf{S}_{\mathbf{x}})_{11} = -\frac{1}{4}i, \qquad (\mathbf{S}_{\mathbf{y}}\mathbf{S}_{\mathbf{x}})_{22} = \frac{1}{4}i,$$

the other elements still being zero.

Finally, by taking the difference of the two matrices obtained in this way, you get

$$\mathbf{S}_{\mathbf{x}}\mathbf{S}_{\mathbf{y}} - \mathbf{S}_{\mathbf{y}}\mathbf{S}_{\mathbf{x}} = i \begin{pmatrix} \frac{1}{2} & 0\\ 0 & -\frac{1}{2} \end{pmatrix} = i\mathbf{S}_{\mathbf{z}}$$

This result tells us that the matrices we have associated with the spin *operators* have exactly similar properties: the commutator $S_x S_y - S_y S_x$, which we found in (1.18) to be equivalent to $i \times S_z$, is echoed

in the matrix relationship

$$\mathbf{S_xS_y} - \mathbf{S_yS_x} = i\mathbf{S_z}.$$

And the same is true, as you can confirm, for the other pairs of operators.

Example 7.2 has shown how matrices can be associated with operators, provided we have a set of basis vectors (like the α and β of a **spin-space**) on which the operators have known effects: when these effects are expressed in matrix notation we say the matrices provide a **matrix representation** of the operators. If you look again at Chapter 2 it will be clear that the idea can be generalized to the case of operators working on the basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots \mathbf{e}_n$ of an *n*-dimensional vector space.

Suppose that in the general case some operator A works on a basis to produce a new set of vectors, $Ae_1, Ae_2, ... Ae_n$ and that these are linear combinations of the original set. For example, we might have

$$Ae_1 = e_1A_{11} + e_2A_{21} + e_3A_{31} \dots + e_nA_{n1}$$

as the expression for the first of the new vectors in terms of the original 'fixed' basis $\mathbf{e}_1 \, \mathbf{e}_2, \dots \mathbf{e}_n$. When all such expressions are collected the result is

$$Ae_{1} = e_{1}A_{11} + e_{2}A_{21} + e_{3}A_{31} \dots + e_{n}A_{n1}$$

$$Ae_{2} = e_{1}A_{12} + e_{2}A_{22} + e_{3}A_{32} \dots + e_{n}A_{n2}$$

$$Ae_{3} = e_{1}A_{13} + e_{2}A_{23} + e_{3}A_{33} \dots + e_{n}A_{n3}$$

$$etc.....$$

– where the A_{ij} are the given numerical coefficients of the original basis vectors in the new vectors Ae_i . The matrix **A** to be associated with the operator **A** is then the square array of numbers on the right in the following equation:

$$A(e_1 e_2 e_3 \dots) = (Ae_1 Ae_2 Ae_3 \dots) = (e_1 e_2 e_3 \dots) \begin{pmatrix} A_{11} & A_{12} & A_{13} & \dots \\ A_{21} & A_{22} & A_{23} & \dots \\ A_{n1} & A_{n2} & A_{n3} & \dots \end{pmatrix}$$

In this way, given a basis (which may be a set of vectors or a set of functions, the operators A, B, ..., X, ..., which act on them may be **represented** by matrices A, B, ..., M. We write $A \rightarrow A, B \rightarrow B, ...$ and, subject to certain mathematical conditions, the matrices 'echo' the properties of the operators: when, for example, AB = C then AB = C. This was the case with the spin operators S_x, S_y, S_z and the spin matrices S_x, S_y, S_z .

We're going to use matrix equations a lot in the rest of this book and don't always want to write them out in full, as in the example above. So let's agree to use a 'shorthand' form: just as we've used a single **boldface** letter **A** to denote a whole square array of numbers – like the one appearing above – we'll use a bold letter also for a whole array of similar *non-numerical* quantities, such as basis vectors. Thus, we can let **e** stand for the whole row matrix of basis vectors in the example above; and with this convention the equation shrinks to almost nothing. It becomes simply

$$\mathbf{A}\mathbf{e} = \mathbf{e}\mathbf{A},\tag{7.5}$$

and yet, as long as you recognize the symbols and know what they stand for, (7.5) means exactly the same as the full equation we started from! So you must examine each symbol carefully to be sure what it means:

Numbers, such as vector components, are usually set in 'italic' type; nonnumerical things, such as vectors and operators, are set in 'sans serif' type (with no 'frills' at the ends of the letters); and whole arrays of elements of either kind are set in bold type.

Matrix representations of the operators are found in all parts of quantum mechanics. Usually, we'll deal with them as we come to them, but first we note that rotations in 'everyday' 3-space are dealt with in just the same way:

Example 7.3 Matrix representation of rotations

In three dimensions we often use Cartesian coordinates by introducing orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ to define the three perpendicular axes. The **position vector** of any point P is then $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ or, in more convenient notation, $\mathbf{r} = r_1\mathbf{e}_1 + r_2\mathbf{e}_2 + r_3\mathbf{e}_3 = \sum_i r_i\mathbf{e}_i$. When basis vectors and components are labelled with numerical indices in this way, it is clear that a vector **r** pointing in any direction can be indicated as a row-column product

$$\mathbf{r} = r_1 \mathbf{e}_1 + r_2 \mathbf{e}_2 + r_3 \mathbf{e}_3 = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3) \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \mathbf{er},$$

where **r** on the right stands for the *column* of components r_1, r_2, r_3 relative to the basis **e**. (It *must* be a 3×1 matrix to match the 1×3 matrix **e** that stands on its left.)

Moreover, if we decide to use a new basis in which $\mathbf{e}_1 \rightarrow \bar{\mathbf{e}}_1 = l\mathbf{e}_1 + m\mathbf{e}_2 + n\mathbf{e}_3$ and similarly for the other basis vectors, then it's easy to do so. (You may remember that l, m, n are the **direction cosines**, defined long ago in Section 5.4 of Book 2.) When all three basis vectors are rotated into new ones we simply get

$$(\bar{\mathbf{e}}_1 \; \bar{\mathbf{e}}_2 \; \bar{\mathbf{e}}_3) = (\mathbf{e}_1 \; \mathbf{e}_2 \; \mathbf{e}_3) \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{pmatrix},$$

where each column in the square matrix *represents* one of the new basis vectors in terms of the 'old' e_1, e_2, e_3 . This change of basis is a **transformation** and its matrix is often denoted by the single symbol **T**. The last result (all of it!) is then written in the short form

 $\bar{\mathbf{e}} = \mathbf{eT},$

where \mathbf{T} is simply the array of numbers describing (through matrix multiplication) how the new vectors are expressed as linear combinations of the old.

Transformations of the kind introduced in Example 7.3 are very common in quantum mechanics. When the basis is **orthonormal**, its vectors being orthogonal and of unit length, the array of scalar products with elements $S_{ij} = (\mathbf{e}_i | \mathbf{e}_j) = 0$ $(j \neq i), = 1$ (j = i) (usually denoted by $S_{ij} = \delta_{ij}$) is not changed in such transformations. In the case where only *real* quantities are allowed, transformation matrices with this property are called

"orthogonal matrices". But when *complex* numbers are admitted the allowed transformations are more general: they are described by **unitary matrices**, for which we'll use the symbol **U**.

After a rotation R, two new basis vectors $\mathbf{e}_i, \mathbf{e}_j$ will become $\bar{\mathbf{e}}_i = \mathsf{R}\mathbf{e}_i = \sum_k \mathbf{e}_k U_{ki}$ and $\bar{\mathbf{e}}_j = \mathsf{R}\mathbf{e}_j = \sum_l \mathbf{e}_l U_{lj}$; here the coefficient U_{ki} is taken from column *i* of the matrix U and U_{lj} from column *j*. For an orthonormal basis, where the metric matrix **S** is the *unit* matrix, the rotated vectors will have a scalar product

$$\bar{S}_{ij} = (\bar{\mathbf{e}}_i | \bar{\mathbf{e}}_j) = \sum_k \sum_l U_{ki}^* (\mathbf{e}_k | \mathbf{e}_l) U_{lj} = \sum_k \sum_l U_{ik}^* S_{kl} U_{lj}$$

Now the matrix with elements U_{ki}^* may be obtained from that with elements U_{ik} by interchanging rows and columns, so $ik \to ki$, and attaching the 'star' to get the complex conjugate. This is an important operation, called "taking the **Hermitian transpose**", and is denoted by attaching a 'dagger' to the given matrix instead of the star. Thus, $U_{ki}^* = (\mathbf{U}^{\dagger})_{ik}$ and the full matrix \mathbf{U}^{\dagger} is obtained from \mathbf{U} by swapping corresponding rows and columns (e.g. first row of \mathbf{U} becomes first column of \mathbf{U}^{\dagger}), which is called "transposing" the matrix, and then taking the complex conjugate of every element. With this notation the equation for \bar{S}_{ij} is seen to come from a product of three matrices on the right-hand side, with the indices following the 'chain-rule' order $ik \to kl \to lj$; as a matrix equation it reads

$$\mathbf{\tilde{S}} = \mathbf{U}^{\dagger} \mathbf{S} \mathbf{U},\tag{7.6}$$

which defines a **unitary transformation**

When the transformation is from one *orthonormal* basis to another, both **S**-matrices become unit matrices and may be removed. The conditions which define any unitary matrix are thus

$$\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{1},\tag{7.7}$$

where **1** is the **unit matrix**. This means that \mathbf{U}^{\dagger} is also the **inverse** of **U**: $\mathbf{U}^{\dagger} = \mathbf{U}^{-1}$ and (7.6) thus implies also $\mathbf{U}\mathbf{U}^{\dagger} = \mathbf{1}$.

7.2 What makes the central field so special?

We'll be specially interested in rotations of a physical system, like an atom or a molecule, and even in Book 1 (Section 6.1) you met the idea of **symmetry**: if, when you turn an object round (e.g. by rotating it through 60° about some fixed axis – but without displacing it), the turned object appears to be unchanged, then it has some kind of 'symmetry'. There may be several **symmetry operations** – or even an infinite number – and when they satisfy certain conditions they form a **symmetry group**. The system studied in the last Chapter is defined by the central field, provided by an atomic nucleus placed at the origin of coordinates; and if an electron is put at Point P it doesn't 'see' any change in the field when the nucleus is rotated. The electron moves in a central field and that is what defines 'the system'. If, instead, you had *two* identical nuclei, on the x-axis and equidistant from the origin, any rotation of the pair around the z-axis would

change the field – except the 180° rotation that just interchanges the nuclei and leaves the system looking exactly as it was. Such rotations would be **symmetry operations**.

The fact that the hydrogen atom has degenerate sets of eigenfunctions, three p-functions with the same energy, five d-functions, seven f-functions, and so on, follows purely from the symmetry of the central field; you don't have to solve any differential equation or do any kind of calculation! Now we want to know why?

Let's denote a general rotation in 3-space by R and think about how it affects a vector $\mathbf{r} = r_1\mathbf{e}_1 + r_2\mathbf{e}_2 + r_3\mathbf{e}_3$. A rotation in space doesn't automatically change *numbers*, like r_1, r_2, r_3 ; it only moves *vectors* – like the basis vectors themselves. So if you want to say $\mathbf{r}' = \mathbf{Rr}$ is the rotated vector corresponding to \mathbf{r} you mean it's related to the rotated basis $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3$ in the same way that \mathbf{r} was related to $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ – which we think of as a standard 'fixed' basis. So, in matrix language, we have to say

$$(\mathbf{e}_1' \ \mathbf{e}_2' \ \mathbf{e}_3') = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3)\mathbf{R},$$
 (7.8)

where the square matrix \mathbf{R} has columns of components which relate the rotated basis vectors to the old; and then express the general vector \mathbf{r}' as

$$\mathbf{r}' = (\mathbf{e}'_1 \ \mathbf{e}'_2 \ \mathbf{e}'_3) \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3) \mathbf{R} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3) \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}.$$

All that looks much simpler, and easier to follow, if we use the shorter notation where a single symbol \mathbf{e} stands for the whole basis; and another, \mathbf{R} , for the whole rotation matrix. Thus

$$\mathbf{r}' = \mathbf{e}'\mathbf{r} = \mathbf{e}\mathbf{R}\mathbf{r} \tag{7.9}$$

and the components of the rotated vector \mathbf{r}' are therefore contained in the column matrix

$$\mathbf{r}' = \mathbf{R}\mathbf{r},\tag{7.10}$$

which gives the components \mathbf{r}' of the rotated vector \mathbf{r}' in terms of the fixed basis. (Remember we agreed that you have to look carefully at every symbol to fully understand what it means!)

When a basis is changed everything else is changed – not just the components of a vector – so we'll need to know also how the square matrices representing operators must be changed. In Section 6.6 we talked about **unitary transformations**, from one orthonormal basis to another, and used the notation

$$\mathbf{e}
ightarrow ar{\mathbf{e}} = \mathbf{e} \mathbf{U}$$

to denote this kind of change – even though it can be thought of as just another rotation. Now let's ask how the components of a *given* vector, which we think of as 'fixed', must be changed if we decide to change the basis.

Calling the fixed vector \mathbf{v} , it can be expressed either as

(i)
$$\mathbf{v} = \mathbf{e}\mathbf{v}$$
, or (ii) $\mathbf{v} = \bar{\mathbf{e}}\bar{\mathbf{v}}$:
it's the same vector, but has to be given new components when referred to the new basis. And now we can say, from (ii) that

$$v = \bar{e}\bar{v} = eU\bar{v}.$$

On comparing this result with (i), the new components are seen to be related to the old by $\mathbf{v} = \mathbf{U}\bar{\mathbf{v}}$. Alternatively, this equation can be 'turned round' by using (7.7), namely $\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{U}\mathbf{U}^{\dagger} = \mathbf{1}$ – for on multiplying both sides by \mathbf{U}^{\dagger} it follows that $\mathbf{U}^{\dagger}\mathbf{v} = \bar{\mathbf{v}}$ Evidently the rules for changing basis vectors and numerical components of a given vector are different.

When the basis is changed by a unitary transformation, to $\bar{\mathbf{e}} = \mathbf{e}\mathbf{U}$, the components \mathbf{v} of any fixed vector \mathbf{v} , must be changed to $\bar{\mathbf{v}} = \mathbf{U}^{\dagger}\mathbf{v}$. A 'fixed' vector is an example of an **invariant**: here we're talking only about different ways of *describing* it. The invariance shows up nicely when we write

$$\mathsf{v} = \mathbf{e}\mathbf{v} = \mathbf{e}\mathbf{U}\;\mathbf{U}^\dagger\mathbf{v} = ar{\mathbf{e}}\;ar{\mathbf{v}}$$

– whenever \mathbf{U} and \mathbf{U}^{\dagger} stand side-by-side they are equivalent to the unit matrix, which doesn't change anything and can be dropped.

Figure 18 summarizes, in pictures, everything we've been saying. In the left-hand picture a basis $\mathbf{e} = (\mathbf{e}_1 \ \mathbf{e}_2)$ is rotated through 30° around a perpendicular axis (\mathbf{e}_3 say) and is changed into $(\bar{\mathbf{e}}_1 \ \bar{\mathbf{e}}_2) = (\mathbf{e}_1 \ \mathbf{e}_2)\mathbf{U}$, where \mathbf{U} is a 2×2 unitary matrix. Any vector \mathbf{v} which is 'locked into the basis' and has components \mathbf{v} will be carried into a new vector \mathbf{v}' , with components $\mathbf{v}' = \mathbf{U}\mathbf{v}$. This describes an **active transformation**: all vectors move along with the basis.

The right-hand picture in Fig.18 shows what happens to a *fixed* vector – nothing! It is **invariant** against the change of basis: the components change only because we're describing the vector in terms of a new basis. This describes a **passive transformation** in which the same vector gets new components $\mathbf{U}^{\dagger}\mathbf{v}$ only because we've referred it to the new basis $\bar{\mathbf{e}}$.



Figure 18. Active and passive transformations (see text)

With these two transformation laws in mind it's easy to change from one representation of operations in a linear vector space to another. Suppose for example you have a representation in which three operators A, B, C have associated matrices A, B, C; so that a

relationship AB = C is 'echoed' in AB = C. And that now you decide to change the basis from \mathbf{e} , with which you set up the original matrices, to a new one $\mathbf{\bar{e}} = U\mathbf{e} = \mathbf{e}\mathbf{U}$ (supposing as usual that the bases are orthonormal, so \mathbf{U} is unitary). The question to be answered is: What new operators (and corresponding matrices will have to be used instead of A, B, C?

So far we've been thinking about operations in 3-space – the real 'everyday' space we live in, where we can visualize the vectors as arrows and the operators as any operation that turns them round. But, as you saw in Chapter 2, most of the things we're talking about can be defined and used just as well for a **linear vector space** in n dimensions. And the same is true even when the vectors are replaced by *functions*, of one or more variables, as you discovered long ago in Section 6.3 of Book 3. No real Mathematician would be happy to go ahead in this way, without worrying about all sorts of details and conditions; but once you've understood the main ideas you can come back later to all that! So let's suppose that everything will work out well (if you don't make any silly blunders), and put the last page or two in a slightly more general and abstract form.

Suppose then that a typical relationship among the operators is AB = C and think first about the action of A on any vector v in the space. Take the basis vectors as $\mathbf{e}_1, \mathbf{e}_2 \dots \mathbf{e}_n$, the whole set being symbolized by \mathbf{e} , and set up a new basis $\mathbf{\bar{e}} = U\mathbf{e} = \mathbf{eU}$. If the operator A sends vector v into $\mathbf{v}' = A\mathbf{v}$ then, after rotating the basis so that $\mathbf{e}_i \to \mathbf{\bar{e}}_i$, the new operator $\mathbf{\bar{A}}$ must change $\mathbf{\bar{v}}$ in a similar way. Thus $\mathbf{\bar{A}}\mathbf{\bar{v}} = \mathbf{\bar{v}}'$. But, since $\mathbf{\bar{v}} = U\mathbf{v}$ and $\mathbf{\bar{v}}' = U\mathbf{v}' = U\mathbf{Av}$, this means $\mathbf{\bar{A}}U\mathbf{v} = U\mathbf{Av}$.

This has to be true for *any* vector \mathbf{v} , so there must be an operator identity $\overline{A}U = UA$. Also, since for unitary operators $UU^{\dagger} = U^{\dagger}U = I$, multiplication of the last equation from the left by U^{\dagger} gives $U^{\dagger}\overline{A}U = A$; and similarly it follows (do it!) that $\overline{A} = UAU^{\dagger}$.

The same conclusion will be reached for all operators of the set A, B, C, ...M, ... so by defining the transform of each operator M of the set as

$$\overline{\mathsf{M}} = \mathsf{U}\mathsf{M}\mathsf{U}^{\dagger},\tag{7.11}$$

all relationships among operators and vectors will be conserved in the transformation. For example, when $\mathsf{AB}=\mathsf{C}$ we'll find

$$\overline{A}\overline{B} = UAU^{\dagger}UBU^{\dagger} = UABU^{\dagger} = UCU^{\dagger} = \overline{C}.$$

Equation (7.11) defines a **similarity transformation**. (Note that $U^{\dagger} = U^{-1}$ only for a transformation that conserves the unit metric: for a more general similarity transformation the U^{\dagger} in (7.11) must be replaced by U^{-1} .)

Now, at last, we can come back to the question asked at the beginning of Section 6.6: Why is the central field so important? And what follows from the fact that you can turn a central-field system round, about one fixed point (the nucleus), without making any noticeable change? The fact is that you can get the forms of all the angle-dependent wave functions, for all possible eigenstates, using symmetry arguments alone! But to do that you need quite a bit of **group theory**.

7.3 A first glimpse of the theory of groups

Back in Book 1 (Chapter 6 you first met some simple **Point Groups** (read about them again!). They are sets of operations which, when applied to a symmetrical object leave it looking exactly the same – although only one point is actually fixed. The group is *finite* if you can count the number of different operations (group 'elements') it contains. And to be a group, for a Mathematician, it must contain an **identity operation**, which does nothing to the object (i.e. leaves it as it was), and an **inverse** of every operation in the group. So if I denotes the identity and R a typical rotation then its inverse, denoted by R^{-1} , must have the property $RR^{-1} = R^{-1}R = I$. In addition, there must be an associative **law of combination** between elements of the group (called the "product"), such that (RS)T = R(ST) for any elements in the group. (This only means you can combine R and S first, as indicated by the parentheses, amd then combine the result with T; or combine S with T first and then combine the result with R; it mustn't make any difference – *as long as the sequence of the elements is not changed.*) Remember (Book 1) that the operations may or may not commute, so their *order* in a product is in general important.

The group of all possible rotations about an origin in 3-space, which is what interests us in talking about an atom, is *not* a finite point group: it contains an infinite number of rotations, through any angle and around any axis in space! Moreover. it is a **continuous group** in the sense that two distinct rotations may differ by an *infinitesimal* angle. The theory of continuous groups is difficult stuff even for professional mathematicians, but luckily we can do without it by using what you know about already.

The Hamiltonian operator for a central-field system with its nucleus fixing the origin of coordinates O is invariant against all rotations about O. If the unit vectors \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 define the x-, y- and z-axes before rotation, then the Hamiltonian for the rotated system will be set up with reference to a basis $\mathbf{e}' = \mathbf{R}\mathbf{e} = \mathbf{e}\mathbf{R}$, where \mathbf{R} is the 3×3 matrix describing the 3-space rotation R. But although the 'rotated Hamiltonian' – call it H' –is related to the 3-space rotation R it's not like an object that you can move around with the operator R. It's really a complicated 'recipe', containing partial differential operators and electric fields etc., and it works on wave functions. All you can say is that H' is the operator set up by the same recipe as H but with respect to the rotated basis ($\mathbf{e}'_1 \mathbf{e}'_2 \mathbf{e}'_3$); and the same is true of a 'rotated wave function' Ψ' . This is so important and general that it deserves a box:

A 'rotated Hamiltonian' H' is an operator set up by the same prescription as H but in the rotated coordinate frame produced by the 3-space operator R. The 'rotation' **induced** in the space of wave functions and operators, by the 3-space rotation R, will be denoted by U_R . With this convention we can write $\Psi' = U_R \Psi$ and $H' = U_R H U_R^{\dagger}$ for transformations that result from the rotation operation R in three dimensions

(7.12)

Now you know how to define 'rotated' operators and wave functions it's possible to talk about their properties. If R is a **symmetry operation** for some object in 3-space (e.g. an atom or a molecule) then the associated operator of the quantum system, namely U_R , will transform its Hamiltonian H into $H' = U_R H U_R^{\dagger}$ and if R is a symmetry operation then H and H' can't be distinguished: H' = H. Thus $H' = U_R H U_R^{\dagger} = H$ and on multiplying from the right by U_R this gives $U_R H = H U_R$. Thus

The Hamiltonian for a quantum system must commute with all the symmetry operations.

Consequently, if Ψ is an eigenstate of H with eigenvalue E then after a symmetry operation $\mathsf{H}\Psi = E\Psi$ becomes

$$\mathsf{U}_{\mathsf{R}}\mathsf{H}\Psi=\mathsf{H}\mathsf{U}_{\mathsf{R}}\Psi=E\mathsf{U}_{\mathsf{R}}\Psi.$$

Thus, the rotated Ψ , namely $\Psi' = U_R \Psi$, satisfies the same eigenvalue equation as before the rotation, with the same energy eigenvalue. It follows that

> Two eigenstates, which differ only through a symmetry operation, must be degenerate.

Example 7.4 Degeneracies of Hydrogen-atom eigenstates

Figure 15 showed energy-level diagrams for some s-, p-, d-, f-states of the Hydrogen atom. The energies of the s-states were calculated in Example 6.4 by solving a differential equation for R(r), the radial factor in the wave function, which gave the results in (6.10) for values up to n = 4 of the principal quantum number. The s-state wave functions were *spherically symmetrical*, being *invariant* under all operations of the rotation group. But the wave functions of other states contained angle-dependent factors and appeared in sets of 3 (for p states), 5 (for d states), 7 (for f states), and so on. The simplest angle-dependent wave functions are $\Psi(2p_x), \Psi(2p_y), \Psi(2p_z)$ and have the forms

$$R(r) \times x$$
, $R(r) \times y$, $R(r) \times z$,

respectively, R(r) being a common radial factor.

From the principles stated (in boxes) just before this Example, it follows that an s state wave function Ψ_s doesn't turn into anything new after a rotation R: $\Psi'_s = \mathsf{R}\Psi_s = \Psi_s$. So Ψ_s remains an energy eigenfunction with the same eigenvalue E_s : the set of functions obtained by rotation R contains only *one* function, which is said to be **non-degenerate**. The same is true for other s functions, with other values of the quantum number n: every set consists of only one eigenfunction and there is no reason for their energies to be the same.

On the other hand there are *three* p-state wave functions for any principal quantum number n, the $2p_z$ function having the form indicated in Fig.17, with 'partners' (not shown) having the same forms but pointing along the other two axes. Pictorially it is clear that these functions behave under rotations just like the basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$: rotations through 90° around any of the three axes, for example, send a function into $-1 \times$ a function of the same form, which is again an eigenfunction with the same energy

eigenvalue. More generally, *any* rotation around the origin produces a transformed function which is a linear combination of the three 2p functions: thus

$$\Psi_{2p_x}' = l\Psi_{2p_x} + m\Psi_{2p_y} + n\Psi_{2p_z}$$

where the coefficients l, m, n are the direction cosines of the rotated x-axis. To summarize, there are three linearly independent 2p functions, with the same energy eigenvalue E_{2p} , and they may be combined with arbitrary coefficients to give the most general 2p state, which is said to be '3-fold degenerate'.

So far, all seems to be in accord with the energy-level diagrams in Fig.15, but what about the d- and f-states? – which are not so easily related to the axial directions in 3-space. For instance, why are there five independent d-functions when we have only three axes in 3-space? To understand this you need to think more carefully about what the 'rotated function' means; the pictures are not always enough. The definition of $\Psi' = U_R \Psi$ is contained in (7.12). When we write $\Psi_{2p_x} = R(r) \times x$ we're talking about a *function of position* and x is the x-coordinate of the point at which it is evaluated. The x-coordinate of any point P (usually called a "field point") is the projection on the e_1 axis of the vector r which points from the origin to Point P, $x = (e_1|\mathbf{r})$. (Remember, from Chapter 2, that this definition allows for vectors whose components may be complex numbers!)

And now you see that the rotated function Ψ'_{2p_x} will be $\Psi'_{2p_x} = R(r) \times x'$, where $x' = (\mathbf{e}'_1 | \mathbf{r})$. Note that we're still thinking of the original point P in 3-space, but the rotated *function* is built around the rotated basis vector \mathbf{e}'_1 . In this way you can rotate *any* function in *any* direction – and you don't need any pictures to show the 'shapes' of the functions. In a later Example you'll see how this works for the d-functions; but here we'll do it generally. Let $\Psi(\mathbf{r})$ be any function of position, indicated by the position vector \mathbf{r} – which contains

the independent variables such as x, y, z. Then after a rotation R the new function, which will be denoted by $R\Psi$, will have the same numerical value at the 'image point P' (with position vector $\mathbf{r}' = \mathbf{R}\mathbf{r}$) as the original function had at the original point P. In symbols this says simply that

$$\Psi'(\mathsf{r}') = \Psi(\mathsf{r}) \text{ or } \mathsf{R}\Psi(\mathsf{R}\mathsf{r}) = \Psi(\mathsf{r}).$$

But this has to be true generally – for any point P – so let's look at a point with $R^{-1}r$ instead of r, where the inverse operator R^{-1} takes you to a 'backwards rotated' point. The last result then becomes $R\psi(RR^{-1}r) = \Psi(R^{-1}r)$, or

$$\mathsf{R}\Psi(\mathsf{r}) = \Psi(\mathsf{R}^{-1}\mathsf{r}). \tag{7.13}$$

In words, this defines the rotated function as the original function evaluated at the *backwards rotated field point*, allowing one to construct the rotated function however general the function may be.

Example 7.5 Effects of some particular rotations

A general 3-space rotation changes the basis according to (7.8), where the elements in the first column of the matrix **R** express the rotated vector \mathbf{e}'_1 in terms of the 'fixed' basis ($\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3$) and similarly for \mathbf{e}'_2 and \mathbf{e}'_3 . Let's choose **R** with elements $R_{21} = 1, R_{12} = 1, R_{33} = 1$ and all other elements zero: this will give

 $\mathbf{e}'_1 = \mathbf{e}_2$, $\mathbf{e}'_2 = \mathbf{e}_1$, $\mathbf{e}'_3 = \mathbf{e}_3$, which describes (check it!) a rotation of 90° around the \mathbf{e}_3 -axis. For the $2p_x$ function (see Example 6.4) the x-factor will become $x' = (e'_1|\mathbf{r}) = (\mathbf{e}_2|\mathbf{r}) = y$) and when this is substituted in Ψ_{2p_x} , with y in place of x, the result is $\Psi'_{2p_x} = \mathsf{U}_R\Psi_{2p_x} = \Psi_{2p_y}$. This confirms what we found in Example 6.4 using the pictures. Note, however, that the transformation operator has been written U_R – not just the ordinary 3-space rotation R – because it's the operator for an operation we've invented; like differential operators (and H in particular) – it works on wave functions in a **function space**. It just happens that the space based on $\Psi(2p_x), \Psi(2p_y), \Psi(2p_z)$ is 3-dimensional, but it's quite different from the space that holds geometric objects, which can be turned round by using operators like R. That's why we give the transformation operator a new name U_R and call it the "rotation *induced* by R".

When we pass to the d-functions the dimension of the function space must be changed. In fact there are *five* independent d-type wave functions with the same radial factor R(r) and these are usually taken to have the angle-dependent factors

$$\frac{x^2 - y^2}{r^2}, \quad \frac{3z^2 - r^2}{r^2}, \quad \frac{xy}{r^2}, \quad \frac{yz}{r^2}, \quad \frac{zx}{r^2}.$$

Two of them are shown in Fig.17 for principal quantum number n = 3. How do they behave under the symmetry operation we used for the p-functions at the beginning of this Example (rotation through 90° about the z-axis)? From the pictures it is clear that the $3d_{z^2}$ function is invariant, while $3d_{zx} \rightarrow 3d_{zy}$ and $3d_{zy} \rightarrow 3d_{zx}$ – are interchanged just like the 3-space vectors \mathbf{e}_1 and \mathbf{e}_2 . On the other hand $3d_{xy}$, which is like $3d_{zx}$ but lying mainly in the xy-plane, is simply multiplied by -1 because the 90° rotation interchanges the positive and negative 'lobes'. (You might have expected two more d-functions, with angle-dependent factors x^2/r^2 and y^2/r^2 , but three functions containing factors x^2 , y^2 , z^2 would not be linearly independent since there is a relationship $x^2 + y^2 + z^2 = r^2$.) You should read again the text following Figure 17.

The important fact is that the first two of the d-functions listed above are only multiplied by ± 1 under rotation through 90° about the z-axis; they do not mix with any of the last three. They remain eigenfunctions of the operator H, with which R commutes, and therefore must have the same energy eigenvalue E_{3d} – as long as the symmetry group contains only rotations around the z-axis. And the same is true for the last three eigenfunctions. However, the full symmetry group contains all rotations around the origin and in general they mix all five d-functions. Both sets of functions must accordingly correspond to the same eigenvalue. But if the central-field symmetry is in some way 'spoilt' and only the z-axis rotations remain true symmetry operations, then this will no longer be so: the set-of-two functions and the set-of-three, slightly changed by the spoilt symmetry, will no longer need to have the same energy. The five-fold degenerate set of energy levels can in this way be 'split' into a two-fold degenerate set and a three-fold degenerate. This effect is found in practice whenever a free atom is put into a field of lower symmetry, due for example to neighbouring atoms in a crystal, and is usually called **crystal field splitting**.

The last two Examples have shown the connection between the degeneracies of the Hydrogen atom eigenstates and the central-field (spherical) symmetry of the system. But we said earlier that even the forms of the angle-dependent wave functions could be derived using only symmetry arguments. Now we want to show *how*, but to do this you need to go a bit deeper into the theory of the rotation group. This is quite difficult so it's going into a separate Chapter, which you can try if you're curious, or skip if you want to keep moving.

Chapter 8

The three-dimensional rotation group

8.1 Representations in spin-space

In Chapter 1 we studied the Stern-Gerlach (SG) experiment, using α and β as basis vectors in a 'spin-space': they stood for states in which the spin component of a particle was 'up' or 'down' with respect to a z-axis in ordinary 3-space. (From now on we'll often use the term 'real space' to stand for the 3-space of our everyday experience.)

We've kept that notation throughout: α and β represent eigenstates of a spin operator S_z and define an orthonormal basis (the vectors being orthogonal and of unit 'length'). With the notation of Chapter 2, where the components of a general vector were first allowed to be complex numbers, that means

$$(\alpha | \alpha) = (\beta | \beta) = 1, \quad (\alpha | \beta) = (\beta | \alpha) = 0.$$

The spin operators S_x, S_y, S_z are related to the axial directions in 3-space (i.e. to basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$) and by considering rotations in space of a physical system we were able (in Chapter 2) to deduce their commutation properties. Now we must ask how the rotations affect the *spin*-space; and to simplify the algebra it's useful to work with matrices $\mathbf{s}_x, \mathbf{s}_y, \mathbf{s}_z$, with $\mathbf{s}_x = 2\mathbf{S}_x$ etc. instead of those defined in (7.3). Thus

$$\mathbf{s}'_x = \begin{pmatrix} 0 & 1\\ 1 & 0 \end{pmatrix}, \ \mathbf{s}'_y = \begin{pmatrix} 0 & -i\\ i & 0 \end{pmatrix}, \ \mathbf{s}'_z = \begin{pmatrix} 1 & 0\\ 0 & -1 \end{pmatrix},$$

where the $\frac{1}{2}$ s have all disappeared (but don't forget to put them back at the end!).

Example 8.1 Spin-space representations of the rotation group

$$(\alpha \ \beta) \rightarrow (\alpha' \ \beta') = (\alpha \ \beta) \mathbf{u}_{\mathsf{R}},$$

Any 3-space rotation R will induce a spin-space rotation, which can be described by a unitary matrix \mathbf{u}_{R} . (A lower-case letter u is used for this very basic 2×2 unitary matrix, upper case U being kept for the general $n \times n$ case.) Thus

where the basis vectors before and after the rotation are set out in 1-row matrices. The most general form of $\boldsymbol{u}_{\mathsf{R}}$ is

$$\mathbf{u}_{\mathsf{R}} = \left(\begin{array}{cc} a & b \\ -b^* & a^* \end{array} \right),$$

where the two complex numbers, a, b, must satisfy the condition $aa^* + bb^* = 1$. This can be checked by noting that the columns represent vectors of unit length with scalar product $a^*b - ba^* = 0$.

On rotating the system, s_x, s_y, s_z , being physical 'objects' like the 3-space vectors e_1, e_2, e_3 , should behave in the same way. They may therefore be collected into 3-element rows and will change according to (remember the new definition of the Pauli matrices, just before this Example)

$$(\mathsf{s}_x \mathsf{s}_y \mathsf{s}_z) \to (\mathsf{s}'_x \mathsf{s}'_y \mathsf{s}'_z) = (\mathsf{s}_x \mathsf{s}_y \mathsf{s}_z) \mathbf{U}_\mathsf{R},$$

where the 3×3 matrix \mathbf{U}_{R} is the one that transforms $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ under the 'real space' rotation R .

In Example 8.1 we defined the basic unitary transformation induced in spin-space by the physical rotation R in (7.8): it is

$$(\alpha \ \beta) \to (\alpha' \ \beta') = (\alpha \ \beta) \mathbf{u}_{\mathsf{R}},\tag{8.1}$$

where \mathbf{u}_{R} is the simple 2 × 2 matrix

$$\mathbf{u}_{\mathsf{R}} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix},\tag{8.2}$$

There are many ways of describing, in terms of matrices, the same 3-space R: each of them provides a **representation** of the group of *all* three-dimensional rotations. The group is *infinite* (containing an infinite number of elements) and *continuous* (for any given rotation one can find an infinite number of others that differ from it by only an **infinitesimal rotation**. The amazing thing is that all different representations of the rotation group can be obtained directly from the special unitary group (SUG) carried by the spin states α and β ; and that they all have a physical significance in terms of **angular momentum**. What we must do now is find how they are related to SUG and to the group of 'real-space' rotations in 3-space defined in (7.8).

Example 8.2 How are the representation matrices related?

$$\mathbf{s}'_x = \mathbf{u}_\mathsf{R} \mathbf{s}_x \mathbf{u}_\mathsf{R}^\dagger$$

. More fully the matrix product on the right (work it out for yourself!) gives the result

$$\mathbf{s}'_{x} = \begin{pmatrix} a & b \\ -b^{*} & a^{*} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a^{*} & -b \\ b^{*} & a \end{pmatrix} = \begin{pmatrix} ab^{*} + ba^{*} & a^{2} - b^{2} \\ -b^{*2} + a^{*2} & -b^{*}a - a^{*}b \end{pmatrix}.$$

Let's start from (8.1) and take the rotated spin operators one at a time. Remember that \mathbf{u}_{R} is the 2 × 2 matrix representing the *operator* \mathbf{u}_{R} which turns \mathbf{s}_x into \mathbf{s}_x' according to $\mathbf{s}_x' = \mathbf{u}_{\mathsf{R}}\mathbf{s}_x$ and that operations in spin-space are 'echoed' by similar operations involving the *matrices*. The rotated operator is therefore represented by the matrix $\mathbf{s}_x' = \mathbf{u}_{\mathsf{R}}\mathbf{s}_x$, where \mathbf{s}_x is the first of the Pauli matrices in (7.3); and after the rotation the first of the transformed Pauli matrices will be, by (7.11),

That's quite a 'messy' result: but as a and b are arbitrary we can choose any numbers we please and then ask what they mean. The easiest choice will be to make \mathbf{u}_{R} diagonal by putting b = 0 and setting $a = e^{i\theta}$, a general unimodular complex number. The matrix you just found then becomes simply

$$\left(\begin{array}{cc} 0 & a^2 \\ a^{*2} & 0 \end{array}\right) = \left(\begin{array}{cc} 0 & \cos 2\theta + i \sin 2\theta \\ \cos 2\theta - i \sin 2\theta & 0 \end{array}\right).$$

(If you don't remember much about complex numbers you can look back at Chapter 4 of Book 2.) The rotated spin operator s'_x is thus represented by the matrix

$$\mathbf{s}'_x = \cos 2\theta \begin{pmatrix} 0 & 1\\ 1 & 0 \end{pmatrix} + i \sin 2\theta \begin{pmatrix} 0 & -i\\ i & 0 \end{pmatrix} = \cos 2\theta \, \mathbf{s}_x - \sin 2\theta \, \mathbf{s}_y.$$

Now this corresponds to the 3-space rotation of basis vectors in which

$$\mathbf{e}_1' = \cos\phi \,\mathbf{e}_1 + \sin\phi \,\mathbf{e}_2$$

and tells us how the parameter θ must be interpreted as a 'real-space' rotation angle: we must choose $\theta = -\frac{1}{2}\phi$.

On the other hand the rotated spin operator s'_y will have a matrix

$$\mathbf{s}_{y}' = \cos 2\theta \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + i \sin 2\theta \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \cos 2\theta \,\mathbf{s}_{y} - \sin 2\theta \,\mathbf{s}_{x}$$

and – simplest of all (prove it!) – $s'_z = s_z$.

The full 3×3 matrix for a rotation through angle ϕ around the z-axis is thus, again with $\theta = -\frac{1}{2}\phi$,

$$\mathbf{R}_{z}(\phi) = \begin{pmatrix} 0 & \cos\frac{1}{2}\phi & 0\\ \sin\frac{1}{2}\phi & 0 & 0\\ 0 & 0 & 1 \end{pmatrix}.$$

All that can be repeated (try it now that you have a model to follow) for a rotation around the y-axis, which means that $\mathbf{e}'_2 = \mathbf{e}_2$ and $\mathbf{s}'_y = \mathbf{s}_y$. This means, in turn, that the parameters a and b in \mathbf{u}_{R} must be real numbers, since $a^2 + b^2 = 1$. So we can choose $a = \cos \theta$, $b = \sin \theta$ for any real angle θ . And when you repeat the above calculation you should find

$$\mathbf{s}_x' = \sin 2\theta \, \mathbf{s}_x + \cos 2\theta \, \mathbf{s}_z,$$

– where rotation around the y-axis mixes the basis vectors \mathbf{e}_1 , \mathbf{e}_3 . The y-axis stays as it was, so $\mathbf{s}'_y = \mathbf{s}_y$; and finally

$$\mathbf{s}_z' = \cos 2\theta \, \mathbf{s}_x + \sin 2\theta \, \mathbf{s}_z$$

The full 3×3 matrix for a rotation through angle ϕ around the y-axis is thus,

$$\mathbf{R}_{y}(\phi) = \begin{pmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{pmatrix}.$$

From Example 8.2 it follows that any 'real-space' rotation of a system, through angle ϕ around the z-axis (e_3), described by the matrix

$$\mathbf{R}_{z}(\phi) = \begin{pmatrix} 0 & \cos\frac{1}{2}\phi & 0\\ \sin\frac{1}{2}\phi & 0 & 0\\ 0 & 0 & 1 \end{pmatrix},$$
(8.3)

corresponds to a unitary transformation of the spin states ($\alpha \beta$) with the 2 × 2 matrix

$$\mathbf{u}_1(\phi) = \begin{pmatrix} a & 0\\ 0 & a^* \end{pmatrix}, \quad (a = e^{i\theta} = e^{-i\phi/2}).$$
(8.4)

And a similar rotation around the y-axis will be described by the matrix

$$\mathbf{R}_{y}(\phi) = \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix}$$
(8.5)

and will correspond to the spin-state transformation with

$$\mathbf{u}_2(\phi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \quad (a = \cos\frac{1}{2}\phi, b = \sin\frac{1}{2}\phi), \tag{8.6}$$

It will be noted that the rotation angle in spin-space is only *half* that in 'real space'; and that the rotated spin states $(\alpha' \beta')$ differ from those before the rotation only by **phase** factors, so their eigenvalues remain $\pm \frac{1}{2}$: e.g. under a z-axis rotation

$$\alpha \to \alpha' = e^{-\frac{1}{2}i\phi}\alpha, \quad \beta \to \beta' = e^{\frac{1}{2}i\phi}\beta.$$

We've now found a way of getting a 3×3 matrix **R** to describe *any* 'real-space' rotation **R** and a corresponding 2×2 unitary matrix \mathbf{u}_{R} that serves to define the same **R**. This tells us how to set up two **matrix representations** of the three-dimensional **rotation group**. To end this Section we need only one more short step: the matrices we found are very special – they are for rotations around the z-axis and the y-axis – but what about the x-axis and the infinite number of axes that point in other directions?

Hundreds of years ago the Swiss mathematician Leonhard Euler met a similar problem: If you have a sphere and mark any point P on its surface, then what is the smallest number of rotations needed to send P into any other point P' on the surface? And how can they be defined? The answer he found was "You need two perpendicular axes and three rotation angles, (α, β, γ) , say".

Example 8.3 The Euler angles of a rotation

Here we'll take the z-axis and the y-axis in our coordinate system as the two 'chosen axes' and, following Euler's prescription, we'll use the spin-space rotations described by (i) $\mathbf{u}_1(\alpha)$, given in (8.4), (ii) $\mathbf{u}_2(\beta)$, given in (8.6), and finally (iii) $\mathbf{u}_1(\gamma)$, again as in (8.4) but with $\phi = \gamma$. (Note that here α and β are just rotation angles (not spin states!) and that they refer to axes *fixed in space*.)

The rotation (i), followed by (ii), followed by (iii), will be denoted by $R(\alpha, \beta, \gamma)$ and will be described in spin-space by the product of corresponding 2 × 2 matrices:

$$\mathbf{u}(\alpha,\beta,\gamma) = \mathbf{u}_1(\gamma)\mathbf{u}_2(\beta)\mathbf{u}_1(\alpha),$$

where you read the matrices from right to left, as always, in the order of the operations they represent.

On putting in the explicit forms and doing the matrix multiplications it follows (if you have the patience to do it!) that

$$\mathbf{u}(\alpha,\beta,\gamma) = \begin{pmatrix} e^{-\frac{1}{2}i(\alpha+\gamma)}\cos\frac{1}{2}\beta & -e^{-\frac{1}{2}i(\alpha-\gamma)}\sin\frac{1}{2}\beta \\ e^{\frac{1}{2}i(\alpha-\gamma)}\sin\frac{1}{2}\beta & e^{\frac{1}{2}i(\alpha+\gamma)}\cos\frac{1}{2}\beta \end{pmatrix}.$$

In dealing with rotation groups, the matrices obtained in Example 8.1 are usually denoted by $\mathbf{D}_{1/2}(\alpha, \beta, \gamma)$, where D comes from the German word "Darstellung", for "representation", and the $\frac{1}{2}$ shows that it is carried by the spin states α , β (with half a unit of angular momentum). To summarize: with the (2×2) matrices defined in (8.4) and (8.6), but with angles $\phi = \alpha, \gamma$ in \mathbf{u}_1 and $\phi = \beta$ in \mathbf{u}_2 , the matrix representing $\mathsf{R}(\alpha, \beta, \gamma)$ will be

$$\mathbf{u}(\alpha,\beta,\gamma) = \mathbf{u}_1(\gamma)\mathbf{u}_2(\beta)\mathbf{u}_1(\alpha) \tag{8.7}$$

and this product now becomes

$$\mathbf{D}_{1/2}(\alpha,\beta,\gamma) = \mathbf{u}(\alpha,\beta,\gamma) = \begin{pmatrix} e^{-\frac{1}{2}i(\alpha+\gamma)}\cos\frac{1}{2}\beta & -e^{-\frac{1}{2}i(\alpha-\gamma)}\sin\frac{1}{2}\beta \\ e^{\frac{1}{2}i(\alpha-\gamma)}\sin\frac{1}{2}\beta & e^{\frac{1}{2}i(\alpha+\gamma)}\cos\frac{1}{2}\beta \end{pmatrix}.$$
 (8.8)

It's important to note the order of the rotations (α, β, γ) in defining the full rotation $\mathsf{R}(\alpha, \beta, \gamma)$ as a sequence of component rotations: the latter appear in reverse order, the α -rotation coming first (on the extreme right) as in any operator product. This applies also to the corresponding representation matrix, given in (8.7). The interpretation of the component rotations must also be kept in mind: the second rotation, through angle β , may be around a 'fixed' y-axis in space (the convention used here), or around a 'new' y-axis – in the system *after* the first rotation, through angle α . You always have to be careful when using conventions – otherwise you can get into trouble! (At least a dozen different conventions are used in defining the Euler angles, so if you find different books seem to disagree don't worry: decide on one convention and stick to it!) Many authors rotate the axes along with the physical system, getting different matrices; but – as Euler showed – it doesn't make any difference to the final result.

The representation $D_{1/2}$, with matrices given in (8.8) is the starting point for generating an infinite set of other representations, D_j with the index j being integral of half-integral. The next Section will show how.

8.2 The standard representations

Given the two spin-space eigenvectors α, β we can express any vector in this 2-space as a linear combination $\theta = \xi \alpha + \eta \beta$, where ξ, η are numerical components. Any real-space rotation $R(\alpha, \beta, \gamma)$ (for short let's just call it R) then induces a corresponding spin-space rotation in which the row of basis vectors ($\alpha \beta$) changes according to

$$(\alpha \beta) \to (\alpha' \beta') = (\alpha \beta) \mathbf{D}_{1/2}(\mathsf{R}).$$
 (8.9)

The components ξ , η of any vector θ , carried along with the basis, are then changed as follows (keep Figure 18 in mind – here, as usual, we're talking about an 'active' transformation):

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} \to \begin{pmatrix} \xi' \\ \eta' \end{pmatrix} = \mathbf{D}_{1/2}(\mathsf{R}) \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \tag{8.10}$$

where $\mathbf{D}_{1/2}(\mathsf{R})$ is the matrix defined in (8.8) in terms of the Euler angles of the rotation. For short, this matrix will be written as in (8.2), being the basic 2 × 2 unitary matrix:

$$\mathbf{u}_{\mathsf{R}} = \left(\begin{array}{cc} a & b \\ -b^* & a^* \end{array}\right),$$

When the components ξ , η change according to (8.10), their distinct products ξ^2 , $\xi\eta$, η^2 , will also be transformed. From (8.10), $\xi \to \xi' = a\xi + b\eta$ and $\eta \to \eta' = -b^*\xi + a^*\eta$ and it follows that

$$\begin{aligned} \xi^2 &\to (a\xi + b\eta)(a\xi + b\eta) &= a^2\xi^2 + b^2\eta^2 + 2ab\xi\eta \\ \xi\eta &\to (a\xi + b\eta)(-b^*\xi + a^*\eta) &= -ab^*\xi^2 + (aa^* - bb^*)\xi\eta + ba^*\eta^2 \\ \eta^2 &\to (-b^*\xi + a^*\eta)(-b^*\xi + a^*\eta) &= b^{*2}\xi^2 + a^{*2}\eta^2 - 2a^*b^*\xi\eta \end{aligned}$$

– linear combinations of the three distinct 'second-degree monomials', ξ^2 , $\xi\eta$, η^2 . If these monomials are interpreted as components of a vector in a 3-space they will turn into new linear combinations of themselves when the physical system is subjected to the rotation R. They will carry another representation of the rotation group. If we denote them by f_1 , f_2 , f_3 and set them out in a column, as we did with ξ and η in (8.10), they will transform under R according to

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \rightarrow \begin{pmatrix} f'_1 \\ f'_2 \\ f'_3 \end{pmatrix} = \begin{pmatrix} a^2 & 2ab & b^2 \\ -ab^* & (aa^* - bb^*) & ba^* \\ b^{*2} & -2a^*b^* & a^{*2} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}.$$
 (8.11)

Unfortunately the square matrix on the right, although it provides a representation of the rotation R, does not give us a *unitary* representation. However, this is easily put right. With the usual metric for vectors with complex elements you can confirm that the columns of the matrix in (8.11) correspond to vectors of unit length, provided you re-normalize the components by attaching suitable factors. To find the factors you can do it the hard way, taking each column one at a time for whatever matrix you're working on (and doing a lot of algebra); or you can do the job generally, for monomials of any degree N, as in the next Example.

Example 8.4 Representations of degree N

Let's consider any 2×2 unitary matrix (like \mathbf{u}_{R} for example) and choose ξ, η as the two elements in any column. The length² of the vector represented by this column will be $l^2 = \xi\xi^* + \eta\eta^* = 1$ for a unitary matrix. If you take the *N*th power of l^2 this will still be 1, but can be written as a (two-term) Binomial Expansion (see Book 3) in the form

$$(\xi\xi^* + \eta\eta^*)^N = \sum_{i,j(i+j=N)} \frac{(\xi\xi^*)^i(\eta\eta^*)^j}{i!j!} = \sum_{i,j(i+j=N)} \frac{(\xi^i\eta^j)}{\sqrt{i!j!}} \frac{(\xi^{*i}\eta^{*j})}{\sqrt{i!j!}}.$$

Next think of the double-index i, j as a single label, and call it m, standing for the i-j pair. The result above can then be written as _____

$$\sum_{m} f_m f_m^* = 1,$$

where f_m is a properly normalized monomial, of degree N, in the variables ξ, η . The sum over all components times their complex conjugates is just the Hermitian scalar product of a vector f with itself – giving its length² = 1. Explicitly, the components themselves are

$$f_m = \frac{(\xi^i \eta^j)}{\sqrt{i!j!}} \quad (i+j=N)$$

and, being associated with the basic unitary matrix \mathbf{u}_{R} , they carry a new representation of the real-space rotation group – just as in the derivation of (8.11). The representations constructed in this way are usually taken as the **standard representations**.

The monomials of degree N in the variables ξ and η , obtained in Example 8.4, can be written in a more symmetrical form by putting N = 2j and defining

$$f_m^{(j)} = \frac{\xi^{(j+m)}\eta^{(j-m)}}{\sqrt{(j+m)!(j-m)!}} \qquad (m=j, \, j-1, \, j-2, \, \dots -j)$$
(8.12)

Here the label m runs over values between j and -j, in integer steps, and every monomial is of degree (j + m) + (j - m) = 2j. There are clearly 2j + 1 terms of this kind and they will provide a 2j + 1-dimensional representation of the group of rotations. The first one, with j = 1, is specially important because it shows how the pair (ξ, η) may be related to the coordinates (x, y, z), which are the components of a position vector in ordinary 3-space.

At this point it's useful to summarize what we've done:

• We set up a **spin-space**, with basis vectors α and β , and described the effect of rotating the physical system (e.g. a single spin-half particle, moving in a central field) on an arbitrary vector $\xi \alpha + \eta \beta$, with numerical components ξ, η . With the **3-space** rotation R there was a corresponding 'rotation' of vectors in spin-space, described by the 2 × 2 unitary matrix \mathbf{u}_{R} in (8.2).

Under the rotation the basis vectors and components are transformed according to

$$(\alpha \ \beta) \rightarrow (\alpha' \ \beta') = (\alpha \ \beta) \mathbf{u}_{\mathsf{R}}, \qquad \mathbf{c} \rightarrow \mathbf{c}' = \mathbf{u}_{\mathsf{R}} \mathbf{c},$$

where the single column **c** contains the components $c_1 = \xi$ and $c_2 = \eta$.

• The rotation in real space is described in a similar way when referred to the 3-space basis vectors \mathbf{e}_1 , \mathbf{e}_2 , \mathbf{e}_3 ; the basis vectors change according to

$$(\mathbf{e}_1 \, \mathbf{e}_2 \, \mathbf{e}_3) \to (\mathbf{e}_1' \, \mathbf{e}_2' \, \mathbf{e}_3') \mathbf{U}_{\mathsf{R}},\tag{8.13}$$

(where the 3×3 matrix \mathbf{U}_{R} is unitary so as to preserve orthonormality), while the components of an arbitrary vector (contained in a single column c) change according to

$$\mathbf{c} \to \mathbf{c}' = \mathbf{U}_{\mathsf{R}} \mathbf{c}.\tag{8.14}$$

The elements of the matrix \mathbf{U}_{R} depend on the Euler angles of the rotation and can be worked out using only real-space geometry.

- What we have found is the connection between the two matrix representations of the same rotation R of the physical system: one, with matrices \mathbf{U}_{R} , in real space; the other, with matrices \mathbf{u}_{R} , in a strange 'spin-space' that we had to invent in order to describe the behaviour of a certain quantum system (the Stern-Gerlach experiment in Chapter 1).
- In general we say a representation is "carried by a set of basis vectors" and when the rotation R carries a set of vectors \mathbf{e}_i into a set of "images" \mathbf{e}'_i we write $\mathbf{e}' = \mathbf{R}\mathbf{e} = \mathbf{e}\mathbf{R}$. Here the bold letters stand for matrices, square for the one associated with the rotation R and a single row for the basis vectors, with \mathbf{R} standing on its right. But the same matrix describes the set of *components* (c_i) of any vector carried along with the basis; and in that case we write $\mathbf{c} \to \mathbf{c}' = \mathbf{R}\mathbf{c}$, with the matrix standing on the left of the column of components. That's why the monomials in, for example, (8.11) are collected in columns, being products of the components (ξ, η) of a vector in spin-space.

In what follows, we're going to invent many other 'strange' spaces, all of them based on the sets of monomials defined in (8.11) and with dimension 2j + 1. When $j = \frac{1}{2}$ we get the spin-space, with dimension 2j + 1 = 2; but taking j = 1 gives, according to (8.11) a space of dimension 3 carried by the second degree products of the vector components (ξ, η) :

$$f_{+1} = \frac{\xi^2}{\sqrt{2}}, \quad f_0 = \xi \eta, \quad f_{-1} = \frac{\eta^2}{\sqrt{2}}.$$
 (8.15)

These are the correctly normalized forms of the monomials used in (8.11) and you can confirm (do it!) that the labels m = +1, m = 0, m = -1 do indeed correspond to the m-values used in (8.12).

Just for the record (it's a lot of work to do it!) you can find the general 3×3 matrix representing the rotation $\mathsf{R}_{\alpha\beta\gamma}$, with Euler angles α, β, γ for the case j = 1. To get it you have to express the parameters a and b in terms of the angles (as in deriving $\mathbf{D}_{1/2}(\mathsf{R}_{\alpha\beta\gamma})$ starting from the properly normalized form of (8.11). This gives you

$$\mathbf{D}_{1}(\mathsf{R}_{\alpha\beta\gamma}) = \begin{pmatrix} \frac{1}{2}e^{-i(\alpha+\gamma)}(1+\cos\beta) & -\frac{1}{2}\sqrt{2}e^{-i\alpha}\sin\beta & \frac{1}{2}e^{-i(\alpha-\gamma)}(1-\cos\beta) \\ \frac{1}{2}\sqrt{2}e^{-i\gamma}\sin\beta & \cos\beta & -\frac{1}{2}\sqrt{2}e^{i\gamma}\sin\beta \\ \frac{1}{2}e^{i(\alpha-\gamma)}(1-\cos\beta) & \frac{1}{2}\sqrt{2}e^{i\alpha}\sin\beta & \frac{1}{2}e^{i(\alpha+\gamma)}(1+\cos\beta) \end{pmatrix}, \quad (8.16)$$

where (reading *left-to-right* along the rows, but *down* the columns) the entries are in descending order of m-values, (m = +1, 0, -1). The matrix representing *any* 3-space rotation follows on putting in values of the angles, but usually the cases of interest are very simple, one or two angles being zero.

For example, for rotation through angle γ about the z-axis, putting $\alpha = \beta = 0$ will give you all you need to in order to show the properties of the basis. In that particular case, the first column of $\mathbf{D}_1(\mathsf{R}_{\alpha\beta\gamma})$ will have elements $e^{-i\gamma}$, 0, 0 and if we denote the *basis vectors*, corresponding to the three monomials in (), by $\mathbf{e}_{+1}^{(1)}$, $\mathbf{e}_{0}^{(1)}$, $\mathbf{e}_{-1}^{(1)}$, the rotated vector $\mathsf{Re}_{+1}^{(1)}$ will be $\mathbf{e}_{+1}^{(1)} \times e^{-i\gamma} + \mathbf{e}_{0}^{(1)} \times 0 + \mathbf{e}_{-1}^{(1)} \times 0$. In the same way one finds that R leaves $\mathbf{e}_{0}^{(1)}$ unchanged, whilst multiplying $\mathbf{e}_{-1}^{(1)}$ by $e^{+i\gamma}$ (check it out!).

Now, although this is a 3-space, it is not defined using the usual Cartesian basis vectors (let's call them, for the moment, \mathbf{e}_x , \mathbf{e}_y , \mathbf{e}_z). In rotation $\mathsf{R}_z(\gamma)$ around the z-axis, through angle γ , the Cartesian unit vectors change as follows:

$$\mathbf{e}_x \to c\mathbf{e}_x + s\mathbf{e}_y, \qquad \mathbf{e}_y \to -s\mathbf{e}_x + c\mathbf{e}_y, \qquad \mathbf{e}_z \to \mathbf{e}_z,$$

where $c = \cos \gamma$, $s = \sin \gamma$. The two sets of basis vectors describe the same 'real space' rotation: they carry **equivalent representations** in which the vectors of each set can be expressed as linear combinations of those in the other set. You should verify, that they can be related by

$$\mathbf{e}_{+1}^{(1)} = -(\mathbf{e}_x + i\mathbf{e}_y)/\sqrt{2}, \quad \mathbf{e}_0^{(1)} = \mathbf{e}_z, \quad \mathbf{e}_{-1}^{(1)} = (\mathbf{e}_x - i\mathbf{e}_y)/\sqrt{2}.$$
 (8.17)

Note that the basis vectors are all of unit length, as well as being orthogonal. For example, $(\mathbf{e}_{+1}^{(1)}|\mathbf{e}_{+1}^{(1)}) = \frac{1}{2}((\mathbf{e}_1 - i\mathbf{e}_2)|(\mathbf{e}_1 + i\mathbf{e}_2)) = \frac{1}{2}(1 + 0 + 0 + 1) = 1$. (Check this, making sure you understand every step!) These linear combinations of the Cartesian basis vectors are called **spherical basis vectors**. The minus sign in the definition of $\mathbf{e}_{+1}^{(1)}$ is another example of a 'phase convention': you've met them in many places (often in connection with the 'step-up' and 'step-down' operators, or because of the difference between right-handed and left-handed coordinate systems). This choice of phase will be justified later.

As indicated in Item 3 following Example 8.4, the monomials listed in (5.2) are the functions of the variables (ξ, η) associated with the abstract vectors $\mathbf{e}_{+1}^{(1)}, \mathbf{e}_{0}^{(1)}, \mathbf{e}_{-1}^{(1)}$; and the correspondence in (8.17) applies also to the components of a general vector

$$\mathsf{V} = V_{+1}^{(1)} \mathsf{e}_{+1}^{(1)} + V_0^{(1)} \mathsf{e}_0^{(1)} + V_{-1}^{(1)} \mathsf{e}_{-1}^{(1)},$$

which can also be expressed as

$$\mathsf{V} = V_x \mathsf{e}_x + V_y \mathsf{e}_y + V_z \mathsf{e}_z$$

The spherical components of a vector V are thus related to its Cartesian components by

$$V_{+1}^{(1)} = -\frac{(x+iy)}{\sqrt{2}}, \qquad V_0^{(1)} = V_z, \qquad V_{-1}^{(1)} = \frac{(x-iy)}{\sqrt{2}}.$$
 (8.18)

In particular, when the vector V is the *position vector* of a point, the equations above give its spherical components in terms of the Cartesian components x, y, z. But these components are listed in (5.2) in terms of the variables ξ and η ; and comparison shows that

$$\xi^2 = -(x+iy), \qquad \xi\eta = z, \qquad \eta^2 = (x-iy).$$
 (8.19)

This is the fundamental correspondence we need for getting the angle-dependent factors in the eigenfunctions of any central-field system, as you'll see in the next Example.

Example 8.5 The p- and d-orbitals

First look back at Section 6.4, where you met the simplest angle-dependent wave functions for the hydrogen atom – the p and d orbitals, which corresponded to angular momentum quantum number L = 1, 2, repectively. Remember also how we got the basis functions for a representation of the rotation group with j = l = 1, listed in (5.2). These are the correctly normalized second-degree products of the components (ξ, η) , as defined generally in (8.11), and they span a space of dimension 2j + 1 = 3. The three basis functions in that space are shown again below, with a superscript (l = 1) indicating the orbital angular momentum quantum number:

$$f_{+1}^{(1)} = \frac{\xi^2}{\sqrt{2}}, \quad f_0^{(1)} = \xi\eta, \quad f_{-1}^{(1)} = \frac{\eta^2}{\sqrt{2}}.$$

The p- and d-functions given in Section 6.4 were expressed in terms of Cartesian coordinates, which show most clearly their geometrical forms; but in talking about their behaviour under rotations we have been using the ξ, η variables. Now that we know from (8.21) how the two sets of variables are related, it is easy to express the p-functions listed above in terms of Cartesian coordinates. We simply substitute

$$\xi^{2} = -(x + iy), \qquad \xi\eta = z, \qquad \eta^{2} = (x - iy)$$

and find at once

$$\begin{aligned} f_{+1}^{(1)} &= -\frac{(x+iy)}{\sqrt{2}} \\ f_{0}^{(1)} &= z, \\ f_{-1}^{(1)} &= \frac{(x-iy)}{\sqrt{2}}. \end{aligned}$$

What about the d-functions? They were mentioned briefly in Section 6.4 and it seemed strange that only *five* of them were needed, although there are six second-degree products of the variables x, y, z. Now things are clearer.

On putting l = 2 as an orbital angular momentum quantum number we'll be thinking of eigenfunctions that carry the representation D₂ of the rotation group. This has dimension 2l + 1 = 5 and from (8.12) it follows that the monomials f_m of degree 2j = 4 are

$$f_{+2} = \frac{\xi^4}{\sqrt{4!}}, \quad f_{+1} = \frac{\xi^3 \eta}{\sqrt{3!}}, \quad f_0 = \frac{\xi^2 \eta^2}{\sqrt{2}\sqrt{2}}, \quad f_{-1} = \frac{\xi \eta^2}{\sqrt{3!}}, \quad f_{-2} = \frac{\eta^4}{\sqrt{4!}}$$

When these are used as basis functions the five of them are sufficient to describe *all* the transformations arising from ordinary rotations of the system in *three*-space. Any other combinations of the five energy eigenfunctions to which they lead will be redundant (not needed at all!) and may be 'thrown away'. The question now is how to find five *linearly independent* mixtures and to express them in terms of the Cartesian coordinates. This can be done as in the case of the p-functions. The first one, for example, is

$$f_{+2}^{(2)} = \frac{\xi^4}{\sqrt{4!}} = \frac{(\xi^2)(\xi^2)}{2\sqrt{6}} = \frac{(x+iy)^2}{2\sqrt{6}} = \frac{(x^2 - y^2 + 2ixy)}{2\sqrt{6}},$$

while the second is

$$f_{+1} = \frac{\xi^3 \eta}{\sqrt{3!}} = \frac{(\xi^2)(\xi\eta)}{\sqrt{6}} = \frac{-(x+iy)z}{\sqrt{6}}.$$

Example 8.5 has shown how the basis functions for any representation D_j of the rotation group can be obtained in terms of the Cartesian coordinates x, y, z.

On the other hand, direct solution of the Schrödinger equation, using polar coordinates, leads to a wave function of the **separated form** given in (6.13), namely $\Psi(r, \theta, \phi) = R(r)Y(\theta, \phi)$. In Example 6.4 we found how to determine the radial factor R(r), which gives the spherically symmetric (s-type) functions with $Y(\theta, \phi) = 1$. But now we want to find the **spherical harmonic** factors $Y(\theta, \phi)$ for the p- and d-states.

To express the p-functions, given above in Cartesian form, in terms of the angles θ, ϕ , we need to use (see Fig.16)

$$x = r \sin \theta \cos \phi, \ y = r \sin \theta \sin \phi, \ z = r \cos \phi.$$

On substituting these values for x, y, z, we find for the first p-function

$$r^{-1}f_{+1}^{(1)} = -\frac{\sin\theta(\cos\phi + i\sin\phi)}{\sqrt{2}} = -\frac{\sin\theta(e^{i\phi})}{\sqrt{2}}$$

where you've remembered that $\cos \phi + i \sin \phi = e^{i\phi}$.

This is the spherical harmonic $Y_{l,m}(\theta, \phi)$ for the wave function with quantum numbers l = 1, m = 1. You can do the same (do it!) for $f_0^{(0)}$ and $f_{-1}^{(1)}$, to obtain the harmonics $Y_{1,0}$ and $Y_{1,-1}$.

More generally, the energy eigenfunctions of the Hamiltonian operator H in (6.1) would follow as solutions of Schrödinger's equation (6.2), but in Section 6.4 we didn't even try to find the angle-dependent solutions because the operator ∇^2 is too complicated to handle without knowing quite a lot about partial differential equations. Now, after learning about the symmetry properties of functions, you know how to do it another way.

Take next the case of the functions $f_m^{(l)}$ with l = 2. They will carry one form of the representation D₂, and will serve to describe how the wave functions behave under rotations. To express them in terms of the angles θ, ϕ , defined in Figure 16, we do as we did for the p-type functions with l = 1, substituting

$$x = r\sin\theta\cos\phi, \ y = r\sin\theta\sin\phi, \ z = r\cos\theta$$

in the functions $f_{+2}^{(2)}$ and $f_{+1}^{(2)}$ obtained in Example 8.5: namely

$$f_{+2}^{(2)} = \frac{(x^2 - y^2 + 2ixy)}{2\sqrt{6}}, \qquad f_{+1}^{(2)} = \frac{-(x + iy)z}{\sqrt{6}}.$$

In terms of the angles θ, ϕ , we have $x^2 - y^2 = r^2 \sin^2 \theta (\cos^2 \phi - \sin^2 \phi)$ and $xy = r^2 \sin^2 \theta (\cos \phi \sin \phi)$; so, if you remember your geometry from Section 4 of Book 2, you can rewrite these as

$$x^{2} - y^{2} = r^{2} \sin^{2} \theta \cos 2\phi, \qquad xy = \frac{1}{2}r^{2} \sin^{2} \theta \sin 2\phi$$

On substituting in $f_{+2}^{(2)}$ this gives

$$r^{-2}f_{+2}^{(2)} = \sin^2\theta(\cos 2\phi + i\sin 2\phi) = \sin^2\theta e^{2i\phi}$$

and in the same way

$$r^{-2}f_{+1}^{(2)} = -\sin\theta(\cos\phi + i\sin\phi) = -\sin\theta e^{i\phi}.$$

These functions have 'partners', for given |m|-values, obtained by changing the sign of *i*. The spherical harmonics $Y_{l,m}(\theta, \phi)$ follow on removing the radial factor $(r^{-1} \text{ or } r^{-2})$. For the functions listed above they are

$$Y_{1,0} = N_1 \cos \theta,$$

$$Y_{1,\pm 1} = \mp N_1 \sin \theta e^{\pm i\phi}$$

$$Y_{2,0} = N_2 (3\cos^2 \theta - 1)/2,$$

$$Y_{2,\pm 1} = \mp N_2 (\sqrt{3}/2) \cos \theta \sin \theta e^{\pm i\phi}$$

$$Y_{2,\pm 2} = N_2 (\sqrt{3/8}) \sin^2 \theta e^{\pm 2i\phi},$$

where N_1, N_2 are normalizing factors for functions of given l, and are defined generally as $N_l = \sqrt{(2l+1)/4\pi}$.

(You needn't worry about the numerical factors, which just ensure that the angle-dependent factors in the wave functions give unity when $|\Psi|^2$ is integrated over all values of θ, ϕ .).

Note that the wave functions, when written using polar coordinates and the 'separated form' (6.13), are often *complex*. But the p- and d-type wave functions used in Section 6.4 were $p_x = f(r)x$, $p_y = f(r)y$, $p_z = f(r)z$ etc. where f(r) depends only on the radial distance r. Evidently these Cartesian forms must be expressible as *mixtures* of the complex functions; and the right mixtures are clear from Example 8.5, where the Cartesian functions, indeed, appeared first. Thus, as you can easily verify (with a bit of algebra!),

$$p_{z} = R(r) \times N_{1}rY_{(l,0)}$$

$$p_{x} = R(r) \times N_{1}r(-Y_{(1,1)} + Y_{(1,-1)})/\sqrt{2},$$

$$p_{y} = R(r) \times iN_{1}r(Y_{(1,1)} + Y_{(1,-1)})/\sqrt{2},$$

$$d_{z^{2}} = R(r) \times N_{2}r^{2}Y_{(2,0)}$$

$$d_{x^{2}-y^{2}} = R(r) \times N_{2}r^{2}(-Y_{(2,2)} + Y_{(2,-2)})/\sqrt{2},$$

$$d_{xy} = R(r) \times iN_{2}r^{2}(-Y_{(2,2)} + Y_{(2,-2)})/\sqrt{2},$$

$$d_{yz} = (r) \times iN_{2}r^{2}(Y_{(2,1)} + Y_{(2,-1)})/\sqrt{2},$$

$$d_{xz} = R(r) \times N_{2}r^{2}(-Y_{(2,1)} + Y_{(2,-1)})/\sqrt{2}.$$
(8.20)

Here the radial factor R(r) is common to all the wave functions of given type (p or d), while the angle-dependent factor after the '×' contains the spherical harmonics. But note

that an extra factor (of r or r^2) is present in the p- or d-type harmonics. This means that the differential equation we solved in Section 6.2, to find the energy levels, will change as R(r) is replaced by rR(r) or $r^2R(r)$. As a result, the energies of the p and d states will, in general, depend on the principal quantum number n: only for a strictly Coulomb field will s, p and d states in the same shell have the same energy (as indicated in Fig.16). That's why, in Chapter 6, this kind of degeneracy was called "accidental".

You may be wondering if the l and m labels in the angle-dependent wave functions are {really angular momentum quantum numbers, for which we used L and M in earlier Sections. In fact they *are*: there's a very deep connection which goes down to the roots of quantum mechanics. It's worth spending a few pages to bring in one more very fundamental idea

8.3 The infinitesimal operators

Any rotation can be described as the result of a series of *infinitesimal* rotations, so it should be possible to study the rotation group in terms of the **infinitesimal operators** which define vanishingly small rotations about given axes.

Let's take a small rotation, about the z-axis in real 3-space, through an angle $\delta\phi$. This changes the Cartesian basis vectors according to

$$\mathsf{R}_{z}(\delta\phi)(\mathsf{e}_{1}\,\mathsf{e}_{2}\,\mathsf{e}_{3}) = (\mathsf{e}_{1}\,\mathsf{e}_{2}\,\mathsf{e}_{3}) \left[\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) + \delta\phi \left(\begin{array}{ccc} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right) + \mathcal{O}(\delta\phi^{2}) \right]$$

where $\mathcal{O}(\delta\phi^2)$ stands for terms of order $\delta\phi^2$ and higher, which will vanish when $\delta\phi \to 0$ and will be thrown away from now on.

The first matrix on the right describes the unit operator 1, while the second describes a tiny rotation $\delta\phi$ around the z-axis. The equation may thus be written $\mathsf{R}_z(\delta\phi) = 1 + \delta\phi\mathsf{D}_z$ where D_z is the infinitesimal operator represented by the second matrix. Usually, however, it's more convenient to work with *Hermitian* matrices – and we can get one by defining $\mathsf{J}_z = i\mathsf{D}_z$, which corresponds to $\mathsf{J}_z = i\mathsf{D}_z$. On doing the same for the other two axes we obtain three *unitary* matrices:

$$\mathbf{J}_{x} = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{J}_{y} = i \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \mathbf{J}_{z} = i \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
(8.21)

These matrices give us a representation of the infinitesimal operators J_x , J_y , J_z , so if we know how they combine under matrix multiplication we can write down similar rules for combining the operators. We'll try this idea out out in the next Example.

Example 8.5 Rules for combining infinitesimal rotations

Take the first two matrices in (8.21) and form the matrix product $\mathbf{J}_x \mathbf{J}_y$. For example, the 21-element will be the sum of products $(\mathbf{J}_x)_{2k} \times (\mathbf{J}_y)_{k1}$ as k runs along Row 2 of \mathbf{J}_x and down Column 1 of \mathbf{J}_y : this gives $i^2 \times (0^2 + 0^2 + -1^2) = -1 \times 1 = -1$, so $(\mathbf{J}_x \mathbf{J}_y)_{21} = -1$ and this is the same as $i(\mathbf{J}_z)_{21}$. But if you change the order of \mathbf{J}_x and \mathbf{J}_y the result is seen to be zero. And if you do all that for all pairs of rows and columns (you'll get quicker as you go on!) the result you'll find is that the elements of the difference $\mathbf{J}_x \mathbf{J}_y - \mathbf{J}_y \mathbf{J}_x$ exactly match those of the single matrix $i\mathbf{J}_z$:

$$\mathbf{J}_x \mathbf{J}_y - \mathbf{J}_y \mathbf{J}_x = i \mathbf{J}_z.$$

The Example gave the first of three results which you'll surely recognise:

$$\begin{aligned}
\mathbf{J}_x \mathbf{J}_y - \mathbf{J}_y \mathbf{J}_x &= i \mathbf{J}_z \\
\mathbf{J}_y \mathbf{J}_z - \mathbf{J}_z \mathbf{J}_y &= i \mathbf{J}_x \\
\mathbf{J}_z \mathbf{J}_x - \mathbf{J}_x \mathbf{J}_z &= i \mathbf{J}_y
\end{aligned}$$
(8.22)

- they are the commutation rules for angular momentum! But they came out simply from the properties of ordinary '3-space' rotations and are thus more primitive than the rules we found in quantum mechanics, using the position and momentum operators in Schrödinger's representation. Since the matrices provide a representation of the rotations, the operators satisfy the same rules:

$$J_x J_y - J_y J_x = i J_z$$

$$J_y J_z - J_z J_y = i J_x$$

$$J_z J_x - J_x J_z = i J_y.$$
(8.23)

Now perhaps you remember from long ago (Book 2, Chapter 4) that infinitesimal rotations can be combined in an infinite sequence to describe a rotation through any finite angle ϕ . Here, at the beginning of this Section, we studied the effect of the operator $\mathsf{R}_z(\delta\phi) =$ $1 + \delta\phi\mathsf{D}_z$ on the basis vectors $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3$. If we divide ϕ into n small steps, $\delta\phi = \phi/n$, the z-axis rotation can be performed as a sequence of n repetitions of $(1 - i\frac{\phi}{n}\mathsf{J}_z)$. On letting $n \to \infty$, the result will be

$$\mathsf{R}_{z}(\phi) = \lim_{n \to \infty} \left(1 - i\frac{\phi}{n} \mathsf{J}_{z} \right)^{n} = \exp(-i\phi \mathsf{J}_{z}), \tag{8.24}$$

where the exponential of the operator means the series

$$\exp(-i\phi \mathsf{J}_z) = 1 - i\phi \mathsf{J}_z + \frac{1}{2}\phi^2 \mathsf{J}_z^2 + \dots .$$
(8.25)

The same argument applies to rotations around other axes and a general rotation, with Euler angles α, β, γ can therefore be expressed in terms of the infinitesimal operators in the form

$$\mathsf{R}(\alpha,\beta,\gamma) = \exp -i\alpha \mathsf{J}_z \exp -i\beta \mathsf{J}_y \exp -i\gamma \mathsf{J}_z.$$
(8.26)

The J-operators we're talking about all describe infinitesimal rotations in the 'real space' that carries the matrix representation D_j with j = 1; and the matrices for finite rotations can be constructed simply by putting \mathbf{J}_z and \mathbf{J}_y , in place of the operators, in the last equation. The result will be, after doing the matrix multiplications, $\mathbf{D}_1(\mathsf{R}_{\alpha\beta\gamma})$ as given in (8.16).

At first sight it seems strange that so much of quantum mechanics can be obtained just by thinking about turning things round in ordinary space! So let's check it out using the spin-space, with basis vectors α , β . For example, an infinitesimal rotation through $\delta\phi$ around the z- axis in real space induces a spin-space rotation in which, according to (8.16), $\alpha \to e^{-i\delta\phi/2}\alpha \approx (1 - \frac{1}{2}i\delta\phi)\alpha$ and $\beta \to e^{+i\delta\phi/2}\beta \approx (1 + \frac{1}{2}i\delta\phi)\beta$. This may be written

$$(\alpha \ \beta) \to (\alpha \ \beta) \left[\left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) + i\delta\phi \left(\begin{array}{cc} -\frac{1}{2} & 0 \\ 0 & +\frac{1}{2} \end{array} \right) + \dots \right]$$

– with neglect of terms of order $\delta \phi^2$ and higher. Now *i* times the coefficient of $\delta \phi$ is the matrix representing the infinitesimal operator that we have denoted by J_z . A similar analysis applied to the other infinitesimal operators gives the complete set in which

$$\mathsf{J}_x \to \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \quad \mathsf{J}_y \to \begin{pmatrix} 0 & -\frac{1}{2}i \\ \frac{1}{2}i & 0 \end{pmatrix}, \quad \mathsf{J}_z \to \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}.$$
(8.27)

So now you see the Pauli matrices in a new light! They represent the infinitesimal rotations in spin-space induced by those in real space around the three coordinate axes. They can be written

$$\mathbf{S}_{x}^{1/2} = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} = \mathbf{J}_{x}^{1/2}, \quad \mathbf{S}_{y}^{1/2} = \begin{pmatrix} 0 & -\frac{1}{2}i \\ \frac{1}{2}i & 0 \end{pmatrix} = \mathbf{J}_{y}^{1/2}, \quad \mathbf{S}_{z}^{1/2} = \begin{pmatrix} \frac{1}{2}i & 0 \\ 0 & -\frac{1}{2}i \end{pmatrix} = \mathbf{J}_{z}^{1/2},$$
(8.28)

where a superscript (1/2) has been added to show that the matrices belong to the representation $D_{1/2}$ associated with a spin- $\frac{1}{2}$ system. These labels are not always needed, because the J-operators and their representative matrices have the same commutation properties (8.23) for all the representation spaces D_j (of dimension 2j + 1) to which they refer.

You should note, however, that the basis vectors which carry the reprentations have important j-dependent properties. In particular, the real-space rotation $\mathsf{R}_z(\phi)$ induces a rotation in the space of D_j , with basis vectors $\mathbf{e}_m^{(j)}$. For rotations through angle ϕ around the z-axis, the basis vectors of D_j are simply multiplied by phase factors $\exp -im\phi$, the label m going from m = j down to m = -j: thus, $\mathsf{R}_z(\phi) \, \mathsf{e}_m^{(j)} = \exp -im\phi \, \mathsf{e}_m^{(j)}$.

On expressing $R_z(\phi)$ in terms of the infinitesimal operator J_z this gives

$$(1 - i\phi J_z + \dots) \mathbf{e}_m^{(j)} = (1 - im\phi + \dots) \mathbf{e}_m^{(j)}$$

and term-by-term comparison of the two sides of this equation shows that $J_z e_m^{(j)} = m e_m^{(j)}$. In other words, each basis vector of D_j represents an eigenstate of the infinitesimal operator J_z with eigenvalue equal to the state label m. The properties of the J-operators seem to

be just the same as those of the angular momentum operators, studied in the first three Examples of Chapter 6, where they were denoted by L_x, L_y, L_z with M as the eigenvalue of L_z . But here the values of the index j include $j = \frac{1}{2}$, corresponding to *spin* angular momentum.

We can continue by looking at the effect of rotations about the other two axes; and those of the corresponding rotations induced in the space of D_j . It's a fairly long story, but the results are easily summarized: the infinitesimal operators J_x and J_y work on $\mathbf{e}_m^{(j)}$ to give mixtures of the basis vectors 'on either side' (with $m \to m \pm 1$).

In full

$$\begin{aligned} \mathsf{J}_{x}\mathsf{e}_{m}^{(j)} &= \frac{1}{2}[(j+m+1)(j-m)]^{1/2}\mathsf{e}_{m+1}^{(j)} + \frac{1}{2}[(j-m+1)(j+m)]^{1/2}\mathsf{e}_{m-1}^{(j)} \\ \mathsf{J}_{y}\mathsf{e}_{m}^{(j)} &= -\frac{1}{2}i[(j+m+1)(j-m)]^{1/2}\mathsf{e}_{m+1}^{(j)} + \frac{1}{2}i[(j-m+1)(j+m)]^{1/2}\mathsf{e}_{m-1}^{(j)} \end{aligned}$$

and this is true whichever representation space we're using (which is why the J-operators have not been given a superscript j).

It is also possible to define operators with simpler properties, namely $J^+ = J_x + iJ_y$ and $J^- = J_x - iJ_y$. These 'shift operators' are like the step-up and step-down operators in Chapter 6: they act on $\mathbf{e}_m^{(j)}$ to shift the basis vector into the one with index $m \to m \pm 1$. Thus

$$\begin{split} \mathsf{J}^+ \mathsf{e}_m^{(j)} &= [(j+m+1)(j-m)]^{\frac{1}{2}} \mathsf{e}_{m+1}^{(j)}, \\ \mathsf{J}^- \mathsf{e}_m^{(j)} &= [(j-m+1)(j+m)]^{\frac{1}{2}} \mathsf{e}_{m-1}^{(j)}. \end{split}$$

You're probably beginning to guess what comes next.

If you set up an operator $J^2 = J_x^2 + J_y^2 + J_z^2$ it will turn out to behave just like the angular momentum operator L^2 . In fact the basis vectors of the representation D_j are eigenvectors of both J_z and J^2

$$\mathbf{J}^{2}\mathbf{e}_{m}^{(j)} = j(j+1)\mathbf{e}_{m}^{(j)}, \quad \mathbf{J}_{z}\mathbf{e}_{m}^{(j)} = m\mathbf{e}_{m}^{(j)} \quad (m=j,j-1,\,\dots\,-j).$$
(8.29)

The remarkable thing about all this is that we've not done any Physics! It all came out of a study of rotations in ordinary 3-dimensional space and the results were derived using only **geometry**!

To end this Chapter let's go back to the central-field problem and the Schrödinger equation.

8.4 The connection with Physics

When we first worked on the central field problem, in Chapter 6, we were using Schrödinger's representation in which the state of a 1-electron system was described by a **wave func-tion**; the dynamical quantities were represented by differential operators, acting on the independent variables in the wave function (usually the Cartesian coordinates of the electron's position); and the eigenstates of definite energy were found by solving Schrödinger's

equation $H\Psi = E\Psi$. We've come a long way since then, moving towards a more general approach in which the abstract relationships among the operators became all-important. But in the end we usually have to come back to Schrödinger's formalism in order to calculate wave functions and all the observable properties they can give us.

The first step towards this end is to connect the more abstract geometrical ideas, developed in the last two Chapters, with Schrödinger's formalism. The most basic ideas relate to the **invariance** of an isolated physical system, and its properties, when it suffers a spatial displacement of some kind. A *linear* displacement, in which all points in the system are translated by the same amount in the same direction, has no effect – because all positions in space are equivalent; and a *rotational* dispacement has no effect – because all directions are equivalent and simply turning something round does not change its internal structure in any way.

First let's deal with translations, using D_x for the operation of displacing the system through a small amount dx in the x-direction. We know from (7.13) how the wave function must be transformed when the system moves to its new position: $\Psi \to \Psi' = U\Psi$ where U is a unitary operator and can be written $U = \exp i D$ in which D_x is Hermitian. (Remember, this follows because the normalization of Ψ mustn't change, so $(\Psi'|\Psi') = (U\Psi|U\Psi) =$ $(\Psi|U^{\dagger}U\Psi)$. This reduces to $(\Psi|\Psi)$ when

$$\mathsf{U}^{\dagger}\mathsf{U} = \exp{-i\mathsf{D}_{x}^{\dagger}}\exp{i\mathsf{D}_{x}} = \exp{i(\mathsf{D}_{x}-\mathsf{D}_{x}^{\dagger})} = 1.$$

– which is satisfied when D_x is Hermitian.)

On expanding the exponential, only up to the term in dx, we find

$$\Psi' \approx (1+i\mathsf{D}_x)\Psi(x,y,z) = \Psi(x,y,z) + i\mathsf{D}_x\Psi(x,y,z)$$
$$= \Psi(x,y,z) + i\Psi(x-\mathrm{d}x,y,z) = \Psi(x,y,z) - i\mathrm{d}x\frac{\partial\Psi}{\partial x},$$

where D_x acting on the *function* is equivalent to the inverse operation applied to the *variable*, as shown in (7.13).

In other words, going to the limit as $dx \to 0$,

$$\mathsf{D}_x \Psi = \frac{1}{i} \frac{\partial \Psi}{\partial x}$$

But this is just the recipe for getting the effect of the **momentum operator** p_x when it works on $\Psi(x, y, z)$ and we can therefore write

$$\mathsf{D}_x = \mathsf{p}_x = \frac{1}{i} \frac{\partial}{\partial x} \tag{8.30}$$

and the same argument applies when x is replaced by y or z.

The Hamiltonian for the translated system is also invariant, which means that

$$\mathsf{U}\mathsf{H}\mathsf{U}^{\dagger} = \mathsf{H}, \text{ or } \mathsf{U}\mathsf{H} = \mathsf{H}\mathsf{U}.$$

And the commutation of H with the operators describing translation of the system through space tells us that the momentum components (p_x, p_y, p_z) can also have definite values in an eigenstate of the energy. What about *rotations* in space? In the next Example we'll find that similar arguments apply.

Example 8.6 Connecting infinitesimal rotations with angular momentum

When the 'displacement' of the system is angular we must look at the effect of rotation around an axis. We'll take rotation through angle ϕ about the z-axis, calling it $\mathsf{R}_z(\phi)$ and ask how the components x, y, z of the position vector $\mathsf{r} = x \mathsf{e}_1 + y \mathsf{e}_2 + z \mathsf{e}_3$ change in the rotation. For real Cartesian basis vectors we know that (see for example (8.3)) $\mathsf{r} \to \mathsf{r}' = \mathbf{eR}$, where \mathbf{R} is the square matrix describing the rotation $\mathsf{R}_z(\phi)$. Thus

 $\mathbf{e}_1 \rightarrow \mathbf{e}_1' = \mathbf{e}_1 \cos \phi + \mathbf{e}_2 \sin \phi, \quad \mathbf{e}_2 \rightarrow \mathbf{e}_2' = -\mathbf{e}_1 \sin \phi + \mathbf{e}_2 \cos \phi, \quad \mathbf{e}_3' \rightarrow \mathbf{e}_3.$

Now the 'rotated function', according to (7.13), is obtained from the original by changing the argument r to $r' = R^{-1}r$ for the 'backwards rotated' field point, $r' = x'e_1 + y'e_2 + z'e_3$; but it's easy to show that, for real Cartesian basis vectors, the coordinates follow the same transformation law as the e-vectors. It then follows that

$$\mathsf{R}_z(\phi)\Psi(x,y,z) = \Psi(x',y',z'), \quad \text{with} \quad x' = x\cos\phi + y\sin\phi, \quad y' = -x\sin\phi + y\cos\phi, \quad z' \to z.$$

If you substitute these values into $\Psi(x', y', z')$ and go to the limit where the angle ϕ is infinitesimal (so that $\cos \phi \to 1$, $\sin \phi \to \phi$) you'll see that, to first order, $x' = x + y\phi = x + dx$ and $y' = y - x\phi = y + dy$. You can then go ahead as in the case of translation operations (just before this Example), making a Taylor expansion up to first order in $dx = y\phi$ and $dy = -x\phi$, and you will find in the limit

$$\mathsf{R}_{z}(\phi)\Psi(x,y,z) = \Psi(x,y,z) + \left(y\phi\frac{\partial\Psi}{\partial x} - x\phi\frac{\partial\Psi}{\partial y}\right).$$

Example 8.6 has shown that an infinitesimal rotation of the physical system, in real space, induces a corresponding rotation in **wave-function space**, which can be written as (don't forget that -i = 1/i!) $\mathsf{R}_z(\phi) = 1 - i\phi\mathsf{L}_z$ where L_z is used to denote the operator

$$\mathsf{L}_z = \frac{1}{i} \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right)$$

The symbol L_z has been used for the infinitesimal operator, previously denoted by J_z , when it refers to a function space with index 2j as an even integer. The operator L_z is evidently the one associated with the z-component of orbital angular momentum $R_z(\phi)$ in Schrödinger's representation. If equations (8.30) and (8.4) look a bit odd it's because you've forgotten that we've been using *dimensionless units* ('atomic units'): L_z is an operator for angular momentum, with the same physical dimensions as Planck's constant of action \hbar , so if you want to put the units back you should replace 1/i in (8.4) by \hbar/i . (Check that this is true in (8.30) also.)

All the operators introduced in this Chapter have their origins in geometry alone! The amazing thing about them is that they have all the properties of the operators we've been using in Schrödinger's form of quantum mechanics. What we've discovered is this:

Whenever a physical displacement of a system leaves its Hamiltonian unchanged, it means there is a corresponding operator that commutes with H. When the displacement is simply translation through space, the operator describes a **linear momentum**; when it is a rotation in space, the operator describes an **angular momentum**. In both cases their eigenvalues may be 'constants of the motion'.

(8.31)

So what are the 'new' operators (like L_x, L_y, L_z – or even p_x, p_y, p_z) that came out from our study of geometrical operations in ordinary 3-space? They really are exactly the same differential operators we've been using in quantum mechanics!

Chapters 6 and 7 exposed many of the general features of central-field systems, in particular those that arise from the commutation properties of the Hamiltonian operator H and the angular momentum operators L^2 and L_z .

The present Chapter 8 has concentrated on the **symmetry properties** of systems that are invariant under all 3-space rotations around one fixed point (e.g. the nucleus of a central-field system). The rotations form a **group** and the infinitesimal operators J_x, J_y, J_z have been shown to have the same commutation properties as angular momentum operators in general. In particular, $J^2 = J_x^2 + J_y^2 + J_z^2$ and J_z – associated with rotation operations in ordinary 3-space – behave just like L^2 and L_z , which are differential operators working on Schrödinger's wave functions. And finally we found the connection: with the abstract operators (for which we used symbols alone, with certain rules for combining them) we often find many **'realizations'** in which the symbols stand for definite mathematical operations (such as differentiating functions of variables, or moving vectors in space).

You've now seen how Schrödinger's differential operators provide just one way of describing the abstract angular momentum operators; algebraic operations on the monomials given in (8.12) provide another. And the algebraic approach is more powerful because it allows for the presence of **spin**. When *j* is a whole number the J-operators behave exactly like the orbital angular momentum operators, being expressible in terms of differential operators and working on *functions* (of x, y, z, or r, θ, ϕ). In that case we simply re-name the operators J^2 , J_z as L^2 , L_z and their eigenvectors as eigen*functions*. In that way we found the spherical harmonics $Y_m^{(l)}(\theta, \phi)$

But when j is half an odd integer, like $\frac{1}{2}$, there *are* no differential operators and instead we must fall back on the algebraic approach. In fact, that's the way all the properties of angular momentum have been derived in earlier Sections; and when you deal with *many*particle systems like atoms and molecules (Book 12) you'll have to think about **angular momentum coupling**, where two kinds of angular momentum are combined into one *total* angular momentum. You had a first taste of all this in Section 2.2, where the spins of two particles were coupled to give a resultant total spin, but now you've built up most of the mathematical 'machinery' for dealing with real many-electron systems. It remains only to put the 'finishing touches' to what you've done so far.

Chapter 9

Finishing touches –the Dirac formalism

9.1 The bra-ket notation

Apart from his formulation of the Uncertainty Principle, little has been said about Werner Heisenberg though he was one of the founders of quantum mechanics and played a key role in its development. He believed that the theory should be based purely on observable quantities, such as energy states, thus excluding Bohr's electron orbits and Schrödinger's wave functions. By setting out the observed energies, $E_1, E_2, E_3, ...,$ of the H atom electronic states, along the diagonal of a square table of values, and the transition energies $E_{ij} = E_i - E_j$ at the intersection of Row *i* and Column *j*, he could collect all the observable values in one place. Such a table is a **matrix** (as you now know, though the physicists of the day knew little about them) and Heisenberg began to argue with them in his **matrix mechanics**. Max Born (1882-1970), with a young collaborator Pascual Jordan, saw what Heisenberg was trying to do and put his attempts on a mathematically sound basis. Their resultant paper "On quantum mechanics", along with its sequel by Born, Heisenberg and Jordan (published in 1925), showed that different formulations of the theory were mathematically equivalent; that what really mattered were the abstract **commutation relations**.

At the same time there appeared a publication by a young Englishman, which led to exactly the same conclusions. Paul Adrian Dirac (1902 - 1982) made many great contributions to quantum physics: four years after his 1925 paper he published a first version of his famous book, "The Principles of Quantum Mechanics", in which he invented and used a highly original formalism to establish all the basic principles of the subject. In later editions he completed his development of quantum mechanics. Dirac's notation is now in almost universal use. Here we'll try to introduce it in the simplest possible way by summarizing and re-writing what you should know already.

In Chapter 1 we first used a state vector $\psi = c_1 \alpha + c_2 \beta$ to indicate a 'mixture' of two possible states of definite spin $S_z = \pm \frac{1}{2}$ ('up' or 'down'), the numerical coefficients indicating their relative probabilities $(|c_1|^2, |c_2|^2)$. In Chapter 2 this picture was generalized

to a system with an observable, A say, which could take many possible values, by writing the state vector (the name we'll normally use) as

$$\Psi = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + \dots + c_n \mathbf{e}_n + \dots$$
(9.1)

– a linear combination of possible states in which the **basis vector** \mathbf{e}_i symbolized the state in which A would have the definite value A_i . Dirac would have written this state vector as

$$|\Psi\rangle = c_1|1\rangle + c_2|2\rangle + \dots + c_n|n\rangle + \dots$$

where the pointed bracket '| \rangle ' reminds one that we're talking about a state *vector*, the space inside holding the name of the state or at least the index that labels it.

Here, when we're talking about Schrödinger's wave functions, expressed as linear combinations of basis functions, we'll write instead

$$|\Psi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle + \dots + c_n|\psi_n\rangle, +\dots$$
(9.2)

where the basis of this **function space** is usually infinite, though again only the first n members are shown.

As in Chapter 2, we also define a **Hermitian scalar product**. With Dirac's notation, $\langle \Psi | \Psi' \rangle = \langle \Psi' | \Psi \rangle^*$ for any two elements Ψ, Ψ' of the space. This is generally a complex number, but becomes real when $\Psi' = \Psi$ and gives the square of the 'length' of the vector. Any vector may be multiplied by a number c, in general complex, and it should be noted that

$$\langle \Psi | c \Psi' \rangle = c \langle \Psi | \Psi' \rangle, \quad \langle c \Psi | \Psi' \rangle = c^* \langle \Psi | \Psi' \rangle,$$

$$(9.3)$$

where the first form applies when the c appears in the right-hand vector of the scalar product, but the second when it appears on the left.

The whole set of scalar products, for all pairs of basis vectors, determines the **metric** of the space, usually collected in a square matrix **M** with elements (Dirac notatation) $M_{ij} = \langle i | j \rangle$. Normally, the basis is chosen to be **orthonormal**: in terms of the convenient 'delta symbol' this means

$$\langle i|j\rangle = \delta_{ij}, = 1 \ (i=j), = 0 \ (\text{otherwise}), \tag{9.4}$$

which makes \mathbf{M} into an infinite **unit matrix** and greatly simplifies the formalism. The length² of the vector (9.2) is then

$$|\Psi|^2 = \langle \Psi | \Psi \rangle = \sum_{i,j} c_i^* c_j \langle i | j \rangle = \sum_i c_i^* c_i,$$

where the coefficient that comes from the left-hand vector in the scalar product is the one that carries the 'star' according to (9.3). The length of any vector is thus sure to be a *real* number.

Dirac had the bright idea of looking at both vectors in a scalar product as independent elements, writing $\langle \Psi | \Psi \rangle = \langle \Psi | \times | \Psi \rangle$ where

$$|\Psi\rangle = c_1|1\rangle + c_2|2\rangle + \dots + c_j|j\rangle...$$

is called a "ket vector", while

$$\langle \Psi | = c_1^* \langle 1 | + c_2^* \langle 2 | + \dots + c_j^* \langle j | \dots$$

is a "bra vector". When the two are put together they make the "bra-ket" which is a scalar product (a single number). Note that the bra vector carries the 'starred' (complex conjugate) components and is often said to be the 'dual' of the corresponding ket vector. Note also that the components themselves may be written as scalar products: taking the scalar product of any vector $|\Psi\rangle$, with $\langle j|$ from the left, gives

$$\langle \psi_j | \Psi \rangle = c_j \langle j | j \rangle = c_j, \tag{9.5}$$

- since only the *j*th term remains when the basis is orthonormal.

The notation is easily extended to operators and their matrix elements. Thus, an operator A can be represented by a matrix A, with elements

$$A_{ij} = \langle i | \mathsf{A} | j \rangle,$$

where Dirac notation is used for the bra and ket vectors and the second vertical bar is put in just to separate the three parts (vector|operator|vector).

When operators are applied in succession, C = AB, the product C is represented by the matrix C with elements $C_{ij} = \sum_k A_{ik}B_{kj}$, which becomes in bra-ket form

$$\langle i|\mathsf{C}|j\rangle = \sum_{k} \langle i|\mathsf{A}|k\rangle \langle k|\mathsf{B}|j\rangle.$$

Clearly, even the 'ket-bra' product $|k\rangle\langle k|$, can be given a meaning:

$$|k\rangle\langle k|\Psi\rangle = |k\rangle \times (\langle k|\Psi\rangle)$$

and is therefore a vector in the 'direction' of the kth unit vector $|k\rangle$ with numerical coefficient $c_k = \langle k | \Psi \rangle$ i.e. it is the k-component of $|\Psi\rangle$. This result may be written

$$\mathsf{P}_k|\Psi\rangle = c_k|k\rangle,\tag{9.6}$$

where P_k is a **projection operator**, as first used in (2.2) of Chapter 2. If we project *all* components out of any given vector $|\Psi\rangle$ and add the results we get back the whole vector. In other words

$$\sum_{k} \mathsf{P}_{k} |\Psi\rangle = \left(\sum_{k} \mathsf{P}_{k}\right) |\Psi\rangle = |\Psi\rangle,$$
$$\sum_{(\text{all } k)} \mathsf{P}_{k} = \mathsf{I}, \tag{9.7}$$

which means

Here I is the **unit**, **or identity**, **operator** which doesn't change anything and may often be replaced by 1.

The property (9.7) is very useful: for example, the matrix representing an operator product ABC has an ij-element $\langle i|\text{ABC}|j\rangle$ but an identity operator $\sum_{k}|k\rangle\langle k|$ (with any index k, as long as it runs over all basis vectors) may be inserted between any two adjacent factors to give

$$\langle i | \mathsf{ABC} | j \rangle = \sum_{k,l} \langle i | \mathsf{A} | k \rangle \langle k | \mathsf{B} | l \rangle \langle l | \mathsf{C} | j \rangle.$$

In this way the 'chain rule' for getting matrix products is automatically respected. The eigenvalue problem for a harmonic oscillator (see Section 4.4) provides a simple example of the use of Dirac notation. This approach leads directly to the 'matrix mechanics' of Heisenberg, Born and Jordan.

9.2 Heisenberg's 'matrix mechanics'

As in Example 4.6, which you may want to read again, the eigenvalue equation $H\Psi = E\Psi$ for the linear oscillator contains the Hamiltonian $H = (1/2m)(p^2 + \alpha^2 x^2)$, where the constant $\alpha = m\omega$, ω being the angular frequency predicted by classical physics.

Here we'll start again from the beginning, supposing only that the position and momentum operators satisfy the commutator equation

$$xp - px = (i\hbar)I.$$

Let's write 2mH in the form

$$2m\mathsf{H} = \mathsf{p}^2 + m^2 \omega^2 \mathsf{x}^2 \tag{9.8}$$

and 'factorize' it by defining a new operator η and its adjoint η^{\dagger} :

$$\eta = (\mathbf{p} + im\omega\mathbf{x})/\sqrt{2m\hbar\omega}$$

$$\eta^{\dagger} = (\mathbf{p} - im\omega\mathbf{x})/\sqrt{2m\hbar\omega}.$$
(9.9)

The product $\eta \eta^{\dagger}$ doesn't give exactly the Hamiltonian 2mH but it follows easily that

$$\hbar \omega \eta \eta^{\dagger} = \mathsf{H} - \frac{1}{2} \hbar \omega \mathsf{I}.$$
 (A)
$$\hbar \omega \eta^{\dagger} \eta = \mathsf{H} + \frac{1}{2} \hbar \omega \mathsf{I}.$$
 (B)

(look back at Example 4.6 if you need to) and that there is a new commutation rule – which will be needed later:

$$\eta^{\dagger}\eta - \eta\eta^{\dagger} = 1. \qquad (C)$$

From (A) and (B) it follows that η and η^{\dagger} have properties similar to those of the 'stepdown' and 'step-up' operators L^- and L^+ used in Example 6.1 (read it again if you need to). In fact, supposing $|E_i\rangle$ to be an eigenket of H with energy E_i and writing it on the right of all terms in (A) you easily find (do it!)

$$\mathsf{H}(\eta|E_i\rangle) = (E_i - \hbar\omega)(\eta|E_i\rangle). \tag{9.10}$$

This shows that $\eta | E_i \rangle$ is an eigenket of H, but with energy reduced by an amount $\hbar \omega$, to E_j with j = i - 1. Starting instead from (B), and making similar steps, the corresponding result is

$$\mathsf{H}(\eta^{\dagger}|E_i\rangle) = (E_i + \hbar\omega)(\eta^{\dagger}|E_i\rangle), \tag{9.11}$$

which shows that $\eta^{\dagger}|E_i\rangle$) is an eigenket of H, but with energy stepped-up by $\hbar\omega$, to E_j with j = i + 1. Now it's possible to find the matrices to associate with all the operators.

Example 9.1 The matrix representation

To construct the representation we need an *orthonormal* set of basis vectors: the eigenkets $|E_0\rangle, |E_1\rangle, \dots, |E_n\rangle, \dots$ are orthogonal for different values of the integer n, as they belong to different energy eigenvalues (see Section 4.7) but they have not so far been normalized. To normalize the vector $|E_n\rangle$ we need to know the square of its length, $\langle E_n | E_n \rangle$, so as to get a normalizing factor $c = \langle E_n | E_n \rangle^{-1/2}$: for then $c\psi_n$ will be normalized, even if $|E_n\rangle$ is not.

It will be enough to look at two neighbouring kets, with eigenvalues E_{n-1} and E_n . If we apply the step-down operator to $|E_n\rangle$, the result will be $\eta|E_n\rangle = c_n|E_{n-1}\rangle$. But with Dirac's conventions the corresponding bra-vector will be $\langle E_n | \eta^{\dagger}$ (write the two factors in reverse order, taking the adjoint of each) and their scalar product will be $\langle E_n | \eta^{\dagger} \eta | E_n \rangle = c_n^* c_n \langle E_{n-1} | E_{n-1} \rangle = |c_n|^2$ – where the E_{n-1} eigenket has been taken as normalized.

Now from the property $\eta^{\dagger}\eta = H/(\hbar\omega) - \frac{1}{2}$ (see (A) just before this Example), along with the result we already found, that $H|E_n\rangle = (E_0 + n\hbar\omega)|E_n\rangle$, it follows that

$$\langle E_n | \eta^{\dagger} \eta | E_n \rangle = \langle E_n | (\mathsf{H}/\hbar\omega) - \frac{1}{2} | E_n \rangle = (n + \frac{1}{2}) - \frac{1}{2} = n.$$

Putting it all together we see that $c_n\eta$ is the correctly normalized step-down operator, which leads from $|E_n\rangle$ to $|E_{n-1}\rangle$ without spoiling the normalization. Consequently $\eta|E_n\rangle = \sqrt{n}|E_n\rangle$ and the nn-element of the matrix **H** will be E_n , all off-diagonal elements being zero . In obtaining this result we discovered that $\eta^{\dagger}\eta$ had another important property: its expectation value in any eigenstate of H is $\langle E_n | \eta^{\dagger}\eta | E_n \rangle = n$ and is thus a 'counting operator'. When applied to any eigenket $|E_n\rangle$ it gives the number of energy quanta $(\hbar\omega)$ in that eigenstate. Like H, N is represented in this basis by a matrix with non-zero elements only on the diagonal, all others being zero. It is often called the **number operator**: $N = \eta^{\dagger}\eta$.

On the other hand, the 'step-down' equation $\eta |E_i\rangle = \sqrt{n_i}|E_{i-1}\rangle$ shows that the ji-element of η , namely $\langle E_j |\eta|E_i\rangle$, is 0 unless j=i-1, and is then $\sqrt{n_i}$. And in the same way $(\eta^{\dagger}|E_i\rangle = (E_i + \hbar\omega)(\eta^{\dagger}|E_i\rangle)$, shows that the ji-element of η^{\dagger} , namely $\langle E_j |\eta^{\dagger}|E_i\rangle$, is 0 unless j=i+1, and is then $\sqrt{n_{i+1}}$.

These results are enough to give the matrices that represent all the operators.

Let's now write out the matrices associated with all the operators, using the results above. The *n*th diagonal elements of the matrices **H** and **N**, associated with H, N, are clearly $H_{nn} = E_n$ and $N_{nn} = n$, where n = 0 labels the first row and first column. With an obvious abbreviation, we write

$$\mathbf{H} = \operatorname{diag}\left(E_0 \, E_1 \, E_2 \dots E_n \dots\right) \quad \mathbf{N} = \operatorname{diag}\left(0 \, 1 \, 2 \dots n \dots\right), \tag{9.12}$$

where only the diagonal elements are shown.

The non-zero elements of the matrix associated with the step-down operator, η , appear only in positions with row-column indices (n-1, n), where n labels a *diagonal* element; and similarly those for the step-up operator appear only at points (n + 1, n). From the results obtained in Example 9.1, it follows that

$$\eta \rightarrow \begin{pmatrix} \mathbf{0} & \sqrt{1} & 0 & 0 & 0 & \dots \\ 0 & \mathbf{0} & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & \mathbf{0} & \sqrt{3} & 0 & \dots \\ 0 & 0 & 0 & \mathbf{0} & \sqrt{4} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad \eta^{\dagger} \rightarrow \begin{pmatrix} \mathbf{0} & 0 & 0 & 0 & 0 & \dots \\ \sqrt{1} & \mathbf{0} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & \mathbf{0} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{3} & \mathbf{0} & 0 & \dots \\ 0 & 0 & \sqrt{4} & \mathbf{0} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$
(9.13)

From the definition of the operators η and η^{\dagger} at the beginning of the Section it follows that they determine also the matrices **x** and **p** associated with the position and momentum operators **x** and **p**. In fact, as (9.9) shows, taking the sum and difference of the matrices in (9.13) should give (apart from a numerical factor in each) the matrices representing **p** and **x**, respectively.

Let's therefore write the results below, with p on the left and x on the right:

$$\frac{\mathbf{p}}{iC_p} = \frac{\mathbf{x}}{C_x} = \begin{pmatrix} \mathbf{0} & -\sqrt{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ \sqrt{1} & \mathbf{0} & -\sqrt{2} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \sqrt{2} & \mathbf{0} & -\sqrt{3} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \sqrt{3} & \mathbf{0} & -\sqrt{4} & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{4} & \mathbf{0} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}, \quad \begin{pmatrix} \mathbf{0} & \sqrt{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\ \sqrt{1} & \mathbf{0} & \sqrt{2} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \sqrt{2} & \mathbf{0} & \sqrt{3} & \mathbf{0} & \dots \\ \mathbf{0} & \mathbf{0} & \sqrt{3} & \mathbf{0} & \sqrt{4} \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{4} & \mathbf{0} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

$$(9.14)$$

If you look back at (9.9) you'll easily be able to identify the constants C_p and C_x . But why did we slip in the $i = \sqrt{-1}$? It was done simply because, without it, the matrix **p** would have been *real* and *antisymmetric* – while all the matrices associated with observables have been supposed **Hermitian**. By adding a 'phase factor' *i* in the definition of η , and of course -i in its adjoint η^{\dagger} , this convention is respected; and it doesn't make a scrap of difference to the properties of the operators (e.g. $i\eta \times -i\eta^{\dagger} = \eta\eta^{\dagger}$)!

Perhaps it may seem that the Dirac notation wasn't really very useful; after all, you could have got the same results without all those kets and bras (just as Heisenberg, Born and Jordan did). But its advantages become more obvious when we start thinking of *continuous* variables, that can take *any* values – not just those that are discrete and 'quantized'. In that case we can no longer even write out the matrices!

9.3 Generalizations. Observables with continuous values

When we first passed from state *vectors* to state *functions* by introducing the **wave** function $\Psi(x)$ we took the first big step in going from quantities (X, say) that could

take only discrete 'quantized' values $(X_1, X_2, ..., X_i,)$ to those that could take any values – even in the whole range $-\infty$ to $+\infty$. We did it in Section 3.2, where X_k (we used x_k) stood for the kth value of the x-coordinate of a particle's position: to describe the 'state' of the particle, when we're not sure where it will be found. We divided the whole range into tiny intervals $(x_k, x_k + \delta)$ and used p_k for the fractional number of times n_k/N we find it there in a long series of N observations. The corresponding state vector is indicated by

$$\Psi = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + \dots c_k \mathbf{e}_k \dots = \sum_k c_k \mathbf{e}_k,$$

where the component c_k is chosen to make $p_k = |c_k|^2$ (allowing the components to be real or complex numbers) and the **basis vectors** are supposed to be **orthonormal**, fixing the metric of the space to make the squared length of Ψ a real number:

$$|\Psi|^2 = |c_1|^2 + |c_2|^2 + \dots |c_k|^2 \dots = \sum_k |c_k|^2 = \sum_k p_k.$$

To be sure the particle is *somewhere* along the x-axis the vector must be of unit length, $\sum_k p_k = 1$, corresponding to certainty. The vector is then **normalized**. In Chapter 3 (Example 3.1) we noted how this condition could be expressed when the variable xcould take continuous values. The probability of observing the particle in the kth interval then becomes $|\Psi(x_k)|^2 \delta$, where $|\Psi(x_k)|^2$ is a continuous function of x evaluated at point $x = x_k$ and is called a **probability density**; it is a probability *per unit range of* x and thus gives the probability of the particle being within the interval when multiplied by the δ . In calculus notation, δ may be replaced by the differential dx and the normalization condition $\sum_k p_k = 1$, for a variable with discrete values, will be replaced by $\int p(x)dx = 1$, where the definite integral (see Book 3) is taken over the whole range of x values – even when not shown. In terms of the wave function, the correspondence is thus

$$|\Psi|^2 = \sum_k |c_k|^2 \quad \to \quad |\Psi|^2 = \int |\Psi(x)|^2 \mathrm{d}x.$$
 (9.15)

In other words, to pass from discrete values of a vector component (c_k) to continuous values, we treat the function value $\Psi(x)$ as a 'continuously variable component' of the state vector, with x playing the part of the label k, and summation over k being replaced by integration over x.

This idea can be used more generally, as the next few Examples show, when we express the eigenvalue equation in Dirac notation.

Example 9.2 Schrödinger's equation in Dirac form. The problem

In abstract form, the energy eigenvalues for the stationary states of a system follow from the operator equation $H\Psi = E\Psi$, where H is the Hamiltonian operator which in some way changes the state vector Ψ into a new vector Ψ' . (You can think of this as some kind of 'rotation' $\Psi \to \Psi'$ -pointing in a different 'direction' in the vector space.) When the equation is satisfied, the new vector doesn't change its direction but only its length – being multiplied by a real factor E, which is one of the eigenvalues. In Schrödinger's

formulation of quantum mechanics, $\Psi = \Psi(x)$ for a particle moving along the x-axis; and H is represented in terms of position and momentum operators x, p (see the Examples in Chapter 4).

In Dirac notation the equation to be solved would be written as $\mathsf{H}|\Psi\rangle = E|\Psi\rangle$, but this doesn't help much until the symbols are given some meaning. Schrödinger did that by interpreting Ψ as a function of position (x) and expressing H in terms of the operators listed in (3.16). If we write $\Psi(x) = \langle x|\Psi\rangle$ (Dirac notation!), thinking of it as a sort of 'x-component' of the state vector, the normalization condition (9.15) contains

$$\langle \Psi | \Psi \rangle = \int \langle \Psi | x \rangle \langle x | \Psi \rangle \mathrm{d}x.$$

(Think about this for a minute or two! If $\langle x|\Psi\rangle$ means the 'x-component' of Ψ , labelled by the continuous variable x, then $\langle \Psi|x\rangle$ means its complex conjugate ($\Psi(x)^*$) and the whole expression, with the integration over x, gives the squared length of the state vector Ψ .)

But how can we deal with the eigenvalue equation $\mathsf{H}|\Psi\rangle = E|\Psi\rangle$? If the vector Ψ had a k-component, along the basis vector \mathbf{e}_k of a finite dimensional space like the ones we're used to, then the components would be collected in a column of numbers \mathbf{c} and any rotation of the vector, $\Psi \to \Psi' = \mathsf{H}\Psi$, would be described by a square matrix \mathbf{H} . The eigenvalue equation would then become a **matrix equation**: $\mathbf{H}\mathbf{c} = E\mathbf{c}$ or in terms of components

$$\sum_{k} H_{jk} c_k = E c_j \quad (\text{all } j).$$

In that case there would be no difficulty in using Dirac notation: every element of the matrix **H** involves two components (indices j, say, for the row, and k for the column). The jk-element could then be written $H_{jk} = \langle j | \mathsf{H} | k \rangle$ and the matrix representation would become

$$\sum_k \langle j | \mathsf{H} | k \rangle \langle k | \Psi \rangle = E \langle j | \Psi \rangle \quad (\text{all } j),$$

where there is one equation for every j-component (giving a big set of simultaneous equations).

The trouble is that we don't know how to deal with the case in which the components are labelled by continuous variables – on top of which the number of equations is infinite and even uncountable!

For a continuous representation, in which the 'components' of Ψ are written in the form $\langle k | \Psi \rangle$ with the row index j replaced by a variable such as x, we might guess that the eigenvalue equation would be written

$$\int \langle x | \mathsf{H} | x' \rangle \mathrm{d}x' \langle x' | \Psi \rangle = E \langle x | \Psi \rangle, \qquad (9.16)$$

in which the two values of the variable, x and x', replace the row- and column-indices, j and k. The two-variable matrix element $\langle x|\mathsf{H}|x'\rangle$ is called an "integral kernel" and (9.16) an **integral equation**. How can we make sense of this equation – which looks so different from the Schrödinger equation we've been using so far?

The form used so far has been simply $H\Psi(x) = E\Psi(x)$, where H stands for a *differential* operator:

$$\mathsf{H} \to -(\hbar^2/2m)\mathrm{d}^2/\mathrm{d}x^2 + V(x).$$

There is no difficulty in expressing $\Psi(x)$ in Dirac notation as $\langle x|\Psi\rangle$, but how should we interpret the left-hand side of (9.16)? The potential energy function V(x) simply multiplies $\langle x|\Psi\rangle$, whereas the corresponding term in (9.16) would be written as

$$\int \langle x | \mathsf{V} | x' \rangle \mathrm{d}x' \langle x' | \Psi \rangle,$$

which contains an *integration*. It seems, in fact, that no integration is needed: for example the operator V leaves the function $\langle x|\Psi\rangle$ unchanged, multiplying it only by the PE V(x). In other words, the term $\nabla\Psi(x)$ (Schrödinger form), or $\int \nabla |x'\rangle dx' \langle x'|\Psi\rangle$, as it appears in (9.16), is equivalent to $V(x)\langle x|\Psi\rangle$: the integration over x' is 'cancelled' and the new variable x' is set equal to x.

There's a neat way of expressing this: we *invent* a new function $\delta(x - x')$, called the "delta function", with just the property we need. It is defined by

$$\int \delta(x - x')f(x')dx' = f(x)$$
(9.17)

and it acts on a given function f(x) by changing the variable x to a new variable x', and then integrating over the whole range of interest (even for all x' between $\pm \infty$, to give you back the function you started from! It must have a very sharp peak when $x' \to x$, must fall rapidly to zero outside the peak, and must be normalized so that

$$\int \delta(x - x') f(x') dx' \approx f(x) \int \delta(x - x') dx' = f(x).$$

(Here f(x') varies only slowly compared with $\delta(x - x')$ in the region of the peak.) Professional mathematicians try not to use the delta function but it's a convenient way of avoiding difficult analysis. (And you've already met similar functions in Chapter 5 (e.g. in Example 5.2), so you know they exist!) A rigorous analysis gives at least one concrete example of a function with the properties we need: it is

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ikx} \mathrm{d}k, \qquad (9.18)$$

where k and x are any real variables lying in the interval $(-\infty, +\infty)$. But the 'equality' applies only when each side multiplies a well-behaved function f(x) and is followed by an integration over the whole interval.

Thus, we can say (on replacing x by x - x') that

$$\int \delta(x - x') dx' = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x - x')} dk dx', \qquad (9.19)$$

All this just confirms what we could have guessed already: that when an equation, in a discrete representation, contains a ket-bra factor such as such as $|k\rangle\langle k|$, summed over all k, this factor is a representation of the identity operator I and may be removed; and that a similar interpretation is valid even when discrete indices are replaced by continuous variables.

To summarize, the two integral kernels $\langle x|V|x'\rangle$ and $\langle x|T|x'\rangle$, which together make up that for the Hamiltonian $\langle x|H|x'\rangle$, take the delta-function forms

$$\langle x|\mathbf{V}|x'\rangle = V(x)\delta(x-x'), \quad \langle x|\mathbf{T}|x'\rangle = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\,\delta(x-x'). \tag{9.20}$$

Before ending the Section, we should check that on using these forms in (9.16) we get back the usual Schrödinger equation in the x-representation.

Example 9.3 Is the recipe correct?

It will be enough to look at one of the kernels in (9.18). Let's take the one that looks more difficult – the one that should give the effect of the kinetic energy operator. When T works on the wave function $\Psi(x)$, or in Dirac notation $\langle x|\Psi\rangle$, it should be equivalent to differentiating twice with respect to x and multiplying the result by $-\hbar^2/2m$.

Operating with T on the state vector $|\Psi\rangle$ will give a new ket, $\mathsf{T}|\Psi\rangle$, represented by the function ('*x*-component')

$$\langle x|\mathsf{T}|\Psi\rangle = \int \langle x|\mathsf{T}|x'\rangle \langle x'|\Psi\rangle \mathrm{d}x' = -(\hbar^2/2m)(\mathrm{d}^2/\mathrm{d}x^2) \int \mathrm{d}x'\delta(x-x')\langle x'|\Psi\rangle,$$

where the operators work, as usual, in the order right-to-left. The delta operator works first, its property (9.17) giving $\int dx' \delta(x-x') \langle x' | \Psi \rangle = \langle x | \Psi \rangle$. And you are left with

$$\langle x|\mathsf{T}|\Psi\rangle = -(\hbar^2/2m)(\mathrm{d}^2/\mathrm{d}x^2)\langle x|\Psi\rangle.$$

So writing the Schrödinger equation as an integral equation, with suitably defined integral kernels, confirms that it simply reproduces the differential-operator form.

Of course, that's no big deal! – getting back what you already know, with a different notation. But in the next (and last!) Section you'll find something quite new – the prepresentation, in which the roles of the momentum and position operators p and x are reversed (a possibility we noted in Section 4.1).

9.4 The p-representation: momentum space

Suppose we want to change from one 'language' to another e.g. from one in which the coordinate operators x, y, z are represented by the multipliers x, y, z, to another in which it is the *momentum* operators p_x, p_y, p_z that simply multiply by the corresponding variables p_x, p_y, p_z . This means using **transformation theory** and it can be done very conveniently with Dirac notation.

In this Section we'll take the simple case of a 1-dimensional system, a particle moving along the x-axis. And, as usual, we'll use just p for the momentum variable p_x , as there's only one. Remember also that most of the work we need to do has been done already in earlier Sections: you know about basis change (see for example Sections 7.1 and 7.2) and how all your equations change if you decide to set up and use a new basis. So here we'll start by just reviewing things briefly and changing the notation a bit.

The state vector Ψ is something 'physical', which doesn't depend on how we describe it. But now we're using Dirac notation we'll keep it as a label for the corresponding ket or bra, $|\Psi\rangle$ or $\langle\Psi|$. The basis vectors, which we've often called $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_i$, etc. will now be called $\psi_1, \psi_2, \dots, \psi_i$, etc., and will also serve to label the corresponding ket and bra vectors, e.g. $|\psi_i\rangle$, $\langle\psi_i|$. The components of $|\Psi\rangle$, relative to the basis vectors $|\psi_i\rangle$ will thus become $\langle \psi_i | \Psi \rangle$, replacing the c_i in the expansion $\Psi = c_1 \psi_1 + c_2 \psi_2 + ...$, as used in earlier chapters. The *matrix elements*, relative to this basis, of an operator such as H will now be written $\langle \psi_i | \mathbf{H} | \psi_j \rangle$ – a two-index quantity, corresponding to the row (i) and column (j) indices of an element of the square matrix **H**.

To pass from the basis $(\psi_1, \psi_2, \dots, \psi_i, \dots)$ to a new basis $\bar{\psi} = (\bar{\psi}_1, \bar{\psi}_2, \dots, \bar{\psi}_i, \dots)$, the new vectors must be related to the old: thus, for example, the new vector $\bar{\psi}_i$ will be a linear combination

$$\bar{\psi}_i = \sum_j \psi_j U_{ji},\tag{9.21}$$

where U_{ji} is the ji-element of a square matrix **U**, whose columns contain the expansion coefficients of the new vectors in terms of the old. As you know from Section 7.1, in order to conserve orthonormality **U** must be a **unitary matrix** (obtained by transposing rows and columns and taking the complex conjugate of every element) denoted by \mathbf{U}^{\dagger} . It has the property $\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{U}\mathbf{U}^{\dagger} = \mathbf{1}$ and is thefore its own **inverse**.

When the basis vectors are changed according to (9.21) the components of any given vector must be changed according to

$$c_i \to \bar{c}_i = \sum_j U_{ij}^{\dagger} c_j \tag{9.22}$$

and the matrix elements of any operator such as H must transform as follows

$$H_{ij} \to \bar{H}_{ij} = \sum_{k,l} U_{ik}^{\dagger} H_{kl} U_{lj}.$$
(9.23)

Let's now express these effects of basis change in Dirac notation.

Example 9.4 Effects of basis change

Note first that c_i is the component of $|\Psi\rangle$ along $|\psi_i\rangle$, so $c_i = \langle \psi_i |\Psi\rangle$; and that U_{ji} in (9.21) is a matrix element of the *operator* U that rotates the whole basis, so $U_{ji} = \langle \psi_j | \mathsf{U} | \psi_i \rangle$. Thus, the basis change (9.21) becomes

$$|\psi_i\rangle \rightarrow |\bar{\psi}_i\rangle = \sum_j |\psi_j\rangle \langle \psi_j |\mathsf{U}|\psi_i\rangle.$$

Similarly, the components of the ket vector $|\Psi\rangle$ will change to

$$\langle \psi_i | \Psi \rangle \to \langle \bar{\psi}_i | \Psi \rangle = \sum_j \langle \psi_i | \mathsf{U}^\dagger | \psi_j \rangle \langle \psi_j | \Psi \rangle = \sum_j \langle \bar{\psi}_i | \psi_j \rangle \langle \psi_j | \Psi \rangle,$$

where the last step follows because (with Dirac notation) $\langle \psi_i | \mathsf{U}^{\dagger}$ means the bra vector corresponding to the ket $\mathsf{U} | \psi_i \rangle$ (remember? – take the adjoints of the two factors and reverse their order).

Finally, the matrix elements of the operator H change according to

$$\begin{split} \langle \psi_i | \mathsf{H} | \psi_j \rangle &\to \langle \bar{\psi}_i | \mathsf{H} | \bar{\psi}_j \rangle &= \sum_{k,l} \langle \psi_i | \mathsf{U}^{\dagger} | \psi_k \rangle \langle \psi_k | \mathsf{H} | \psi_l \rangle \langle \psi_l | \mathsf{U} | \psi_j \rangle \\ &= \sum_{k,l} \langle \bar{\psi}_i | \psi_k \rangle \langle \psi_k | \mathsf{H} | \psi_l \rangle \langle \psi_l | \bar{\psi}_j \rangle. \end{split}$$
Notice that, in this equation and the one before it, the operator U – which changes the basis – has been eliminated: once you've set up the basis you want to use you only need to get the scalar products like $\langle \bar{\psi}_i | \psi_k \rangle$ between the new and old basis vectors.

To summarize the conclusions from Example 9.4, a change of basis is essentially a change of 'language' in which the eigenvalue equation $H|\Psi\rangle = E|\Psi\rangle$ stays the same and gives the same solutions, but the operator and its eigenkets take very different forms.

Thus, in the basis change (9.21), the state vector $|\Psi\rangle$, with components $c_i = \langle \psi_i | \Psi \rangle$, will take a similar form but with the new components

$$\langle \bar{\psi}_i | \Psi \rangle = \sum_j \langle \bar{\psi}_i | \psi_j \rangle \langle \psi_j | \Psi \rangle, \qquad (9.24)$$

connected with the original basis through the scalar products $\langle \psi_i | \psi_j \rangle$ between the new and old kets.

The matrix elements of the operator H must be transformed in a similar way: in terms of the new basis they become

$$\langle \bar{\psi}_i | \mathsf{H} | \bar{\psi}_j \rangle = \sum_{k,l} \langle \bar{\psi}_i | \psi_k \rangle \langle \psi_k | \mathsf{H} | \psi_l \rangle \langle \psi_l | \bar{\psi}_j \rangle.$$
(9.25)

The beauty of the bra-ket formalism is that all these results could have been derived without any work, just by inserting the identity operator on either side of the operator in the matrix element you need. Thus, making spaces on either side of H, you need $\langle \bar{\psi}_i |$ H $|\bar{\psi}_j \rangle$ and putting $\sum_k |\psi_k\rangle \langle \psi_k|$ in the first space and $\sum_l |\psi_l\rangle \langle \psi_l|$ in the second gives you exactly the result in (9.25). The notation does the work for you!

Now let's get back to momentum space, where the basic kets are labelled by a continuous variable p and the eigenket $|\Psi\rangle$ is represented by the function $\langle p|\Psi\rangle$. In Schrödinger language $|\Psi\rangle$ was represented as a function of *position* x, which could be expressed in terms of basis functions $\psi_i(x)$, or in Dirac notation $\langle x|\psi_i\rangle$. But now we want to express everything in terms of p, so we must look for a new basis, with functions $\langle p|\bar{\psi}_i\rangle$. In fact we have one already – we found sets of Schrödinger eigenfunctions for a free particle in Section 4.6. For a particle moving along the x-axis we wrote them in the form $\psi_k(x) = N \exp ikx$, where N is a normalizing factor and k is a **wavenumber**, related to the momentum by $p = k\hbar$.

When the basis functions are labelled by continuous variables, x and p, the transformation equations take the corresponding forms:

$$\langle x|\Psi\rangle \to \langle p|\Psi\rangle = \int_x \langle p|x\rangle \mathrm{d}x \langle x|\Psi\rangle,$$
(9.26)

$$\langle x|\mathbf{H}|x'\rangle \to \langle p|\mathbf{H}|p'\rangle = \int_{x,x'} \langle p|x\rangle \mathrm{d}x \langle x|\mathbf{H}|x'\rangle \langle x'|p'\rangle \langle, \qquad (9.27)$$

where the bars are no longer needed, the two representations being indicated by the variables x (old) and p (new). Note that there is only one integration (over x) in (9.26), but there are two (over x and x') in (9.27)

Those are the basic laws of transformation theory. Thus $\langle x|\Psi\rangle$ is the Schrödinger wave function for state $|\Psi\rangle$ while $\langle p|\Psi\rangle$ is its corresponding wave function in momentum space. The link between the two representations is provided by the eigenfunctions for a free particle, namely $\langle p|x\rangle$, which is the Schrödinger function for a state of definite momentum p. (And don't forget that $\langle x|p\rangle = \langle p|x\rangle^*$. What does it mean?)

So now you have all that's needed. To show how well it all works let's go back briefly to Section 5.4, where (just before equation (5.20)) we guessed that there may be another representation, similar to Schrödinger's but with the roles of x and p reversed. Now we can see how to get it.

Example 9.5 An alternative to Schrödinger's representation

In Schrödinger language the basic position and momentum operators, x, p, for a 1-dimensional system, are represented by the two integral kernels (cf.(9.20)

$$\langle x|\mathbf{x}|\mathbf{x}'\rangle = x\delta(x-x'), \quad \langle x|\mathbf{p}|x'\rangle = -i\hbar\frac{\mathrm{d}}{\mathrm{d}x}\,\delta(x-x')$$

To express them in the language of momentum space we need the 'connecting' functions $\langle p|x\rangle$ and $\langle x'|p'\rangle\langle$, as used in the transformation equation (9.27). To transform the momentum operator, for example, we need its matrix elements

$$\langle p | \mathbf{p} | p' \rangle = \int_{x,x'} \langle p | x \rangle \mathrm{d}x \langle x | \mathbf{p} | x' \rangle \langle x' | p' \rangle \langle$$

and that means knowing the scalar products such as $\langle x|p\rangle$, which connect the two representations. In fact, we know them already because $|p\rangle$ is a momentum eigenket for a free particle moving along the x-axis. And $\langle x|p\rangle$ is its Schrödinger wave function, which we obtained in Example 4.1: it has the form (not yet normalized) $\langle x|p\rangle = \exp(ikx)$, with wavenumber $k = p/\hbar$.

In Chapter 4 the wave function was normalized over a certain region of space (e.g. a finite box, over which $\int |\Psi|^2 dx = 1$) but when the region becomes infinite another convention must be used. Here we make use instead of $\int_{p,x'} \langle x|p \rangle dp \langle p|x' \rangle dx'$. This contains the identity operator in the middle, which may be removed after doing the p-integration, giving (remember $p = k\hbar$)

$$\int_{p} \langle x|p \rangle \mathrm{d}p \langle p|x' \rangle = \hbar \int_{k} \exp ik(x-x') \,\mathrm{d}k = 2\pi\hbar\delta(x-x').$$

On putting x' = x and doing the remaining integration over x this gives $\int_x |\langle x|p\rangle|^2 dx = 2\pi\hbar$. The required normalizing factor for the momentum-space wave function $\langle x|p\rangle$ is thus $(2\pi\hbar)^{-1/2}$. This is an example of **delta-function normalization**.

The momentum-space matrix element of \boldsymbol{p} now follows on using the correctly normalized free-particle functions: it is

$$\begin{split} \int_{x,x'} \langle p|x \rangle \mathrm{d}x \langle x|\mathbf{p}|x' \rangle \langle x'|p' \rangle &= \int_{x,x'} \frac{\exp -ikx}{\sqrt{2\pi\hbar}} \left[\frac{\hbar}{i} \frac{\mathrm{d}}{\mathrm{d}x} \delta(x-x') \right] \frac{\exp ik'x'}{\sqrt{2\pi\hbar}} \mathrm{d}x' \mathrm{d}x \\ &= \frac{1}{2\pi\hbar} \frac{\hbar}{i} \int_{x} (ik') \exp[-i(k-k')x] \,\mathrm{d}x, \end{split}$$

where it was noted that the δ -function in the x' integration had the effect of putting x' = x; and that the $(d/dx) \exp(-ik'x)$ then gives a further factor ik'. Nearly finished! From (9.19) $\int \exp[-i(k-k')x] dx = 2\pi\delta(k-k')$ and a simple change of variable from k to Ak (see Book 3 Section 4.3) tells you that $\delta(Ax) = (1/A)\delta(x)$. It follows from above, on introducing $p = \hbar k$, that

$$\int \exp[-i(p-p')x] dx = (1/\hbar)\delta(p-p')$$

and this gives the final result (check it!)

$$\langle p|\mathbf{p}|p'\rangle = \frac{1}{2\pi\hbar}\frac{\hbar}{i}(ik)2\pi\,\hbar\delta(p-p') = \hbar k = p.$$

In words, the transformed momentum operator is 'diagonal' (the kernel having only elements $\langle p|\mathbf{p}|p'\rangle$ with p' = p) and follows on multiplying the resultant "momentum space wave function" by the momentum variable p – just as we guessed in Chapter 5.

Now that you've seen how to represent the operator \mathbf{p} in momentum space as an **integral operator**, by getting the integral kernel $\langle p|\mathbf{p}|p'\rangle$, you can follow the same pattern to get the kernel $\langle p|\mathbf{x}|p'\rangle$. On putting the two results together you should find

~ (

· · · ·

$$\langle p | \mathbf{p} | p' \rangle = p \delta(p - p') \langle p | \mathbf{x} | p' \rangle = -\frac{\hbar}{i} \frac{\mathrm{d}}{\mathrm{d}p} \delta(p - p').$$
 (9.28)

The second result should be compared with

$$\langle x | \mathbf{p} | x' \rangle = \frac{\hbar}{i} \frac{\mathrm{d}}{\mathrm{d}x} \delta(x - x').$$

As you can see, the p-representative of x is obtained from the x-representative of p by swapping the roles of the variables and changing the sign of i – in line with the results we guessed in Chapter 5.

To end this Section, we pass from a 1-dimensional system to one in 3-dimensional space. Instead of one pair of conjugate variables x and $p(=p_x)$ we now have three pairs, but the operators for different pairs commute and this makes generalization easy. If we work in the usual coordinate (Schrödinger) representation the basic kets will be $|x, y, z\rangle$, representing states in which all three coordinates are simultaneously definite, while in momentum space the basic eigenkets will be $|p_x, p_y, pz\rangle$ – corresponding to states of definite linear momentum. As in Section 4.6, the eigenstates will in each case be of product form, with one factor for each of the x-, y- and z-directions; remember the momentum eigenfunctions (Section 4.6), which have the (unnormalized) form

$$\Psi_{p_x, p_y, p_z}(x, y, z) = \exp i\hbar (k_x x + k_y y + k_z z)$$

and may be written in Dirac notation as a product $\langle x|p_x\rangle\langle y|p_y\rangle\langle z|p_z\rangle$.

Again, it is convenient to introduce a **wave vector**, writing k for the vector with components k_x, k_y, k_z and the 3-space momentum eigenfunctions are then more neatly written as

$$\langle \mathbf{r} | \mathbf{p} \rangle = \frac{\exp i \mathbf{p} \cdot \mathbf{r}}{2\pi\hbar)^{3/2}},$$
(9.29)

where normalizing factors are now included in the denominator. Remember, \mathbf{r} is the usual 3-space position vector, with components x, y, z, while \mathbf{p} is the momentum vector, $\mathbf{p} = \hbar \mathbf{k}$. The results obtained in Example 9.5 are easily extended to systems in 3-space on using the 'connecting functions' given in (9.29) and remembering that $\langle \mathbf{p} | \mathbf{r} \rangle = \langle \mathbf{r} | \mathbf{p} \rangle^*$. Thus, the kinetic energy operator $\mathsf{T} = \mathbf{p}^2/2m$ keeps the 'diagonal' form and becomes

$$\langle p|\mathsf{T}|p'\rangle = p^2/2m\delta(\mathsf{p}-\mathsf{p}'),$$

as you would expect. But the PE function V(x, y, z) has an associated operator V, which is a simple multiplier in coordinate space, is usually difficult to transform into momentum space: the kernel $\langle p_x, p_y, p_z | \mathsf{V} | p'_x, p'_y, p'_z \rangle$ transforms into something simple only in very special cases, like $\mathsf{V} = \mathsf{x} -$ which gives the differential operator in (9.28). In other cases you have to work it out 'from scratch' using the general equation

$$\langle \mathbf{p} | \mathsf{V} | \mathbf{p}' \rangle = \int \mathbf{p} | \mathbf{r} \rangle \langle \mathbf{r} | \mathsf{V} | \mathbf{r}' \rangle \langle \mathbf{r}' | \mathbf{p}' \rangle \mathrm{d} \mathbf{r} \mathrm{d} \mathbf{r}'$$

where $\mathbf{r}, \mathbf{r}', \mathbf{p}$ are short for the triplets of labels (x, y, z), (x', y', z'), and (p_x, p_y, p_z) , respectively; while $\int d\mathbf{r} d\mathbf{r}'$ means you have to do six integrations. That can be done easily when most of them involve delta-functions, as in the case of $\langle p|\mathbf{x}|p'\rangle$; but if you try to get the corresponding momentum-space kernel for, say, a Coulomb potential function V(r), with $r = (x^2 + y^2 + z^2)^{-1/2}$, you'll soon find how difficult it can be. That's why nearly all current calculations of atomic and molecular electronic structures are made using, in one way or another, Schrödinger's coordinate-space representation.

Looking back –

To get so far into Physics you must have started years ago! And you did, because Book 1 ("Number and symbols: from counting to abstract algebras") gave you a 'key' for getting into all of the 'Physical Sciences'. Symbols are just the marks you make on paper: they can stand for *anything* – words, numbers, displacements through space (that we called "vectors"), the "operators" (that turn things round in space), and so on – until you decide what you want them to stand for and what rules you will use to combine them. Before that they are completely *abstract*.

In Book 1 you learnt how to use symbols in many ways; and you go on doing that in the whole of Science. Book 2 was about Space – the space we live in and some of the 'spaces' we had to *invent* in order to express our ideas in pictures. You took your first steps into 'real' Physics in Book 4; going from Galileo and his experiments on falling bodies; to Newton and his formulation of the 'equations of motion' (using the language you learnt in Book 3 – the differential calculus). These equations apply to all 'everyday' objects, from thrown stones to planets moving round the Sun, or satellites moving round the Earth.

In the present Book 11 you have often been reminded of the things you have studied before. That's why in "Looking back" we're now looking further back than usual – not just to the beginning of this book, but to the beginning of the Series. All areas of Science are linked and inter-dependent, especially the Physical Sciences, which depend heavily on measurement and mathematical argument. The break between 'classical' physics and 'quantum' physics really started in Book 5, where we first began to talk about the **elementary particles** from which all matter is built up. These particles are much too small to see and handle in the laboratory: you can only *imagine* that they exist and then do '*thought experiments*' to discover how matter *would* behave if the particles really did exist and followed the laws of quantum mechanics, rather than Newton's laws. The trouble was, at the beginning of the last century, that nobody knew what the laws were – they had to be guessed and then, much later, confirmed when they led to conclusions in agreement with what was observed

Quantum mechanics is full of thought experiments! Book 12 started by looking at just one of them, the simplest you can imagine, where you think of a beam of 'spinning' particles being 'tested' to see whether their spin 'vectors' point 'up' or 'down'. The beam could be a stream of electrons (which have spin as well as mass and electric charge), like the one in your television set. The details don't matter – what is important is that each particle carries a **'two-valued observable'** and that you can separate them into two groups by passing the beam between the (specially-shaped) poles of a powerful magnet. The whole book followed from that one experiment, the "Stern-Gerlach (SG) experiment".

Let's review the main steps you took on your unusual journey through the theory ...

• Starting from the SG experiment, Chapter 1 (Observation and measurement) introduced most of the essential ideas of quantum mechanics, in just about a dozen pages!. It showed how the results from the experiment could be described in a simple picture, with a **state vector** to indicate the spin state of the particle, before the 'observation', and two unit vectors to indicate the two possible results – up-spin or down-spin, labelled $+\frac{1}{2}$ and $-\frac{1}{2}$. These were called **eigenvalues**, the corresponding unit vectors being **eigenvectors**.

You learnt how to give such things *meaning* in terms of the **projection operators** you first used in Section 6.2 of Book 1; and went on to show how an operator could be associated with any observable, in such a way that its average (or 'expectation') value in a long series of experiments could be related to the orientation of the state vector.

Finally, you noted an **invariance principle** – that the whole experiment could be turned round in space without making any difference to the results – to find the properties of the **spin operators** S_x , S_y , S_z . So much came out from so little!

• Chapter 2 showed how all you'd learnt could be applied to measurements on an observable X, say, with any number of possible values, $X_1, X_2, ..., X_n$, say, just by changing the dimensions of the vector space that contains the (unknown) state vector. Instead of n = 2, for the 2-dimensional spin-space, you now have n arbitrary (as big as you please); and what you did in 2-space was repeated with little change in n-space. You knew quite a lot about many-dimensional spaces already so there was no problem in defining the operator X to associate with any observable X in terms of its eigenvalues and corresponding eigenvectors. And the pictorial representation of the state vector, and of what the operators were doing to it, was also easy to visualize.

Examples showed how the spin operators, with the properties deduced in Chapter 1, could be used to find spin operators for an imaginary *two*-spin system and to show that they satisfied the same **commutation relations** as for a single spin. You used them to find new spin eigenstates and important new operators, the **step-up** and **step-down** operators which could change an eigenstate into one with higher or lower quantum number.

Finally, you had to admit state vectors with *complex* components by making a small change to the **metric**.

• The next chapter got you into Wave mechanics, Schrödinger's form of quantum mechanics. When the observable we've been calling X is continuous, like the Cartesian coordinates x, y, z of a particle, the state vector Ψ becomes a wave function. For a particle moving along the x-axis we used $\Psi(x)$ and, instead of the k-th component c_k of the vector Ψ , we had to think of $\Psi(x_k)$ as a sort of 'continuous component' evaluated at the point $x = x_k$. And, instead of taking $|c_k|^2$ as the **probability** of finding a particular value x_k of the observable, we had to use $|\Psi(x_k)|^2$ – though this quantity is a probability density, a probability per unit range, of finding the particle 'at' x_k .

In three dimensions, the wave function is $\Psi(x, y, z)$ and $|\Psi(x, y, z)|^2 dx dy dz$ the probability of finding the particle in a tiny 'box' of volume dx dy dz: $|\Psi(x, y, z)|^2$ is now a probability per unit *volume*.

From that interpretation (due to Born) you were able to define not only the probable value (\bar{x} , say) of x in a state with wave function $\Psi(x)$ but also the degree of **uncertainty** in that value. And from there you learnt that for any two observables, A, B with operators A, B, there was a limit to the precision with which they could both be measured in any given state Ψ . The more accurately you know A, the less you can say about B: the product of the uncertainties ΔA and ΔB must be at least $\frac{1}{2}(\Psi, iC\Psi)$, where iC is the **commutator** AB – BA of the two operators. Only when their operators commute (C = 0), can A and B both be accurately known in a given state Ψ .

With a state function $\Psi(x, y, z)$ you found that by choosing x and \mathbf{p}_x as, respectively, $\mathbf{x} \to x$, (multiply by x); and $\mathbf{p}_x \to (\hbar/i)(\partial/\partial x)$, the commutator for this **conjugate pair** came out as a constant \hbar . And similarly for the other two conjugate pairs associated with y and z coordinates of position and their corresponding momentum components. The operators for *different* directions in space thus *commute*; but those for the *same* direction form conjugate pairs with non-zero commutators. That was really all you needed to know about Schrödinger's great invention!

- In Chapter 4 you were able to solve the eigenvalue equation for a number of simple 1-particle systems, using Schrödinger's wave mechanics. Most of these examples involved setting up and solving very simple differential equations (much easier than the stuff in the earlier chapters!) which you'll use again and again. But one, for the harmonic oscillator, was also solved by a symbolic method based only on the commutation rules just to show that sometimes it *can* be done. The chapter ended with a review of some general properties of eigenstates, to be kept for reference.
- Remember that in the early days of quantum mechanics it seemed impossible to decide whether an electron, say, was 'really' a particle or 'really' a wave? In Chapter 5 you saw how the riddle was solved by allowing $\Psi(x, y, z)$ to depend also on *time* and to move through space. You started from what you'd learnt about **travelling** waves in Book 10, which allowed you to define $\Psi(x, y, z, t)$ and to find a differential equation it must satisfy: Schrödinger's equation including the time.

By combining an infinite number of travelling wave you were able to construct a strongly localized 'pulse' (like the one shown on the front cover of the book) and to show that it moved like a classical particle, but with coordinates and momentum components replaced by their *expectation values*. So Newton's law that "force=mass×acceleration" still works, even at a sub-atomic level – provided all quantities are interpreted as expectation values.

You also learnt that wave functions of the 'separated form' $\Psi(x, y, z)f(t)$ could be found when $\Psi(x, y, z)$ satisfied the energy eigenvalue equation and the time entered only through the complex exponential phase factor f(t). This confirmed that the energy E was a **constant of the motion** and it followed that other quantities could also remain constant, provided their operators commuted with H. For a system like an atom, the most important quantities of this kind relate to **orbital angular momentum**: you were able to find a complete set of commuting operators, H, L^2, L_z , for motion in a central field, and to explore the correspondence between 'classical' and 'quantum' pictures of the motion.

• Chapter 6 opened with a more complete treatment of the angular momentum operators, using only the abstract **commutation relations** satisfied by L_x, L_y, L_z , which apply to *any* kind of angular momentum. As the operators for different components do not commute, Ψ could be an eigenfunction of only one of them – which you chose arbitrarily as L_z , thus fixing one 'special' direction in space. The states of 'maximal knowledge' then satisfied $L^2\Psi = L(L+1)\Psi$ and $L_z\Psi = M\Psi$ where the **quantum numbers** L, M took integer values, M going down in unit steps from M = +L to M = -L. The energy states could thus be classified according to values of E, L, M: they were named s-states, p-states, d-states, f-states for L = 0, 1, 2, 3, respectively.

The first few 'families' of atomic energy levels were depicted in Figure 15, where levels of given L are (2L+1)-fold degenerate for the allowed values of M. You were able to find the *energy* eigenvalues only for the s-states, where the ∇^2 operator took a simple angle-independent form, but there are 'accidental' degeneracies between corresponding levels in the different families. The existence of true degeneracies results from the **spherical symmetry** of the Hamiltonian: changing the perfect symmetry, even slightly, 'splits' some of the levels. This is revealed in the **atomic spectra**, which result from *electronic transitions* between different levels. All this dependence on symmetry called for a proper explanation.

• In the following Chapter 7 ("More on symmetry") you came to grips with what the idea of 'symmetry' really means and how it can be used. Again you had met symmetry long ago, in Section 6.1 of Book 1, and by now you knew how operations like turning things round in space could be *represented*; so it was no surprise to find that **matrices** could be associated with operations in such a way that combining operations, by performing them one after another, could be 'echoed' in combining the matrices under **matrix multiplication**. You were reminded of such things in Section 7.1, with worked Examples getting you as far as basis change and unitary transformations.

The special feature of central-field systems is their **spherical symmetry**: rotating the system around the nucleus makes no apparent change – the Hamiltonian H is **invariant** against all such rotations. The basis vectors in 3-space carry a matrix representation of the **rotation group** and the representation could be changed by choosing a new set of basis vectors; you were able to distinguish two interpretations of such transformations ('active' and 'passive').

You learnt how a 3-space symmetry operation R could *induce* a related operation U_R in, for example, the Schrödinger space of wave functions and differential operators; and were able to understand why the s-,p-,d-, and f-type atomic orbitals came out in **degenerate sets** containing, respectively, 1,3,5,7 functions.

• Chapter 8, on the 3-dimensional rotation group, looked even more difficult then anything you'd met so far and you may have decided to stop at this point. You already have all the basic principles you're likely to need in the *applications* of quantum mechanics – even to many-electron systems like atoms and molecules (which you can find in Book 12).

But it's worth at least glancing at Section 8.3, because it will help you to get a deeper understanding of all you've done. It shows how **infinitesimal rotations** in space, out of which all finite rotations can be built by repetition, have exactly the same commutation properties as the Schrödinger operators L_x , L_y , L_z for angular momentum components. And similarly **infinitesimal displacements** through space, along the x-, y- and z-axes, have the same properties when acting on a wave function as the *linear* momentum operators p_x , p_y , p_z . The conclusions were:

A system whose H is invariant against linear displacements through space may have operators $\mathbf{p}_x, \mathbf{p}_y, \mathbf{p}_z$, that commute with H: the corresponding quantities E, p_x , $p_y p_z$ can the all be simultaneous constants of the motion. (The simplest example is a particle in free space, moving in a definite direction with definite momentum and energy.)

A system whose H is invariant against *angular* displacements around some axis may have operators L_x, L_y, L_z , that commute with H: E and any one of the corresponding quantities $L_x, L_y L_z$ can then be simultaneous constants of the motion, depending on which axis is chosen. (Think of a particle in a central field, moving with definite energy and with angular momentum of definite *magnitude* around one axis.)

In fact, Schrödinger didn't really have to *invent* the operators of wave mechanics: they were there already, expressing the properties of the space we live in – they just had to be *discovered*!

• Perhaps Chapter 9 was also too difficult for a first reading; but it showed how it was possible to unify all that had been done previously in quantum mechanics by using a simple and beautiful new formalism. Wave mechanics and matrix mechanics were just different ways of expressing the underlying theory: both were related by Dirac's **transformation theory**. And in Section 9.4 it was shown how a transformation to **momentum space** could give yet another version of quantum mechanics.

We end this book on the principles of quantum mechanics, marking the end of the "quantum revolution", by quoting Dirac himself:

"The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved."

Since then (1930), when it was necessary to do the mathematics on a mechanical calculator with a handle, enormous progress has been made and the difficulties forseen by Dirac have been largely removed. A good understanding of the electronic structure and properties of atoms and molecules, at least at an approximate level, can be obtained and used in whole new areas of Science. Some such applications will be the subject of Book 12.

Index

Absorption and emission 50,89 Action 34,49 Angular momentum eigenstates of 80-82 operator properties 75-76 orbital motion in central field 77-78,145 spin 8-12 see also Commutation relations Assembly 1 Atomic orbitals 85,87,91,147 Atomic units 53-54 Basis vectors 7,9,13,56,94,126 spherical 114 Bohr radius 54 Bound states 85.87 Boundary conditions 40-42 periodic 55 Bra-ket notation (Dirac) 125 Commutator, of Hermitian operators 30 Commutation relations angular momentum operators 74-75 infinitesimal operators 118-121,147 position and momentum operators 36 spin operators 11 Complete set (of operators) 27 Components of state vector 5,13,56see also Expansion coefficients Constants of the motion 74,77-78,145 Correspondence principle 33,60 Definite values 38,78 Degeneracy 51, Delta function (Dirac) 133 Delta symbol 97,126 Differential equation 37,46,52,57,84 recurrence relation for 84 Differential operators 34,36 Diffraction 59

Dirac formalism 125-139 Dirac transformation theory 134-137 Direction cosines 9,97 Dispersion (wave packet) 66 Duality 59 Dynamical variables 35,67 and associated operators 33-36 compatible/incompatible 35 conjugate pairs of 35 maximal set 36 Ehrenfest's theorem 73 **Eigenfunctions 56-57** Eigenstates 3 of 2-spin system 16-19 Eigenstates (General properties) 56-58 Eigenvalue equation 7-8 Eigenvalues 3,7-8 of Hermitian operators 57-58 **Eigenvectors 7-8** Energy conservation 43 Energy level diagrams 44,47, Expansion coefficients (components) 5 in Dirac notation 127 significance of 14 Expectation values 6,8,25,69-73 Function space 56-58 Group (definition of) 102 and matrix representations 94-98 see also Rotation group (3-dimensional) Group velocity (wave packets) 65-67 Hamiltonian, function and operator 38 Harmonic oscillator 43-51 Hartree energy unit 53,89 Heisenberg 125 matrix mechanics 128-130 Uncertainty Principle 30-33 Hermite polynomials 48

Hermitian matrix 20 Hermitian operator 21 Hermitian scalar product 20,56 Hermitian symmetry 57 Hermitian transpose 98 Hilbert space (*see* Function space Hydrogen atom 51-54,79-93 angle-dependent wave functions 90 energy levels for s,p,d,f states 88 forms of atomic orbitals 87,91 spectral series 88 s-type wave functions 85

Ideal measurement 1, Induced rotation 102,107,120-121 Infinitesimal operators 118-121,147 for translation (\rightarrow linear momentum) 122 for rotation (\rightarrow angular momentum) 123 Integral equation 132 Integral kernel 132-133 Integral operator 13 for delta function 133 for KE and PE 133 Interaction, radiation-matter 89 Invariance principle 9,141 Invariance under symmetry operations 122-124 Inverse matrix 98 Ionization energy 85

Kernel *see* Integral kernel Ket *see* Bra-ket notation Kinetic energy (KE) 38

Laplacian operator 38,52 Linear dependence 56 Linear operators 56 Linear vector space 56

Macroscopic/Microscopic 1 Magnetic moment of spin 2 Mathematical induction 71 Matrix adjoint *see* Hermitian transpose diagonal form of 129 Hermitian 20 Hermitian transpose 98 inverse 98 metric 20

orthogonal 98 unit 98,1 unitary 98 Matrix conventions 96-97 Matrix element (Dirac form) 127 Matrix mechanics 125,128 Matrix product (definition) 95 Matrix representation of operators 95-97 of rotations 97 see also Rotation group (3-dimensional) Maximal set 36 Mean value see Expectation value Metric matrix 20 Momentum (linear) and associated operator 36 eigenstates for free particle 39-42 see also Angular momentum Momentum space 134-139 Newtonian dynamics as a limit 74 Nodal surface 87 Normalization 14 delta-function 137 of vectors 131 of wave functions 131 of free-particle eigenfunctions 54-55 Number operator 129 Observables 1-2 and associated operators 7.14-15 see also Dynamical variables **Operators** 5 adjoint 21 anticommuting 9 commuting 9 differential 5 for 2-spin system 15-17 Hermitian 21 idempotent 27 identity (unit) 8,27,127 linear 8,56 projection 5,13self-adjoint 21 step-up/step-down 11-12,21-22,49,80-82 unitary 101 Orthogonality 4,14, Orthonormal basis set 57,97,126

Oscillator see Harmonic Oscillator

Partial differentiation 36 Particle in a box 38-43 Particle in free space Particle in parabolic field 45-51 Pauli matrices 94 Periodic boundary conditions 54-55 Periodic function 40 Phase factors 10,11,68,82 Photon 59 Planck's constant (rationalized, \hbar) 8 as unit of angular momentum 34 Potential energy (PE) 38 Polar coordinates 90-91 Probability 14,41,143 Probability density 25,47,54,131,147 Probability distribution 28, Gaussian 29 Projection operator 5,13-15, ket-bra form of 127 p-space see Momentum space Pure state 27

Quantum of energy 49,89 Quantum mechanics 7,8 Quantum numbers 12,41,87

Radiation, absorption and emission 97-99 origin of spectra 92-93 Recurrence relation 84 Root-mean-square deviation 30 Rotation Euler angles of 109-110 infinitesimal operators of 118-121 of a function 102,104of an operator 102 Rotation Group (3-dimensional) 106-124 representations carried by monomials 111-113spherical and Cartesian bases 114 spherical harmonics in p- and d-orbitals 116-118 spin-space representations of 106-109

Scalar product 5, of functions 57,126

of vectors 5,14-15, Hermitian 20,56,126 Schrödinger representation 24-36 Schrödinger equation 38 for free particle 39,54 for harmonic oscillator 43-51 for hydrogen atom 51-54,79-93 for particle in a box 38-43including the time 67-69 integral equation form of 132-134 matrix form of 132 Self-adjoint operator see Hermitian operator Separation of the variables 42,51,90 Simultaneous eigenstates 78 Spectral series 88 Balmer, Lyman, etc 90 Spherical basis vectors 114 Spherical harmonics 126,127,124 in atomic orbitals 117 Spherical symmetry 146-147 Spin 2 derivation of operator properties 8-12 eigenstates of S_z and S^2 Spin-space representations 106-109 State 3 represented by a vector 4,8,13,125,134stationary 37,58 States, pure and mixed, 27 Stern-Gerlach experiments 2-4, Sturm-Liouville equations 57 Symmetry and degeneracy 103-105 Symmetry group 98 Symmetry operation 99,103 Transformation, as basis change 99-100 active and passive 100 of operators and vectors 101

Uncertainty 29-30 Uncertainty principle 30-33 Unitary matrix 98 Unitary operator 101 Unitary transformation 99-101

Transformation (Dirac theory) 134-137

Vector space (basic properties) 56 (see also Function space) Wave amplitude 46 frequency 62 period 62 phase velocity 62profile 60 standing 62 travelling 63velocity see phase velocity wavelength 39,61 wave number 39 Wave equation see Schrödinger equation Wave function 26,33 Wave Mechanics 33-Wave packets 63 Gaussian 64-65 group velocity of 65-67 motion of $69\mathchar`-74$ see also Ehrenfest's theorem Wave vector 138