Elementary Quantum Mechanics

Han-Sen Choong

March 2, 2025

Chapter 1

Chapter 1: Exoskeleton

1.1 Preface

The sudden, explosive rise of quantum physics is probably both one of the most exciting events in human intellectual history. Born in the ashes of a Newtonian model believed for two centuries to have been unquestionable, and forged by a cohort of Physicists who lopsidedly dominate the historical halls of fame in both number and scope of achievement, quantum physics has never been uncontentious, never obvious and never immediately complete. Yet in just 50 years, it ascended to full prominence, and its ideas, practices and results took to centre stage. It was, really, a revolution.

This revolution, speaking from an academic point of view, was radical for two main reasons. The first was that physics had to fully accept that validity was derived from result and result alone. No matter how offensive the theories of quantum mechanics were to human intuition, the truth, if it lay beyond human imagination, nevertheless had to be accepted. The priority was achieving the results corroborated empirically even when everything seemed wrong to common sense. This mindset is critical to any quantum mechanics student, especially when they are starting off and unable to grasp the nuances of quantum mechanics well enough to even fathom debating the philosophical side of its rules. Indeed, it is the approach I take in this book. When a theory can make predictions accurate to the nearest ten billionth, as seen in quantum electrodynamics, and is not invalidated in any obvious way, it must be at least relevant. Studying quantum mechanics is central to

science, and it will be still for quite some time. That should suffice for us, and we will not question the model at all here.

Secondly, quantum mechanics was a revolution in the history of applied mathematics. When the earliest serious formulations of quantum mechanics came in the mid 1920's, what belied them was mathematics: and not only mathematics, but fringe, new and diverse mathematics being unified to a single purpose: to create a theory which could explain the unexplainable results Physicists were obtaining at the time. The genius of Dirac, Heisenberg, Born and other founders was that they could draw from a purely abstract discipline and stamp it onto the real, physical world. Mathematics belying physics has always been a well-understood fact, but the level we are are working at is much more sophisticated than for classical, Newtonian, mechanics. Even a surface incursion into quantum mechanics will start reaching towards the deepest vestiges of mathematics in all corners.

This will all present a challenge to the reader, if this is the first time they have handled any theory of this level of complexity. In this sea of randomly drawn abstractions and deeply anti-human results, confusion is a serious threat. It is quite easy in quantum mechanics to read 400 pages of textbook and still have many unanswered questions about its basic rules. Assuming one can learn quantum mechanics as one might learn differential equations – crunching problems and becoming an expert through practising the mathematics – is naive. There are rules to learn, and concepts to understand; we cannot try to jump logically across the entire subject, if we cannot understand what the rules do and do not impose on the nature of our manipulations.

Thus the aim of this book is to avoid confusion at all costs, and to be pragmatic. Most, if not all, of quantum mechanics textbooks work on the basis that students reading those textbooks are propped up by a rigorous undergraduate semester and experienced professors explaining the nuances in lectures and seminars and even practical experiments while working with the same textbook. Self-learning quantum mechanics from such books, particularly for less mathematically secure students, is risky, because they are far too brief on explanation in their bid to prioritise swathes of algebra and computations, assuming experienced professors will do the rest for their usual readers. It is by no means impossible, especially for those trained in undergraduate mathematics, but it is likely ill-advised for the rest of self-learners

to jump straight into quantum mechanics via standard textbooks like those. Conversely, to those for which all the mathematics here comes in the blink of an eye, or who have finished studying a quantum mechanics textbook or course already, this book will comparatively less useful, because it involves very detailed and explicit proofs and explanations which are optimised for those learning quantum mechanics completely from the book, rather then those looking for a reference textbook. If you like, this book can serve to be a quick map of the mathematics of quantum systems; in the author's opinion, such a map is more helpful in quantum mechanics than virtually any other subject, because of the number of non-intuitive, and interlinking, assumptions made in its formulation.

In the end, this book is but a stepping stone. A stepping stone which I believe is very useful for a new student, to set principles in order before entering the true deep end, but also a stepping stone which is written to be a stepping stone. This book won't give you everything, but, simply by painstakingly avoiding a whole lot of bad, I hope it can give a self-learning student new to quantum mechanics a whole lot of good.

I hope you will enjoy the book.

Thank you, Han-Sen Choong

1.2 What is a Physical Model?

One of the most difficult things in starting to learn quantum mechanics is that it can be very difficult to tell between concrete rules (of reality), chosen assumptions and mathematical logic. It is my belief that a new student of quantum mechanics must be able to distinguish between these as early as possible, because otherwise there are dangers of confusions where one questions assumptions as if they are concrete logical consequences when in the very first place they cannot be subject to such inspection. Thus, I think, it is useful to understand quantum mechanics as a discipline before delving into its details so that there is at all times in our learning of quantum mechanics this awareness of why exactly its components are arranged as they are. So let us begin with a question. What is the difference between quantum mechanics and quantum physics?

The question may have been in the mind of the reader with the loose treatment of both terms in the Preface, and indeed if they have heard both terms elsewhere without knowing their distinction. In fact, the term 'quantum physics' will soon become extinct in the rest of this book, because the book is a quantum mechanics textbook– not a quantum physics textbook. The difference is uncontentious; every classification would put quantum mechanics as *the rules* and quantum physics as *the application*. The latter is more advanced and difficult, but it is also wholly predicated on the former.

Quantum mechanics, easiest thought of then as the pure mathematical component to the study of quantum phenomena (sans experimentation), is what we call a **model**. Specifically, quantum mechanics is a model of physical reality, and thus we might refer to it as a physical model. However, it is the rationale of modelling which is a key characteristic of quantum mechanics as an academic discipline. In modelling, we take a situation where we have some computation or prediction we want to make, and we take the tools we have available to us from mathematics to create a system which is to make such predictions without any additional predictions which conflict with our desired subject. We choose the rules of our model based on what makes sense – one would never model the probability of a train arriving x minutes late with an exponential, for example, because the probability of a train arriving 100 minutes late should not be exponentially greater than it arriving 10 minutes late. How perfect are our rules? That would depend on the purpose of the prediction we are trying to make. In the situation of a train's arrival time, knowing the train has a probability 0.498263 of arriving at a specific time is no more useful than knowing it has 50% probability, rounded to the nearest percentage point. In quantum mechanics, on such microscopic scales, we would like to be as accurate as possible, but yet our greater priority is making sure there are no glaring discrepancies between predictions made by the quantum model and physical experimental results. This latter condition of being true to real world experimental results is clearly prerequisite for any physical model, and the reason why quantum mechanics needed to be created in the first place was because the previously widespread model of Newtonian classical mechanics was proven incompatible with the results of new and radical quantum experiments.

Now we can understand our task in this book as engaging with the model

of quantum mechanics. Such consists of working with and understanding the rules the original Physicists creating the quantum model set out. These rules are called the **postulates** of quantum mechanics. This word 'postulate' is critical – as it means that these rules are *assumed*, just like any mathematical model would assume rules and be tested on how well the resultant model fit its requirements. Thus the postulates of quantum mechanics are fundamentally different to other 'rules' the reader might have come across in earlier mathematical studies. Those rules – like the Pythagoras theorem – are hard and concrete, and arise *logically* from the definitions of a right-angled triangle, and side length. These rules, the quantum mechanical postulates, are *assumed* rules by which we choose to abide! The difference is of course critical. Pythagoras' Theorem is not an assumption: it drops out of a construction we hold to be true. These quantum mechanical rules, however, don't drop out of logic. They were chosen, by the founders of quantum mechanics, to predict reality, and they were kept mainly because they worked.

Now we return to the ideas of the Preface. In all modelling, the rules we assume are simply validated by the results of the predictions of the model compared to the results obtained in that which we are modelling. If the model is accurate and performs its functions well, we can then trust its rules are at least reasonably sound, and that studying that model is useful and relevant. This is the story of quantum mechanics, which is indeed the most successful mathematical model ever created. The Postulates we will be learning, in this book about quantum mechanics, are what define this quantum model we will be working with.

Now, it is useful to undergo an understanding of what physical modelling– in many ways, essentially no different to theoretical physics– really aims to do, because most students will not have learnt this explicitly before. Physical modelling (physics) has two aims– problems to answer:

1. The State Problem

2. The Time-Evolution Problem

If a physical model can fully answer both these problems, then it is considered complete.

The State Problem

The state problem concerns our ability to describe any state at any given time. It is therefore a problem of description, encapsulation and extraction.

Description is the most obvious criterion in the solution of the state problem. It pertains to our ability to observe a state and be able to put it down into description– usually, mathematical description. It is only after this that we can bring anything successful to practical experiment, as there is no point trying to find data on states if there is no standard system whereby states can be distinguished and results can be recorded down.

Linked with that description component of the state problem, then, is the extraction problem, which is another key component of the state problem. If we have some state we have observed, and we have recorded its information down as some solution to the problem of description, we have gone from an observation to an abstract representation. However, we also want to be able to take an abstract representation and be able to at least imagine or compute the characteristics of the physical state it represents. Consider a situation, for example, where we can isolate air molecules and work out their speed in a simulation, and we decide to colour any air molecule with velocity above 550m/s red in our simulation, all air molecules with measured velocity 525-550m/s yellow, and all below 525m/s blue. That would be an attempt to tackle the description component of the state problem. However, it would also clearly be incomplete as a model of reality. The reason for that is simple: given a physical state (specifically, an air molecule having a certain velocity), we can certainly describe it using our colouring system. However, given a certain colour, we cannot at all work out anything about its velocity, other than how its value relates in magnitude to the value of 550m/s. In this case, it is clear that we can describe states better than we can extract information from a description, and it is clear this loss in reversability of information is a problem.

That example, and the final component of the state problem in encapsulation of information, will become more important very soon. For now, it should be very clear that the state problem is quite relevant, not always straightforward (especially after the groundbreaking results of the Stern Gerlach experiment we are about to study) and deserving of a place as one of the two most

important problems in Physics.

Time Evolution and Observables

The definition of the time-evolution problem is luckily much easier. Given a particular state, we want to be able to calculate how that state will evolve in a certain period of time into a new state (subject perhaps to time-evolving external conditions). This is the physics problem all readers will be familiar with: whether it be through acceleration-time graphs or via other common time-based experiments.

Such time-evolution is usually governed by an equation. In quantum mechanics, this is, as we will discover in Chapter 5, the widely famous Schrödinger Equation. Equally, we note that a time-evolution equation clearly creates a greater onus on the results of the state problem in our model to be mathematical objects, so that these can actually be run through the equations we have.

We here have built an understanding of the central aims our physical model needs to meet, and in framing this beforehad, such discussions should no longer seem to come out of nowhere when we reach them by Chapter 3.

1.3 A Structure

If the reader is a student who feels high–school mathematics is trivially within their grasp, as indeed most students who dare start learning quantum mechanics are, you may start reading Chapter 2 immediately. If you are a high–school student and do not think you have learnt every corner of the syllabus yet– start with Chapter 10, where a list of prerequisite mathematics is found in reasonable detail. As described there, you should by no means deprive yourself of a chance to get an early foothold into this significant subject even if out of just interest, but you should also open up some simultaneous study of those areas listed using textbooks written to present those areas, rather than trusting 20 bullet-pointed pages of a quantum mechanics textbook to teach you those crucial foundational concepts which really require more familiarisation.

Now, this section will present an insight into how this book will set about

accomplishing its task of providing a pragmatic entry into quantum mechanics which avoids chronic confusions, building off the Preface.

The first thing to make clear is that this is a mathematics textbook. Certainly, it *much* more verbose than a traditional mathematics textbook, because I prioritise, as I have already stated, an understanding and clear picture of the principles over the cleanliness of logical arguments. However, none of the mathematics here is watered-down, inaccurate or simplified. Instead, results which are too mathematically complicated to prove rigorously are either delegated to the many larger, comprehensive textbooks existing on the subject, or presented as a result with the lack of proof clearly flagged.

The second thing to note is that there will be a notation switch in Chapter 7. This structure is rare in quantum mechanics books: most of the time, an author chooses one notation and sticks to it. The reason I have chosen to set up this book with a notation switch is because I believe the conventional quantum mechanical notation is too good to not be learnt, but at the same time too novel and therefore distracting to focus on right at the beginning. The reader should bear in mind that no area in this book is wrong, and the notation shift will only introduce new perspectives on the algebraic manipulations of quantum mechanics, rather than replacing or correcting ideas covered in preceding chapters. I simply prefer that the reader struggles a bit with some extra reformulation there than struggles with notation, mathematics and conceptual understanding all at the same time, and in any case the demand of that switch is as much aesthetic as anything else, which should be acceptable, especially as the reformulated notation will be the one any more advanced books than this one will continue to use.

The book can be split into two constituent parts, therefore: before and after Chapter 7. Chapter 2 starts by providing motivation for the developments of quantum mechanics with the exemplifying Stern Gerlach experiment, and will be critical in laying out the new physical results which had to be accounted for. Chapters 3, 4, 5 and 6 (quantum states, quantum observables. quantum simultaneous states and time-evolution respectively) lay out the quantum mechanical model and its rules with detailed study of subsequent implications both physically and algebraically.

The second part of this book can be characterised by the remaining chap-

ters, which treat the mathematics more seriously and vigorously than the first part of the book. It will take the model we built in Chapters 3, 4, 5, 6 and recodify it in a way which allows us to understand the algebraic manipulations accessible to us in a more lucid and detailed manner. Finally, after doing all these things, the reader will be ready to move onto physical problems for the first time, and with some practice, will meet the end of the content of this book in Chapters 8 and 9. The closing problem will be the solution of the hydrogen atom, and we will witness a number – and a perfectly accurate number at that – drop out of the quantum mechanical model for the energy levels of hydrogen, which, astonishingly, is corroborated by empirical evidence. That should provide a satisfying conclusion to the journey of starting off with quantum mechanics, before the final chapter has been included as a conclusion to the book which I believe best sets a student on a path moving onwards from this text to learn more advanced quantum mechanics. Hopefully the reader will be in very robust shape after all is said and done to press onto more advanced textbooks when their mathematics allows it, and then I will have successfully completed my task in writing this pragmatic exposition.

Contents

Chapter 2

Chapter 2: The Stern Gerlach Experiment

In the Exoskeleton to this book, I presented the idea of modelling very explicitly to the reader, because I think that a more authoritative overview on the *procedure* of what we are trying to do is important and often neglected by compendium-style textbooks. Considering this is a mathematics textbook, we have opened up with a whole lot of words. However, this is, to my estimation, the best way, and the result is that the reader should get a 'bird's-eye' view of the subject which will be a helpful guide throughout, before we open the algebra and our horizons immediately become narrowed down for some time.

To complete our 'bird's-eye' view over quantum mechanics, then, we will now investigate one of the radical experiments of the early 20's which necessitated the birth of quantum mechanics to take the mantle from a classical model which was suddenly proven woefully inadequate. I have chosen the Stern Gerlach experiment, involving electron deflection in a magnetic field. Other practical experiments, such as the Davisson-Germer modification of Young's Double Slit or the Mach-Zehnder Inferometer, are also often used as opening preludes to the quantum model, but the Stern Gerlach experiment is amongst these the easiest-accessible exhibition of the phenomena we will be trying to mathematically model, without much additional Physics knowledge required. What we herein are trying to understand is what exactly the demands– from physical reality– quantum mechanics had to meet were. It is a model, and models are valuable only based on how well they meet the

demands of the predictions they need to be capable of making. With these extraordinary experimental results, quantum mechanics very much was given a deeply challenging and novel set of demands, so its rules – the Postulates – can be said to be directly motivated by them.

Thus, understanding these motivations provides a stabilising force of unification between the otherwise unexplainable haphazard mathematical abstractions which will follow. The author strongly suggests understanding Planck's (discretised) solution to the blackbody radiation problem, Bohr's new atom model, Young's double slit experiment, De Broglie's matter waves, and Born's radical idea of "probability clouds" before beginning to read this book. These need not be learned out of textbooks or with any length or depth of study at all– only the most superficial familiarity with the concepts is needed– but are very useful to have a basic grasp of in order to understand what new issues quantum mechanics had to address when it was conceived. We cannot cover all the above physics experiments in any true detail, because doing so not only violates the mathematical focus of this book– dissipating the reader's attention with phenomenological distractions– but also requires far more complex physics than can fit into this textbook.

Now, then, if the reader feels introduced to the ideas of inherent probability and superposition, we will press on with a very concrete example to see these principles in real world practice.

2.1 The Stern-Gerlach experiment

The idea of inherent probability has been introduced and is well covered even by non-mathematical media on quantum mechanics. If the reader has familiarised themselves with these background ideas of inherent probability, we can investigate this further, and the consequences it has on the quantum state, by turning to the Stern-Gerlach experiment. The reason this experiment is so useful is because it allows us to focus on a particular *observable*– that is, something we measure in an experiment– which can only take two possible values. This allows greater lucidity in how we organise a systematic investigation into the strange quantum mechanical phenomena of probabilities.

The setup deals with the magnetic properties of silver (Ag) electrons. Specifically, we consider what happens when we pass them through a magnetic field. What we expect is for deflection to occur; this is due to a property of electrons, called spin, which is directly influenced by the magnetic moment exerted on the electrons by the magnetic field. A problem with spin is that it has no classical analogue- that is, we have not seen any such quality with our own eyes in our macroscopic world; indeed, the physicist Pauli describes spin as something 'not describable classically'. Nevertheless, all the reader needs to know is that this quality determines the different ways an electron is deflected by the magnetic field, and that it is a measurable quality. We can reduce each electron to be considered as a single point particle with a particular spin, by performing some net cancellations. So we will make that assumption that the particles we are working with in the experiment are single point particles with some spin values.

Now, after electrons are deflected by this magnetic field, they are detected by an electron detector on the other side, which we can visualise as a simple screen which shows where the electrons land. The expectation classically is that there is a large patch of electrons formed on the detector screen, bounded by the furthest deflections achieved (and possible) during the experiment. This is because we expect that spin is some continuous value which may differ from electron to electron, resulting in slightly different respective deflections by the magnetic field: thus forming in this patch on the detector screen.

However, the results which are achieved are rather different. There are only two positions at which electrons are detected after deflection on the screen. If we proffer that this quality of spin is the only facet which changes the deflection– other than the magnetic field, which is the same for every electron passing through it and constant throughout– then this necessarily means we have observed the *quantisation* of spin: the property of an observable having a finite set of discrete values– here, two. There are two possible deflections, and therefore there must be two possible spins given that nothing else is changing.

In the above we could have made a subtle semantic adjustment and said: "there are two possible deflections, and so the electrons must have one of two possible spins before passing through the magnetic field". We will show that,

while this might seem sensible to classical intuition, it is in fact impossible for us to say this because it is impossible for an electron to have a specific value of spin before passing through the magnetic field! This is a dramatic statement, because under classical intuition every object we study must have some value for any observable at all times. We can change the momentum of a baseball by striking it, but that doesn't mean it did not have a momentum before we struck it and measured the result. Somehow, quantum mechanics tells us that with spin, the electron can exhibit a certain spin value after it passes through the magnetic field, but does not necessarily have one before it does so. This is a prime example of purely quantum behaviour, and was indeed why the Stern Gerlach experiment, which strongly suggests this fact, was so seminal as one of the experiments which necessitated the birth of quantum mechanics.

To understand how the results of the experiment show us this strange behaviour must exist, we will engage in some sort of proof by contradiction. We will assume the opposite is true: that the electron must have a spin value at all times, and show that the results of the experiment lead to a contradiction if this assumption is true. By doing this, we will thereby prove our assumption is incompatible with the results of experiment, and that therefore the electron cannot in fact have a spin value at all times.

In our work, we will focus on the two-dimensional Stern-Gerlach experiment. The primary dimension we focus on will be the y-dimension– or, as we will think of it, the "vertical" direction. Here, the two possible spins in our experiment are "up spin", and "down spin". More interesting things come when we consider a second dimension, so we will use the x-dimension as our secondary dimension, where the two spins are "right spin", and "left spin". Corresponding to these dimensions there are the y-dimension magnetic fields and x-dimension magnetic fields, which deflect electrons in those dimensions alone.

Let us now take our classical assumption on until we reach a contradiction: we will assume that the electron must already have some value of spin before passing through the magnetic field. Thus begins our proof by contradiction.

Let us first consider the purposes of the magnetic fields. It might seem like they are performing a physical action on the electrons- which they are:

they are deflecting them. However, the real purpose of the magnetic fields in this experiment is to *reveal* the spin value of electrons which we assume they possess. By passing a single electron through the y-dimension magnetic field it is either deflected down, if it had down spin, or it is deflected up, if it had up spin. In this way, the action is of passing an Ag electron through the magnetic field is the action of measuring the spin it has, because we reveal the spin through observing the direction of deflection.

The second quality of the magnetic field is that it separates a group of electrons into two new groups corresponding to the two possible spins in its dimension. Note that it can do both of these functions at the same time: it deflects down spin electrons down and up spin electrons up, which both splits them into distinguishable groups, and measures (reveals) their spin. After passing a random beam of electrons through the magnetic field, we will get these two resulting groups split by their initial spin: surely, a useful thing to have.

If this is the case, then we expect successive y-dimension magnetic fields to yield a clear result when applied to a group of electrons. The first y-dimension magnet separates the group of electrons into electrons with up spin and electrons with down spin. Suppose we allow for enough distance to distinguish the two separate deflections of up and down, and place two new magnetic fields at that distance, one in the up deflection position, and one in the down deflection position. Then, we therefore expect the second layer of magnetic fields to perform the same measurement function, but with different results since the original magnetic field originally separated the inital beam into two groups based on the two possible spins. Specifically, the second y-dimension magnetic field in the up-deflected position should measure all its electrons to have up spin, and the second y-dimension magnet in the down-deflection position should measure all its electrons to have down spin. This is of course because they have already been separated into two distinct groups and measuring the spin of all the electrons in those distinct groups will only yield one result, under our assumption that the electrons hold a spin value at all times. We did not know the spin value of the electrons before we passed them through the first magnetic field, but after we do all our revelations are done.

This is indeed what we achieve when we do actually perform the experi-

ment. The first y-dimension magnet splits a group of electrons into two groups corresponding to each spin, and an infinite succession of y-dimension magnetic fields on each of these deflected trajectories will continue to measure the same spin by continuing to deflect them in the same direction. The same holds for successive x-dimension magnetic fields: the first separates the electrons into two groups with left and right spin respectively, and the subsequent magnetic fields will continue to measure the same spins if applied successively along the two initial trajectories obtained by the first magnet.

Of course, physicists were more inspired than just applying the same magnetic field over and over again on the same electrons. We might wonder what would happen if we applied some sequence of y and x magnetic fields. Let us note that electrons passing through a magnetic field in a certain dimension are not deflected in anyway in any other dimension. We implicitly see this in the fact that electrons passing through the y magnetic field follow a straight trajectory with respect to the x-dimension and therefore still all hit the detector screen a certain x distance away from the field.

Our assumption is that an electron at all times holds a value for its spin. We therefore also posit that an electron may at one time hold two simultaneous properties of x spin and y spin, which are independent of each other: just like a ball may have a y axis momentum value and an x axis momentum value at the same time. We test the relationship between the x and y spins of an electron: is an up spin electron more likely to be a right spin electron, for example? Understandings like this are physically relevant, of course- if a electron passing through 'unbiased' magnetic fields was more likely to be deflected right and up or left and down then this would surely have physical implications.

Suppose we isolate a certain y-spin and only consider electrons with that y-spin: say, up spin. To do this experimentally is simple– pass a group of electrons through a y-magnetic field, give sufficient distance for the deflected trajectories to diverge, and block the down-deflected path so that none of those electrons can keep moving forwards and the only ones moving forwards are the up spin electrons. Let us suppose we did this. Then, we can investigate what happens when we place an x-magnetic field in the path of the up-deflected electrons. The point of blocking out the down-deflected electrons in this way is to assure us we will past this point never have to deal

with down spin electrons during our considerations.

The x-magnetic field on any ensemble of electrons will not show considerable bias towards any spin, in fact. We obtain two new deflected trajectories, therefore: up-left and up-right. Again, this all seems good, until we add another y magnetic field, after which a bizarre phenomenon occurs.

If we add another y magnetic field to the trajectory of the up spin electrons which we have just deflected left or right with the x magnetic field, we would expect all the electrons to be deflected upwards, since these are after all electrons with up spin. We know that we have blocked out all the down spin electrons far back, so they cannot escape to intervene with the results of our current experiment.

Such is not the case. What we get instead is that both the up-left and up-right ensembles of electrons have subgroups of electrons which are deflected downwards by the new y magnet! This by all means should absolutely be impossible, given we have filtered out all the down spin electrons already. Clearly, something is wrong.

To keep track of everything following this revelation, we will use a new notation: $(\uparrow)$, $(\downarrow)$, $(\leftarrow)$, $(\rightarrow)$ for different spins, as well as two-fold notations such as $(\uparrow\rightarrow)$ to record when an electron has had its y spin and x spin measured. This is in a sense a crude version of a 'state notation', as we are attempting to label the state of a given electron– at least with respect to its spin properties. The reason this is so useful is, among other things, simply that it saves us a lot of time by avoiding having to write "up-right spin electron" all the time.

This state denotation model is just some arbitrary and meaningless labelling, so one should not carried away with its importance other than for making things more concise. Nevertheless, let's take this state denotation one final step further by adding scalars to the mix. Specifically, we will use scalars to show how many electrons are in each state for a given set. For example, we could have

$$1000e^- = 500(\uparrow) + 500(\downarrow).$$

That is: we have some 1000 electrons, 500 of which are up spin and 500 of which are down spin. Similarly, we could say

$$500(\uparrow) = 250(\uparrow\leftarrow) + 250(\uparrow\rightarrow)$$

or even,

$$500(\uparrow) \implies 250(\uparrow\leftarrow) + 250(\uparrow\rightarrow) + 0(\downarrow\leftarrow) + 0(\downarrow\rightarrow).$$

That is, out of some sample of 500 up spin electrons, we have measured 250 to have up left spin and 250 to have up right spin. We can use the $\implies$ arrow to show that there is a change after we place the electrons in a magnetic field. With this interlude we can now clearly express what is strange about our above result again– given our original classical assumption– by tracking what information was revealed after each step:

1. Pass the electron beam through a y-magnetic field:

   $$1000e^- \implies 500(\uparrow) + 500(\downarrow)$$

   assuming we start with 1000 electrons and the sample is indeed random.

2. Block out the down-deflected channel:

   $$500e^- \implies 500(\uparrow) + 0(\downarrow)$$

   reflecting that we now have 500 electrons since 500 down spin electrons have all been blocked out from the experiment.

3. Pass the electron beam through an x-magnetic field:

   $$500e^- \implies 500(\uparrow) + 0(\downarrow) = 250(\uparrow\leftarrow) + 250(\uparrow\leftarrow) + 0(\downarrow)$$

   assuming again an equal distribution which is experimentally realistic.

4. Block the left spin channel.

   $$250e^- \implies 0(\uparrow\leftarrow) + 250(\uparrow\rightarrow) + 0(\downarrow)$$

5. Finally, pass the electron beam through a y-magnetic field.

   $$250(\uparrow\rightarrow) \implies 125(\uparrow) + 125(\downarrow)$$

The scalars themselves are really not important at all: one needn't think about numbers of electrons, as they are irrelevant in these scenarios. What is relevant is what happens if we track what happens to the ($\downarrow$) count between these above steps.

1. 500($\downarrow$)
2. 0($\downarrow$)
3. 0($\downarrow$)
4. 0($\downarrow$)
5. 125($\downarrow$)

We have followed one single trajectory! So we should not be counting 500, 0, 0, 0, 125 unless down spin electrons have materialised out of thin air (this is impossible, rest assured). We might therefore guess that some of the 0 counts are false. We can consider this:

- In step 2, we have just blocked out all the down-spin electrons. The 0 count surely therefore must be accurate.
- In step 3, we have just passed the electron beam through an x-magnetic field. We might offer 1 possible guess: that the x-magnetic field actually influences the y-spin and that it is here therefore that the subsequent 0 count was wrong.
- In step 4, we blocked out all the left-spin electrons. If the 0 count here is false, that means that either blocking out left-spin electrons switched the y-spin of some of the electrons, or that there already were down spin electrons by this point before we blocked out the left-spin electrons. If there were down spin electrons before step 4 was implemented then this necessarily means again that after step 3 the 0 count was already wrong.

So everything points to the only reasonable explanation being that the 0 count after step 3 of the x-magnetic field was false: and this would only occur if the x-magnetic field actually flipped the y-spin of some electrons. We first remember that the counter detector has only actually measured the number of left spins and right spins after step 3, since an x-magnetic field

does not make any discernible y-deflections. So it is certainly true that if the x-magnetic field did flip some of the electron y-spins we would not have detected this since it would not have manifested in a deflection.

To test this one lifeline for our assumption, there is a cunning modification to the apparatus. We recall that what we did originally was place successive magnetic fields a relatively large distance away from each other such that different deflected trajectories could be given sufficient distances over which to diverge from each other and therefore be discernibly different according to the measurements of detectors. This makes sense if we simply consider the fact that we are considering magnetic deflections of electrons, which are obviously not significant at all in scale.

On the other hand, we can also well imagine that we can let both deflected paths pass through the same magnetic field, since the angle of deflection is small. If we try this, we continue to add to the list of complications. We can use our step by step denotation to track things again:

1. Pass the electron beam through a y-magnetic field:

$$1000e^- \implies 500(\uparrow) + 500(\downarrow)$$

   assuming we start with 1000 electrons and the sample is indeed random.

2. Block out the down-deflected channel:

$$500e^- \implies 500(\uparrow) + 0(\downarrow)$$

   reflecting that we now have 500 electrons since 500 down spin electrons have all been blocked out from the experiment.

3. Pass the electron beam through an x-magnetic field:

$$500e^- = 500(\uparrow) + 0(\downarrow) \implies 250(\uparrow\leftarrow) + 250(\uparrow\rightarrow)$$

   assuming again an equal distribution which is experimentally realistic.

4. Pass both the up right and up left electrons through a new y magnetic field.

$$500e^- = 250(\uparrow\leftarrow) + 250(\uparrow\rightarrow) \implies 500(\uparrow)$$

So if we compare the results of adding a new y magnetic field to the experiment (step 5 in the previous experiment and step 4 here):

- If we remove one of the right or left channels of x-deflected electrons, then both up and down channels emerge from the y-magnetic field applied to the remaining beam.

- If we keep both of the right and left channels, then only one y deflection emerges from the y-magnetic field applied to the remaining beam.

The above effectively cancels out any idea that the x-magnetic field is flipping the y spin of some of the electrons. We had guessed that in the first experiment that the reason why down spin electrons were appearing seemingly out of nowhere (having been blocked out, we thought, earlier) was because the x-magnetic field was flipping these electrons. If this was the case, then adding left spin electrons would not change anything, since the spin of one electron most definitely does not effect the spin of another. Therefore, if the x field really did flip electrons when the left spin electrons were still absent, adding them would simply mean we still have down spin electrons (perhaps more, since presumably some left spin electrons would be flipped too if right spin electrons are). Yet adding these left spin electrons eliminates rather than augments the number of down spin electrons. Since the spin of one electron does not affect the spin of another electron, we thereby infer our hypothesis that an x-magnetic field can flip a y spin is false.

This now undoubtedly therefore suggests a problem with our original assumption that an electron has pre-determined spin values, within which there can be no explanation for certain empirically observed phenomena. Therefore, the assumption that electrons have a spin value at all times leads to a contradiction with empirical results and must be false.

2.2 Remodelling the Stern Gerlach Experiment

Within the next two chapters from this one, the exact reason for our bizarre results will become quite clear. However, the gist is already apparent: it seems like it is impossible for an electron to have pre-determined

spin values at every point in the experiment: and that they only have predetermined values at some very specific points in the experiment. We need to ensure this could be consistent with experimental data, so let us do this.

One clue comes when we repeat the exact same experiment, but with x, y, x magnetic fields rather than y, x, y fields. The exact same set of phenomena occurs, but simply with the sequence of spin dimensions switched around to reflect the new sequence of magnetic fields. The explanation we now proffer– which is finally the correct explanation– is that we cannot know the x and y spins of the electrons simultaneously. This is obviously not a classical explanation, and is our first experience of a quantum phenomenon in action- that there can be observables which cannot simultaneously both hold values for a given state was certainly shocking to physicists of that time and should be equally shocking to us.

This does however, become consistent with all the phenomena we have been getting in the experiments explained. When we have both left and right channels going into the same magnet, out from the magnet emerges one single channel (in our case, up). This is because, when both channels enter the same magnetic field we can't tell which of them have left and which of them have right spin– and therefore we can tell afterwards what their y spin is. However, if we pass only one x-deflected channel through the magnet, then we can tell what the x spin of the input electrons are by what channel we have chosen, and therefore some are deflected down and some are deflected up– we can no longer tell what their y spin is.

The facts above, however, are presented to imply another question which someone trying to cling onto classical common sense would reasonably ask. It may be true that if we stand above the apparatus and only look at detector readings we cannot tell which electrons have which spin when we are passing all of them into the same magnetic field. However, if we were to focus in on electrons one by one, tracking which electrons have which spin, surely we could for example track the x spins of each of the electrons passing into the y field and then each of the y spins when they come out of it: and thereby know the x and y spins simultaneously?

Investigating this leads us to try and repeat the experiment with as few electrons as possible. The first thing we realise, when doing so, is that 10

electrons and 10^8 electrons exhibit the same types of phenomena when we study them, though naturally the groups of electrons are different in quantity for the two scenarios. In other words– intensity (a measure of the number of electrons per unit area in the beam) does not affect the experiment. Well this is good news, since that means we can reduce to single electrons in our electron gun and once and for all understand the phenomenon!

All seems well. The first y magnetic field, let's say, gives the electron to have up spin. The second x magnetic field, might give a right spin. So we know both of the x and y spins simultaneously, we think: seemingly invalidating our quantum hypothesis that we could not do this. To confirm, we put a new y magnetic field to our up-right spin electron. And finally, we get a nasty shock, because (sooner or later, if we repeat the single electron analysis multiple times) the given electron which we measured to have up and then right spin suddenly exhibits a downwards deflection, showing down spin.

This is the probabilistic state. We have already proved that the magnetic fields cannot flip any spins, and only reveal the spin in the relevant dimension where the magnet is set to. Yet this electron has exhibited up and down spin in successive measurements. This leads us to the only possible conclusion. Our overarching point: that an electron cannot have a spin value at all times, must be true. This would be in line with our other quantum hypothesis that we cannot know simultaneously the x and y spins. The electron *cannot* have a spin value at all times, and this means that there are times it does not have a fixed spin value, and in particular we now know that these times where the spin value cannot be known at least include all the times when the spin value in another dimension is known.

Why do electrons still get deflected, if they do not have spin values? The logical explanation is the correct one this time: an electron might not *have* a spin value, but it can *take* a spin value. It turns out that it takes a spin value with specific probabilities, which we will soon learn how to calculate. So what we achieved in the first two measurements on this single electron: an up and right spin, is simply a result of the electron taking some the up value for y spin with some probability ($\frac{1}{2}$ in our example) and then taking the left value for x spin with some probability (also $\frac{1}{2}$ in our example). However, if we measure y then x, then trying to measure y spin again results in a new probabilistic trial: we *could* get up again, but we could also get down

this time. The y spin measurement will be a completely new outcome when we try to measure it again, and therefore we never did know the two spins simultaneously!

A few things are maintained in this world of chaos, fortunately. We recall that applying the same magnetic field to an electron beam over and over again gives the same spin every time. This would not be true if for each magnet we were simply experiencing a probability of $\frac{1}{2}$ that it exhibits up spin and $\frac{1}{2}$ of down spin, since an infinite sequence of magnetic fields would eventually of course invoke that $\frac{1}{2}$ probability of the other spin to the one we are measuring every time. This therefore proves that if we measure the spin of an electron it *takes* a state- here, with probability $\frac{1}{2}$ of taking an up spin state and $\frac{1}{2}$ of taking a down spin state- and stays in this state until we force it out of that measured state with some type of action. The example we have seen for one type of action which takes an electron out of the original state it has taken, shockingly, is trying to measure its x spin. By measuring its x spin we force it into either a left or right state (here, also with probabilities $\frac{1}{2}$), but also revert it into an unmeasured state with respect to its y spin in the process of measuring its x spin.

This term 'unmeasured state' is a clumsy denomination. Quantum physicists came up instead with the term **superposition** of states– literally, the idea of different possible states being added onto each other to make a new state. Here, the superposition is of two states (up and down in the y spin, or left and right in the x spin). However, if we were considering different observables with more than two possibilities- such as position, which has infinite possibilities- then we could have a superposition of infinite possible states. Of course, some of these states in the superposition might have a greater probability of being measured than others: which is why the idea of superpositions and inherent probability are inextricably linked when we are considering the states of quantum systems. Through our long proof by contradiction, we have thus established that:

- A system may exist in a superposition of states.
- A measurement on a system in a superposition forces it into one of the possible states in that superposition.
- There is some probability associated with forcing that superposition

into one of the constituent states.

- There are certain observables, like x and y spin, which get in each other's ways when one tries to force a system out of the respective superpositions.

By the end of the next three chapters, the reason for these things will all be clear.

2.3 Conclusion

This chapter is actually more of a pedagogical exercise than a strict reportage of phenomena. That is why we kept offering false hypotheses instead of the correct answer from the start. The point is that the reader should get a feeling of classical mechanics and their own intuition starting to fail, and the logical links we must make to try and offer explanations for these strange quantum phenomena. The physical set-up described in this chapter will, for the most part, be swept under the rug, but the ideas it has introduced to the reader will on the other hand be central. The most important two ideas, which the reader must remember, especially for the next chapter, are superposition of states, and inherent probabilities.

These will provide good motivation for the mathematical innovations we see in the next chapter, which will begin our journey of quantum mechanics formally.

Chapter 3

Chapter 3: The State Vector and the Quantum State Problem

In the previous Chapter, we discovered two radical results of the early quantum experiments which completely invalidated the Newtonian model of classical mechanics. These were:

1. The probabilistic state: there exist at the very least some measurements for which we can never predict the result with certainty because the very state itself can take different values with different probabilities.

2. The superposition principle: the reason why probabilistic states exist is because they are in some superposition of different possible physical states to which there correspond separate probabilities. In fact, any superposition of physical states should, in theory, create a new possible physical state.

A third result we obtained might be that there are collections of observables for which a state cannot hold simultaneous values, but this will come as a natural consequence after we set out the quantum mechanical model, as we will see in Chapter 4. All these new principles are offensive to our usual intuition, but undoubtedly are present in the Stern Gerlach experiment, as well is in other famous experiments such as the famous Double Slit experiment, and need to be accounted for in a whole new model of physics. With these ideas laid out– ideas pertaining in particular to the quantum *state problem*:

in this chapter, we will meet the first, and most fundamental, postulate of quantum mechanics– which I call the State Postulate– and our study of the model has begun.

3.1 The First Postulate: Necessities

3.1.1 Complications of the Quantum State

The classical model of Newtonian mechanics deals so trivially with the state problem that to those only classically trained the state problem does not even seem like much of a problem at all– much less one of the two most important problems in physics. One measures the momentum of a ball, the mass of the ball, the radius of the ball and angle it is travelling. This completes the classical state problem– at least from the point of view of its basic trajectory: we could also measure its electric charge or angular momentum if we wanted to approach the state problem from a different angle. The measured momentum, mass, radius, angle are the characteristics of the state: which we call values of its observables. We can use equations like $p = mv$ to find further quantities like the velocity when we so desire– or, just find a way to measure them as well. After this, the problem shifts to calculating collisions or terminal velocities or whatever the situation calls for: and all of these problems are of course time evolution questions because we are to predict the future state given events which happen over time (eg, the ball hits another ball, swings on a string or falls down a ramp). With quantum phenomena, however, a certain measurement might tell you everything about that state, if that state happened to have probability 1 of having the given measurement result, but it might also tell you essentially nothing, such as for some state where the obtained measurement actually was initially improbable and the state just happened upon measurement to take an unlikely value out of countless other possible values. The Newtonian procedure of dealing with the state problem – which is to assume a state always has some value for each observable and what only remains is to measure and record it at specific times of interest– is far too simplistic to account for the phenomena of the quantum world.

So the main complication to the state problem, of course, comes with the notion of inherent probability in a given state. This notion would have been

inconceivable in Newtonian mechanics. Probability could only ever be introduced, into experiments by human made contrivances (such as probabilistic machinery) and the current state would remain unchanged until a probabilistic action changed it into a future state. However, even in these cases, these probabilistic differences are not actually pertaining to the inherent state problem, but rather problems of external interference. As far as the state problem is concerned, having a probabilistic state destroys the classical approach. For one, due to the superposition principle, very few states in quantum mechanics can be said to possess a value for an observable: we can no longer measure a particle's momentum and say that is the relevant momentum value of the state if that value is only one of an infinite collection of possible momenta we could have originally measured. We saw this in the Stern Gerlach experiment, where attempting to say a single electron had up spin failed miserably when we realised successive measurements (with a x magnetic field in the middle) could yield down spin measurements as well. The problem of encapsulation is therefore rendered by the results much more difficult than with only classical intuition: because we now have the extra complication of needing to find a concise way to not only store information on far more possibilities for the same state, but also their respective probabilities. The question of being able to store information on infinite (so by definition, impossible to list) possibilities is another concern in itself, and it should be clear to the reader that the classical approach will be insufficient.

Fortunately, contemporary and historic developments and concepts in mathematics proved very useful when trying to deal with these new modelling complications. Probability and how to handle it itself was certainly not unexplored by mathematicians, and the burgeoning field of vector spaces would prove to be useful in answering the superposition principle. With an intellectual jump whose magic we cannot underestimate, Heisenberg, Born and almost simultaneously Dirac realised that they could incorporate these areas of mathematics into a new quantum model of reality. Thus were born the postulates of quantum mechanics, and with them, this seminal new physical model.

3.1.2 One-to-One Correspondences

There was a detail in the previous chapter on Stern Gerlach experiments, which, though commanding little attention, pertains to a key concept which

will be central to this chapter. That was the idea of "state denomination". If we recall, we invented a strange notation– while futilely attempting to cling onto normal classical intuition– in our early attempt to track the spins of different particles. There are many things wrong with the attempted equation:

$$500e^- = 250(\uparrow) + 250(\downarrow)$$

–not least the fact that it is useless in taking into account superpositions and inherent probability. One asks therefore why it was included? This was to introduce a much more useful idea: we will need to start introducing abstract entities to represent *states* themselves. In classical mechanics, we do not do this at all, because the state is treated like a simple background from which to pick attributes rather than anything worth studying in itself. We do not need to perform addition (the mathematical representation of a superposition) on classical states, or calculate inherent probabilities. If we really need to distinguish between them, we treat them like situations or mathematical cases, where symbols are given to observable qualities like velocity v. We might use subscripts like v_a to help us identify which case we are working with– eg, the velocity of ball a. We would never mathematically represent the state itself, however, when we could simply list its physical properties and put those values into equations.

On the other hand, because of these new quantum ideas of summing states together (superposition) and including measurement probabilities in the state, it will be extremely important to have *mathematical entities representing quantum states themselves.*

We will need to consider how to mathematically operate with states so that we will be able to perform physical computations. The first step, as we tackle how to do this, is to learn about vector spaces. Yet before all that, we will give a mathematical definition for the idea of 'labelling' something. It may seem strange to do so, but having these clarifications will be immensely useful because very quickly from now we will need to understand this idea to navigate through labels of labels of labels of objects and being able to track how these things exactly are mathematically connected to each other is very useful.

We have already started using the word 'abstractions' in this text. When

we speak of an abstraction, we mean something which is not physical– in other words, a mathematical entity we indirectly use to describe something physical. This goes beyond labelling with symbols, and can be extended to the use of real mathematically manipulable objects to connect to states as well. The way this is done is through one-to-one correspondences.

A one-to-one correspondence strictly occurs between two sets. It is defined to be the relationship between two sets where for every element in one set, there is one element in the other set, and for every element in the other set, there is an element in the set. Intuitively, the idea is not hard to imagine. Each element can be mapped to a single element in another set, and vice versa, without any elements left out. The conventional way to prove there is a one-to-one correspondence between two sets is to introduce a one-to-one mapping – also known as a bijection – whose existence can be proved if we can prove that for every element in the first set there is a corresponding element in the second, and for every element in the second set there is an element in the first. Take for example

$$A = \{2, 3, 5, 7\}$$

to be the single digit primes. We could set up a one to one mapping with the set

$$B = \{9, 27, 243, 2187\}$$

by using the rule "take 3 to the power of the element" to map an element of A to an element of B and the rule "take the element log base 3" to map an element of B to an element of A. Then there is a one-to-one correspondence between sets A and B since there are two maps, one A to B and one B to A which set up a pairing between elements in A and B such that each pair is unique and no element in either set in left out.

Now we can talk about the uses of one-to-one correspondences, which will be a common point in this book, as quantum mechanics enjoys very many one-to-one correspondences. The first use we will see is in label sets. The idea is simple. If there is a one-to-one correspondence between two sets, we can use elements in one set to label elements in the other set. We know that every label will have something it represents (since each element in the set of 'labels' is mapped to another element in the other set); no label will represent more than one element and cause confusion (since it is a one-to-one

mapping); no element will be unlabelled (since for every element in the other set there is a corresponding element in the set of 'labels'); no element will be labelled by two different labels from the same label set and cause confusion (since it is a one-to-one mapping). Another even more important use we will see is as substitutes: if a one-to-one correspondence exists between two sets, we can substitute the elements of one set for the elements of the other set, since again the correspondence makes it very easy to identify which substitutes are connected to which originals we are trying to describe.

With the two example sets of single digit primes and the results of raising them to the power of 3 above, it makes absolutely no sense to use sets B or A as either labels or substitutes for each other, since they are both sets of numbers, and we do not need to label numbers with other different numbers, and it is equally difficult to see why we might want to substitute numbers for other numbers in any scenario. However, there will be uses later on for one-to-one correspondences for both labelling and substituting for different objects in quantum mechanics. Specifically, one important one-to-one correspondence will be between functions and states. This is because, if we define a certain set of functions properly and have a one-to-one correspondence with physical states, we will be able to extract information and values about that state with certain operations on that function representing it. Clearly, we cannot perform mathematical operations on physical states directly, which are not mathematical objects at all. Nor do we want to describe states in words all the time, as this would be beyond verbose and time-consuming. Thus these mathematical objects both labelling and substituting for physical states is a complete necessity in our new world away from the classical method of ignoring the state problem, and towards our new, quantum realisation that states are very much important and deep objects in themselves.

Thus we understand the goal of this chapter, and the first postulate of quantum mechanics. We cannot perform mathematical operations on a state, but we will need to find some way to achieve this in order to navigate the complex superpositions and probabilities which never existed in the classical model. Thus we want to set up some mathematical entities which can be operated on and which are in one-to-one correspondence with physical states! After that, we can perform mathematical operations on those entities, and interpret the values which result as information about the states they represent. The importance of the substitution and labelling functions of one-to-one corre-

spondences belie nearly all of the quantum mechanical postulates, so having read this section, the reader should now be prepared to handle these ideas without any more confusion. So now, we are ready go through this process of searching for mathematical objects to represent states by one-to-one correspondences in the following section, on vector spaces. One final note before we do so: one-to-one correspondences will in the future be referred to by their mathematical name: **bijections**.

3.2 Vector Spaces

The reader will be vaguely familiar already with the notion of vectors, which may have been mentioned in physics and coordinate geometry. However, they probably will have usually thought of them as objects with magnitude (length) and direction; in particular, geometrically one might have represented vectors with arrows and extrapolated this idea to include the basic operations of vector addition and scalar multiplication of vectors (Figure 1).

This concept of all vectors as geometric arrows is not very helpful here, and must be for the purpose of starting to learn quantum mechanics completely discarded. Instead, we must think of them simply as **objects which can be summed**, which are usually parts of collections, called "vector spaces", with regular structures. Numbers can be summed, but vectors can be objects which are not numbers but which still can to be added to each other; for the purpose of quantum mechanics, having this full space of "things which can be added", but which are not just simply numbers, will prove invaluable. This is because we we need mathematical objects to form bijections (one-to-one correspondences) with physical states: but by the superposition principle any given group of states must be able to be superposed and result some new physical state, and therefore the mathematical objects representing them must be able to be summed together to make a new such mathematical object which represents this new physical superposition state. So it is absolutely essential to bear in mind that functions like x^2 or $(5x-3)^3$ can be vectors just as much as arrows in a coordinate system can be vectors: so long as there is a system where they can be added together and a collection of other objects which fit together into the system. We call this system of objects able to be summed together a vector space, and those objects con-

stituent vectors. To support this new idea of vectors, we will therefore be replacing the misleading arrow notation of $\vec{V}$ with simply V. At first, vectors will feel extremely intangible given that they are only represented by letters giving no indication what they are, but this is fine. Learning about the mathematical system which contains the vectors which will represent physical states is what matters, for now, because there are many operations and ideas to work through.

For a self-contained system working through the lens of linear algebra there exists a need for a vector space: that is, the space containing all relevant vectors where certain mathematical operations can be run without ever requiring any vectors outside the vector space. Similarly, for quantum mechanics there also exists a vector space for specific vectors of interest, (which is called state space). If we can understand this vector space we can then understand the component vectors a lot better: in other words, our formalism of quantum mechanics becomes defined at its farthest boundaries. Therefore, we will start by introducing the definitions mathematicians use to define a valid vector space so we can start to move towards this goal.

A vector space $\mathbb{V}$ is a set of vectors with the following properties:

1. *Null Vector*: Every vector space contains the unique null vector 0, with properties
$$\forall \boldsymbol{\alpha} \in \mathbb{V}, \quad \boldsymbol{\alpha} + 0 = \boldsymbol{\alpha}$$
and
$$\forall \boldsymbol{\alpha} \in \mathbb{V}, \quad 0 \times \boldsymbol{\alpha} = 0.$$
Note that this is not a number, but a vector (though it could be a number if our vector spaces consisted of numbers as constituent vectors). Similarly, the identity vector I (sometimes, 1) is defined to be the vector such that
$$\forall \boldsymbol{\alpha} \in \mathbb{V}, \quad 1 \times \boldsymbol{\alpha} = \boldsymbol{\alpha}.$$
Indeed, the labels 0 and 1 are reasonable since they are very clearly the vector analogs of the numbers 0 and 1.

2. *Additive closure*: For a defined rule for producing a sum of two vectors,
$$\forall \boldsymbol{\alpha}, \beta \in \mathbb{V}, \quad \boldsymbol{\alpha} + \beta \in \mathbb{V}.$$

The fact that the sum of any of the vectors in the vector space is another vector in the vector space means the set is **complete**, and will be a very important point for our requirements, owing to the superposition principle.

3. *Commutative property of addition*: For a defined rule for producing a sum of two vectors,
$$\boldsymbol{\alpha} + \beta = \beta + \boldsymbol{\alpha}.$$

4. *Associative property of addition*: For a defined rule for producing a sum of two vectors,
$$\boldsymbol{\alpha} + (\beta + \gamma) = (\boldsymbol{\alpha} + \beta) + \gamma.$$

5. *Additive inverses*: For every vector $\boldsymbol{\alpha}$, there exists a unique additive inverse vector $-\boldsymbol{\alpha}$ such that
$$\boldsymbol{\alpha} + -\boldsymbol{\alpha} = 0.$$

6. *Distributive property of scalar multiplication*: For a defined rule for vector multiplication by scalars,
$$c(\boldsymbol{\alpha} + \beta) = c\boldsymbol{\alpha} + c\beta.$$
is the requisite distributive property for vectors of scalar multiplication of vectors. There is also a distributive property for scalars in scalar multiplication of vectors, which is similar:
$$(c_1 + c_2)\boldsymbol{\alpha} = c_1\boldsymbol{\alpha} + c_2\boldsymbol{\alpha}.$$

7. *Associative property of scalar multiplication*: For a defined rule for vector multiplication by scalars,
$$c_1(c_2\boldsymbol{\alpha}) = c_1c_2\boldsymbol{\alpha}.$$

Most of these properties are hardly unusual to us and do not require much thought, such as the associative properties, because it is unlikely the reader will have worked with mathematical objects which do not follow them before (note the arithmetic numbers certainly do). The closure of addition is the critical point, however, because it allows us to get a sense of why vectors are

a useful way to denote objects which one might want to add together: they can fit into a vector space which then provides an mathematical enclosure so one knows that no matter how many different constituent vectors they add together, they will never "break" the vector space and end up with a new type of vector outside of the vector space. This enclosed space of objects will be useful for quantum mechanics, as we want to be able to superpose *any* number of states without creating a non-physical state– and therefore, we want to be able to add the mathematical objects which represent physical states without creating a mathematical object which is not in that set of mathematical objects representing physical states.

There exists one final property of vector spaces which make them crucial to quantum mechanics. It pertains to the problem of defining a vector space. One can imagine that in order for additive closure to be a property of a vector space one needs a huge number of vectors to constitute most vector spaces, since we need every possible sum between two or three or any number of vectors in the space. Thus trying to list constituent vectors as a set in order to specify a vector space would be useless and impossible. We need a shorter way to specify a vector space we are working with. The elegant solution to this problem, of accounting for any possible combination of different vectors, is through the concept of a **basis**.

3.2.1 Dimensions of a Vector Space

We have expressed the definition of a vector space through comprehensive discussion, but are no closer yet to having a concrete understanding of what examples might actually be. We will now consider the most common vector space we work with: the Cartesian 2-dimensional plane (figure 2).

There is no need to further explain the coordinate system– only that clearly it does satisfy all the rules above for a vector space (including that of additive inverses, when we include the negative y and x axis). This vector space is known as the coordinate representation of $\mathbb{R}^2$: the 2-dimensional (hence the exponent 2) real-valued (taking real values only) vector space: where the constituent vectors are numbers!

How do we define the dimensions here? Qualitatively, we are familiar with the concept of the x-axis and y-axis, but mathematically it is less easy to say

why they qualify as dimensions. The answer lies in the definition of **linear independence**, whence this idea of a basis we need will follow .

A set of n linearly independent vectors $\boldsymbol{\alpha}_i$ is defined mathematically as follows:

$$\sum_{i=1}^{n} c_i \boldsymbol{\alpha}_i = 0 \Rightarrow \quad \forall i, \quad c_i = 0.$$

In words, there is no nontrivial combination of linearly independent vectors which equals 0 when summed together– a trivial combination would occur where all the multiplicative coefficients are 0. If a linear combination of linearly independent vectors sums to zero then it implies that the coefficients in the combination must be zero. If there exists some combination without all the coefficients equalling 0, then the vectors are not linearly independent.

Let's consider a few examples, using 2×2 square matrices as vectors and the 2×2 null matrix as the 0 vector. If

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{\alpha}_2 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{\alpha}_3 = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}, \ \boldsymbol{\alpha}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

and we set some combination with coefficients $\{c_i\}$

$$\sum_{i=0}^{4} c_i \boldsymbol{\alpha}_i = 0,$$

then

$$c_1\boldsymbol{\alpha}_1 = c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ c_2\boldsymbol{\alpha}_2 = c_2 \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \ c_3\boldsymbol{\alpha}_3 = c_3 \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix},$$
$$c_4\boldsymbol{\alpha}_4 = c_4 \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$
$$\Rightarrow \ \sum_{i=0}^{4} c_i \boldsymbol{\alpha}_i = \begin{bmatrix} c_1 & 2c_2 \\ 3c_3 & 4c_4 \end{bmatrix} \Rightarrow \ \begin{bmatrix} c_1 & 2c_2 \\ 3c_3 & 4c_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so we can see that we necessarily must have

$$c_1 = c_2 = c_3 = c_4 = 0$$

and therefore

$$\boldsymbol{\alpha}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{\alpha}_2 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{\alpha}_3 = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}, \ \boldsymbol{\alpha}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

are linearly independent vectors.

On the contrary, if

$$\beta_1 = \begin{bmatrix} -1 & 0 \\ 5 & 10 \end{bmatrix}, \ \beta_2 = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}, \ \beta_3 = \begin{bmatrix} 2 & 7 \\ 4 & 1 \end{bmatrix}$$

Then $c_1 = 1, \ c_2 = 7, \ c_3 = -3$ gives

$$\begin{aligned} c_1\beta_1 + c_2\beta_2 + c_3\beta_3 &= 1\begin{bmatrix} -1 & 0 \\ 5 & 10 \end{bmatrix} + 7\begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix} - 3\begin{bmatrix} 2 & 7 \\ 4 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 5 & 10 \end{bmatrix} + \begin{bmatrix} 7 & 21 \\ 7 & -7 \end{bmatrix} + \begin{bmatrix} -6 & -21 \\ -12 & -3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0 \end{aligned}$$

which is a nontrivial combination of the vectors β_1, β_2, β_3 as the coefficients are not all 0. Therefore,

$$\beta_1 = \begin{bmatrix} -1 & 0 \\ 5 & 10 \end{bmatrix}, \ \beta_2 = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}, \ \beta_3 = \begin{bmatrix} 2 & 7 \\ 4 & 1 \end{bmatrix}$$

are not linearly independent vectors (we say therefore they are linearly dependent) since there exists at least one nontrivial linear combination of them which sums to make the 0 vector.

Now that we have established this definition of linear dependence, defining the dimensions of a vector space is simple. We have 3 crucial definitions and theorems:

1. An n-dimensional vector space contains n linearly independent vectors.
2. The set of n linearly independent vectors in the n-dimensional vector space is called the **basis** of the vector space.
3. Each vector v in an n-dimensional vector space with basis $\mathbb{B} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ...\boldsymbol{\alpha}_n\}$ can be expressed as

$$v = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i,$$

a unique linear combination of the linearly independent basis vectors. The coefficients c_i of the expansion for v are called the **components** of the vector v in the basis.

Definition 3 contains two assertions. The first statement is that each vector v in an n-dimensional vector space with basis $\mathbb{B} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ...\boldsymbol{\alpha}_n\}$ can be expressed as

$$v = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i$$

for some scalar coefficients $\{c_i\}$. This can be proved by contradiction. Assume that in the n-dimensional vector space with basis $\mathbb{B} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ...\boldsymbol{\alpha}_n\}$ there exists some vector v which is not a linear combination of the basis vectors. Clearly $\forall j, \quad v \neq \boldsymbol{\alpha}_j \in \mathbb{B}$ since otherwise the linear combination

$$1 \times \boldsymbol{\alpha}_j(x) + \sum_{i=0 \neq j}^{n} 0 \times \boldsymbol{\alpha}_i = v$$

would immediately violate the assumptions about v not being able to be expressed as a linear combination of the basis vectors. But then this would mean that all the coefficients c_i are 0 since any coefficient being nonzero would create a valid linear combination of basis vectors equalling the vector v, and so

$$v = \sum_{i=0}^{n} 0 \times \boldsymbol{\alpha}_i + v$$

but then:

$$\sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i + c_{n+1} v = 0 \Rightarrow \quad c_{n+1} v = -\sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i,$$

which is,

$$v = -\sum_{i=0}^{n} \left(\frac{c_i}{c_{n+1}} \right) \boldsymbol{\alpha}_i.$$

This then violates the assumption that v cannot be expressed as a linear combination of the basis vectors since the coefficients c_i/c_{n+1} corresponding to basis vectors $\boldsymbol{\alpha}_i$ would give a valid linear combination producing v. The only way this would not be true would be if $c_{n+1} = 0$. In other words, if we want our assumption to hold true, we get

$$\sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i + c_{n+1} v = 0 \Rightarrow \quad c_{n+1} = 0 \Rightarrow \quad \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i = 0 \Rightarrow \quad \forall \ i, c_i = 0.$$

By definition, therefore, in order for the assumption to be true v must be a vector linearly independent from the other basis vectors since there is no non-trivial combination of the linearly independent basis vectors and the new vector v which sums to the null vector. This however means there is a contradiction since we defined v to be a vector in the n-dimensional space, but adding the new vector v would mean there are $n+1$ linearly independent vectors in an n-dimensional space: defined in Definition 1 to be impossible. Therefore any vector v which is not a linear combination of the basis vectors cannot exist in the vector space; all the vectors in the vector space must be able to be expressed by a linear combination of the basis vectors. We say that the basis (which we must remember is the *set* of basis vectors and so is be treated as a singular entity in reference) **spans** the vector space, because all vectors in the vector space can be created by some linear combination of the constituent basis vectors of that basis.

The second simple but important theorem to prove is that the expansion for any given vector v

$$v = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i$$

is unique. We again prove this by contradiction. Assume

$$v = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i = \sum_{i=0}^{n} c'_i \boldsymbol{\alpha}_i$$

for $\{c_i\} \neq \{c'_i\}$. Then, for basis vectors $\boldsymbol{\alpha}_i \neq 0$,

$$v - v = 0 = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i - \sum_{i=0}^{n} c'_i \boldsymbol{\alpha}_i = \sum_{i=0}^{n} (c_i - c'_i) \boldsymbol{\alpha}_i.$$

However, this combination of basis vectors is equal to 0, but the basis vectors $\{\boldsymbol{\alpha}_i\}$ are linearly independent so necessarily this implies the new coefficients $\{c_i - c'_i\}$ are 0. So

$$\forall i, \;\; (c_i - c'_i) = 0 \Rightarrow \;\; c_i = c'_i$$

a clear contradiction with the original assumption that the coefficients could be different and thereby produce more than one unique way of expressing a vector in the basis. So indeed, each vector in a vector space is uniquely specified by one single linear combination of the basis vectors of the vector space.

New definitions follow:

1. The basis is the set of vectors spanning the vector space.
2. For each constituent vector in the vector space with basis $\{\boldsymbol{\alpha}_i\}$ defined to be $v := \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i$, we call the coefficients $\{c_i\}$ the **components** of the vector in that basis.
3. The sum which uniquely defines a vector of the vector space in a given basis is called its **expansion** in that basis.

Critically, note that we always talk about components *in a basis* and expansions *in a basis*. Without some underlying basis, none of those concepts make sense. In vector spaces with multiple possible bases– which we shall see very much exist– expansions and thereby components for the same vector are very different depending on which of the bases we are working in.

It is now clear why a basis is powerful for its original role– to describe a vector space. This is because with simply the basis vectors (which can be chosen quite easily for some vector spaces) and the components of *any* constituent vector in the space spanned by that basis, we can uniquely specify that vector. Then, with this knowledge of unique expansions in mind, we can define the rules for the simple operations addition, subtraction and multiplication in a vector space:

To multiply a vector v by scalar k, we multiply its components each by scalar k:

$$kv = k\sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i = \sum_{i=0}^{n} k c_i \boldsymbol{\alpha}_i.$$

To sum two vectors $v = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i$ and $w = \sum_{i=0}^{n} c'_i \boldsymbol{\alpha}_i$ we sum their components:

$$v + w = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i + \sum_{i=0}^{n} c'_i \boldsymbol{\alpha}_i = \sum_{i=0}^{n} (c_i + c'_i)\boldsymbol{\alpha}_i,$$

which is a new expansion uniquely specifying a new vector in the vector space, and for subtraction of w from v we subtract the components of w from the corresponding components of v:

$$v - w = \sum_{i=0}^{n} c_i \boldsymbol{\alpha}_i - \sum_{i=0}^{n} c'_i \boldsymbol{\alpha}_i = \sum_{i=0}^{n} (c_i - c'_i)\boldsymbol{\alpha}_i$$

which is also a new expansion uniquely specifying a new vector in the vector space. The additive inverse is obtained by multiplying a vector by the negative identity:

$$w = \sum_{i=0}^{n} c_i' \boldsymbol{\alpha}_i \implies -w = \sum_{i=0}^{n} -c_i' \boldsymbol{\alpha}_i,$$

multiplying by the null vector means multiplying all the coefficients by 0 (creating the null vector, as expected), and multiplying by the identity means doing nothing to the components, resulting in the same vector.

3.3 The State vector

We are finally ready to introduce the mathematical objects chosen to be in bijection with physical states, having hinted heavily that with additive closure vectors in some vector spaces are the answer. This is the First Postulate of Quantum Mechanics.

Postulate 1: The State Vector and its Wavefunctions

Any physical states at time t can be represented by a state vector Ψ_t. State vectors are complex-valued vectors which stand in a bijection with vectors in a Hilbert space, which we call the state space; state vectors can be transformed into unique probability distribution functions, called wavefunctions, to give probabilities of different measurements of different observables occurring.

So there we have it: the mathematical objects we have been alluding to, which are in a bijection with physical states and can therefore be used as substitutes for them in order to perform physical computations. There is a lot to unpack with this postulate, which is central to quantum mechanics and the state problem, so we will now do this systematically.

The state vector is a vector in the vector space we call the state space. While we will cover the state space subsequently, it is absolutely crucial at this point to remember the previous section on vector spaces, which showed

us that all we need chiefly is additive closure and a set of objects can be considered vectors in a vector space. The additive closure property is critical for superpositions, which as previously explained, make this necessary because we need to be able to add any number of states without making an unphysical state, and therefore also need to be able to add any number of state vectors without creating a vector outside of the state space.

The state vector is also a powerful mathematical tool because it can be broken up into basis vectors; an infinite dimensional space like the state space has infinite bases of infinite cardinality. The importance of this pertains to the second part of the postulate. The second part tells us that from the state vector we can generate different unique probability distribution functions, called wavefunctions, to give probabilities of different measurements *for different observables*: the ability to express the state vector in different bases will later be shown to be essential for considering the state represented by the state vector with respect to the different observables we are concerned with. These wavefunctions are the part of the Postulate which are most tangible to us, because we will directly be able to achieve numerical values from them.

We continue now by dealing with this postulate and the rest of its assertions with a bit more detail. This book will not have such a verbose discourse on any other postulate of quantum mechanics than that which is to follow, and there is a serious reason for making so many clarifications prior to starting to consider the mathematics of state vectors– everything else is predicated on this postulate!

1. The functionality of the state vector is very important to understand fully. We recall that in our section on bijections there were two important employments of this relationship between two sets: the first, for labelling, and the second, for substituting. The labelling function of the state vector is very clear. Normally a capital Psi, Ψ, is used, and we can simply say "the state Ψ", just like we can say "ball a". We know that every Ψ is unique and in bijection to each state, so it acts well as a unique label.

   As a substitute for physical states, we have said that the main goal of this chapter is to produce something mathematically manipulable since physical states are not. The main operation we will perform

on it– called the inner product– which will be shown imminently, will be useful for multiple purposes: but the most useful will be to transform the state vector into the probability distribution functions called wavefunctions. We recall from the preliminary on probability that a probability distribution function is a function we can use to encapsulate different possibilities and their probabilities in a single function, so the operation which produces a probability distribution function from the state vector representing a state will be essentially the final step of considering that state with respect to certain observables. Achieving these wavefunctions is normally the final goal of solving a problem in quantum mechanics.

2. Time is an important factor, of course, to the state problem, but it represented here by a subscript rather than a variable in Ψ_t. The subscript is also helpful to remember that each state vector represents a physical state at an instantaneous time t (since any state only exists at a single time before it transforms into a new state). Mainly because we are still covering the state problem, it is unwise to consider time too much while there is so much to learn with regards to the encapsulation and extraction of information; therefore, all of the discussion on quantum states we will have takes place at one instance of time.

   However, there is more commonly in quantum mechanics the notation

   $$\Psi(t),$$

   representing the state vector as a function of time. This doesn't mean the state vector is something like

   $$\Psi(t) = t^2,$$

   but rather that

   $$\Psi(t) = \Psi_t.$$

   In other words, the function notation is just a shorthand of referring to the same isolated system across different moments in time, where inputting a time value gives the state vector at that time. This is the same idea that we are using, but just more concise, so there is no need to get confused if we see this written elsewhere. Again, we are dealing with stationary states, so the notation Ψ_t is valid, intuitive and sufficient especially in these stationary cases.

3. A mathematical vector and a physical circumstance is not the same thing, naturally. Nevertheless, we often call a state vector "a state". The reason is because of the bijection. In those cases, we should say "one represents the other". However, we will still sometimes end up saying "one *is* the other" when this bijection exists– even though they might technically be different types of objects which therefore cannot *be* each other. Such grammatical simplifications are simply because having the bijection eliminates ambiguity. For a physical state, we cannot really say "the hydrogen electron in a magnetic field with one third probability of having momentum p_1, one seven-hundred and eighth probability of having momentum p_2..." since this would be a manifestly inefficient method of specifying it. Thus, a mathematical representation like the state vector encapsulates all these details of a state concisely while being linked to that state with no ambiguity; we therefore call it the state.

   This seems like a terribly semantically pedantic discussion, but there will be times where readers get confused by the difference between A is B and A is in a bijection with B. With this clarification it is hoped such confusions are eliminated in the mind of the reader. So long as one follows the mathematics, the answer to the difference discussed above will always be clear, and this needn't ever evolve into a massive obstacle in understanding. When two things are deemed **equivalent**, the reader should simply consider closely how they may be related to each other, rather than jump to conclusions that they have somehow misunderstood different types of objects. Remember the failed state denomination during the Stern Gerlach experiment: we should not get attached to trying to make mathematical symbols sacred, and should only concern ourselves with the functionalities they provide. Therefore, the phrase 'the state Ψ_1' would be perfectly acceptable.

4. This book will use the letter ψ ubiquitously both in its upper case form (Ψ) and its lower case form (ψ). When a reader sees the capital letter Ψ, we are referring to a state vector. The capital Ψ will never be used to denote anything which is not a state.

5. Though I have said the state space is infinite dimensional, we will be working under the assumption that everything we are doing takes place

> in a discrete space until chapter 7. This doesn't violate the conditions of the postulate, though explaining why exactly this is takes us to discussion of infinities (cf. Chapter 6) which is wholly unnecessary. We just have to accept the assumption that we are working in discrete cases, and it will be enough for us to build up our understanding of what we actually need to understand at this moment. The fact we are working in discrete cases, for example, means that we can use our sigma summation notation very comfortably whilst postponing integrals to another chapter where will be much better prepared and disposed to deal with them.

Now that these clarifications about state vectors have been made, we can move onto studying the state space, as well as the all-important inner product operation, in more detail.

3.3.1 The State Space

The vector space relevant to quantum mechanics is a Hilbert space $\mathscr{H}$, which contains all the state vectors corresponding to possible states. It turns out that we do not have to worry about the name of the space too much: even though Hilbert spaces are special types of vector spaces very much constituting their own mathematical field with much deep analysis and theory, these analyses are matters of mathematical rigor and for our purposes we have neither the luxury nor the need to be so precise. In general, therefore, we will call it, by convention, the state space– since it is the vector space of state vectors– to avoid the exotic 'Hilbert space' name getting in our way and drawing more attention than it needs to.

There is an important idea to understand about the state space and therefore the state vector, however. The human perspective is to try and imagine the vector space and its components in some sort of physical form. It is expected that, despite warnings not to, any reader who has seen the arrow representation of 2 dimensional vectors will be thinking of something similar every time they think about vectors– or perhaps even envisaging other melodramatic depictions of quantum mechanics in popular media.

However, we need to be extremely careful here. Vectors are abstract entities which cannot be imagined physically *until we choose a basis*; this point

must be hugely emphasised. In different bases, the vector takes different forms: based on its unique expansions in those bases! Without picking a basis, the state vector very much exists: it just simply isn't given an explicit form yet. It is usually impossible to predict which form a vector will take in different bases until we know what the bases are and perform some operations, because there is no set pattern each vector follows. To fully deliver this idea of a state vector being able to exist without an explicit mathematical form, a vernacular analogy will be given– to consolidate an understanding for the reader at the cost of temporarily violating the rigorous nature of the text.

Take a vector we probably have seen before in common use: the velocity vector of a moving car. If one asked whether it exists, then obviously the answer would be the affirmative: a moving car (or even a stationary one) must have a velocity vector. However, if someone were to point to a moving car and ask you to write down its velocity vector on a piece of paper, then you would find it impossible. This is because it can take a wide variety of forms depending on how we define our coordinate system– we can use an axis where moving to the left of us is negative, but we can also use one where moving to the right of us is negative (changing the values of the velocity vector we are employing); we can consider it with respect to another moving car or our sedentary selves, but we can also consider it with respect to any fixed point as an origin of movement; in any case we could pick any moving car or any origin of movement and this would give us different *forms* of the velocity vector: because the numbers are different depending on what we are considering the car's motion with respect to. However– no matter what coordinate system we choose, the vector exists. We do not need an origin of displacement for a velocity vector to exist– we only need an origin of displacement for the velocity vector to take a form! **It should be clear that to exist as a mathematical object which can take multiple forms depending on the underlying frame of reference we used is different from taking one specific form**. For the state vector, we have the very same idea: we choose a basis, analogous to defining an axes in the above example, and we specify its components, analogous to the familiar process of specifying coordinates. Both the basis and the components are necessary for us to 'pinpoint' the state vector, but it exists to be expressed in infinite different bases and sets of components regardless of whether or not we give it a certain algebraic specification.

The example of the car velocity vector should enforce a crucial idea of abstract mathematical objects existing but being impossible to express explicitly. In the exact same way, a state vector always exists no matter how we choose to see it or portray it, and to give it a single specific form, we need to pick a basis, which is equivalent to picking a coordinate system in the example above. It does not have a form until we pick a basis. So if we ever see the statement that the state vector is

$$\Psi(x) = cx^2$$

(for example), then we should immediately realise that this statement implies some underlying basis, because there must be some basis in order for it to have this explicit algebraic form. Otherwise, there would be no way we could express the state vector as a function of some variable in this way. Any mathematical equation we run with state vectors will always be implicitly taking place in some basis; that does not at all mean things would look that way in all bases, but one can assume that the basis picked will be the most convenient one– just like we would not pick the Tokyo Tower to be the origin of displacement for the velocity vector of a car in London, because that form would be inconvenient and uninformative. Choosing bases intelligently, we will see, is already in itself an important problem solving tool in quantum mechanics.

Next, beyond simple operations like scalar multiplication, quantum mechanics employs another operation, which will be extremely central to all of quantum mechanics. This is the inner product, which will be as common in quantum mechanics as multiplication in arithmetic.

3.3.2 Inner Products

Suppose we are working in a basis. There then exists a new operation, called the **inner product**, between two state vectors Ψ_1 and Ψ_2 in the state space. This inner product is denoted as (Ψ_1, Ψ_2) and is defined to be:

$$(\Psi_1, \Psi_2) := \sum_{\{x\}} \bar{\psi}_1^*(x)\psi_2 \equiv \sum_{\{i\}} c_i^{(1)*} c_i^{(2)}$$

for components $\{c_i^{(1)}\}$ for Ψ_1 and components $\{c_i^{(2)}\}$ for Ψ_2 .

We will see more often

$$(\Psi_1, \Psi_2) := \sum_{\{i\}} c_i^{(1)*} c_i^{(2)}$$

because there is no need to involve the discrete wavefunction in things when components are easy to track in the discrete case. This form is also very illustrative: it shows us that all the inner product is doing is producing a sum of the products of matching components, with one its complex conjugate. Many readers of this book will be able to understand what I refer to when I say this is essentially the quantum mechanical (really, simply complex valued) equivalent of a vector dot product. However, for less advanced readers who have not met the latter before, this point is not important.

Next, most of the time the inner product is non-commutative- the order matters. In fact, since we have the definition above, it is very easy to see that what we must have is the relationship

$$(\Psi_1, \Psi_2) = (\Psi_2, \Psi_1)^* .$$

regardless if we try to see this in the discrete case or continuous case. This is an essential short-form fact to memorise, as it will return in algebraic manipulations. Now we list a few more facts about the inner product:

IP1. There is a kind of "constant multiple rule" which comes with the inner product. We will always use the short-form of it, but expressing the inner product in sum form makes everything completely clear:

$$(\Psi_1, c\Psi_2) = \sum_{i=1}^{n} c_i^{(1)*} c c_i^{(2)} = c \sum_{i=1}^{n} c_i^{(1)*} c_i^{(2)} = c\,(\Psi_1, \Psi_2)$$

and

$$(c\Psi_1, \Psi_2) = \sum_{i=1}^{n} c^* c_i^{(1)*} c_i^{(2)} = c^* \sum_{i=1}^{n} c_i^{(1)*} c_i^{(2)} = c^*\,(\Psi_1, \Psi_2) .$$

The short-form facts are simply

$$(\Psi_1, c\Psi_2) = c\,(\Psi_1, \Psi_2) , \quad (c\Psi_1, \Psi_2) = c^*\,(\Psi_1, \Psi_2) .$$

IP2. We now define the **norm** of a vector to be the inner product of the vector with itself:

$$(\Psi_1, \Psi_1) := \sum_{i=1}^{n} c_i^{(1)*} c_i^{(1)} = \sum_{i=1}^{n} |c_i^{(1)}|^2.$$

The modulus squared of a complex number is always real nonnegative so the same applies here. The norm of a vector being 0 also would imply that the vector is the null vector since all the components would be 0. Taking this further, we require that every vector in the state space $\mathscr{H}$ must have finite norm, and that every vector which satisfies the other conditions and does have finite norm is a vector in the state space. That is,

$$(\Psi_1, \Psi_1) = \sum_{i=1}^{n} |c_i^{(1)}|^2 < \infty.$$

Next, as

$$(\psi_1, \psi_2) = (\psi_2, \psi_1)^*,$$

we must have

$$(\psi, \psi) = (\psi, \psi)^* \implies (\psi, \psi) \in \mathbb{R}.$$

The *positive semidefinite metric* postulate further states that:

$$(\psi, \psi) \geq 0$$

and

$$(\psi, \psi) = 0 \iff \psi = 0.$$

IP3. Another important rule which is immensely helpful in solving quantum mechanical problems is that inner products can distribute linearly over a sum. This can be again shown by writing out the sum form:

$$\begin{aligned}(\Psi_1, a\Psi_2 + b\Psi_3) &= \sum_{i=1}^{n} c_i^{(1)*}(ac_i^{(2)} + bc_i^{(3)}) = \sum_{i=1}^{n} c_i^{(1)*} a c_i^{(2)} + \sum_{i=1}^{n} c_i^{(1)*} b c_i^{(3)} \\ &= (\Psi_1, a\Psi_2) + (\Psi_1, b\Psi_3) = a\,(\Psi_1, \Psi_2) + b\,(\Psi_1, \Psi_3)\,.\end{aligned}$$

Similarly,

$$\begin{aligned}(\Psi_1, a\Psi_2 + b\Psi_3) &= \sum_{i=1}^{n} (ac_i^{(2)} + bc_i^{(3)})^* c_i^{(1)} = \sum_{i=1}^{n} a^* c_i^{(2)*} c_i^{(1)} + \sum_{i=1}^{n} b^* c_i^{(3)*} c_i^{(1)} \\ &= (\Psi_1, a\Psi_2) + (\Psi_1, b\Psi_3) = a\,(\Psi_1, \Psi_2) + b\,(\Psi_1, \Psi_3)\,.\end{aligned}$$

As the sigma summation continues to distribute over any sum, the facts above can be extended to include more than three vectors in an inner product. The short-form facts are:

$$\left(\Psi_1, \sum_i c_i \Psi_i\right) = \sum_i [c_i \left(\Psi_1, \Psi_i\right)]$$

and

$$\left(\sum_i c_i \Psi_i, \Psi_1\right) = \sum_i [c_i^* \left(\Psi_i, \Psi_1\right)]$$

S4. The above then leads to more implications. If we define a linear combination of state space vectors ψ_1 and ψ_2 with coefficients c_1 and c_2, then

$$c_1\Psi_1 + c_2\Psi_2 := \Psi$$

is in the state space as

$$\begin{aligned}(\Psi, \Psi) &= (c_1\Psi_1 + c_2\Psi_2, c_1\Psi_1 + c_2\Psi_2) \\ &= c_1^* c_1 \left(\Psi_1, \Psi_1\right) + c_1^* c_2 \left(\Psi_1, \Psi_2\right) + c_2^* c_1 \left(\Psi_2, \Psi_1\right) + c_2^* c_2 \left(\Psi_2, \Psi_2\right)\end{aligned}$$

by separating the summation into the sum of these separate inner products (sum) by rule IP 3 of inner products, all the constants and inner products above must be finite so the norm of Ψ, a linear combination of state space vectors, is finite. This therefore means that for any basis vectors of the state space, a linear combination of them is also in the state space. This is the formal mathematical justification for why in the state space all possible state vector additions and therefore physical state superpositions are possible.

3.3.3 Orthonormality

The inner product is an operation which is defined for state space vectors in a set basis, matching-and-multiplying their components in the basis – or, alternatively, the values of their component functions, or wavefunctions, and producing a finite number. As it is so critical to all quantum mechanical calculation, it is necessary to be fluent with the above rules of the inner product as they will not always be repeated. If in doubt, writing out an inner product into the explicit sum form should be revelatory for those who are not yet so fluent.

Now we return to the idea that if we were working in an infinite dimensional vector space, like the state space, there are infinite bases which can span the vector space. This then poses the question of how to choose which bases we want to work in. Ultimately, the answer to this question cannot be given until we reach discussion on the representations of observables, since we want most of the time to work from the perspective of different observables when we are solving a problem; however, to every basis we have a process can be undertaken to make the basis substantially more convenient to work with, while still spanning the same space as a basis. The concept is of an **orthonormal basis**: the attribute of basis vectors being orthogonal to each other and also normalised. To understand how these vectors span the we will need to break down these two characteristics into simple definitions.

Two vectors $\boldsymbol{\alpha}$ and β are **orthogonal** if the following is true:

$$(\boldsymbol{\alpha}, \beta) = 0$$

we also know that this means that

$$(\beta, \boldsymbol{\alpha}) = (\boldsymbol{\alpha}, \beta)^* = 0$$

and therefore the order of the inner product does not matter for orthogonal vectors, since the complex conjugate of 0 is still 0. Meanwhile, a singular vector $\tilde{\boldsymbol{\alpha}}$ is said to be **normalised** if it has norm 1:

$$(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\alpha}}) = 1.$$

So for an **orthonormal basis** of a vector space $\mathbb{O} = \{\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, ..., \tilde{\boldsymbol{\alpha}}_n\}$:

$$\forall\ \tilde{\boldsymbol{\alpha}}_i, \tilde{\boldsymbol{\alpha}}_j \in \mathbb{O}, \quad (\tilde{\boldsymbol{\alpha}}_i, \tilde{\boldsymbol{\alpha}}_j) = \delta_{ij}$$

where δ_{ij} is the **Kronecker delta** (which will appear often in quantum mechanics):

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}.$$

In other words, an orthonormal basis set is a basis where all the vectors are orthogonal to each other and themselves are normalised (note that orthogonality is a property shared by two or more vectors, while normalisation

is a property of single vectors). If it is not clear now how an orthonormal basis will make manipulating inner products and therefore wavefunctions much easier, it will become clear in the close future. First, though, we will prove that every basis of linearly independent vectors can be converted to an orthonormal basis. This is in fact a famous theorem: the Gram-Schmidt Theorem, and it is a highly-useful fact to know that we can perform it on any set of linearly independent vectors.

Gram-Schmidt Theorem: Every basis of linearly independent vectors can be transformed by a defined procedure into an orthonormal set.

Proof- The Gram Schmidt Process: Let $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ...\boldsymbol{\alpha}_n\}$ be a linearly independent basis. We will start by normalising the first vector using a simple procedure.

Take the first vector, denoted without loss of generality as $\boldsymbol{\alpha}_1$. Take the norm of the vector,

$$|\boldsymbol{\alpha}_1| = \sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}.$$

By the positive semidefinite metric,

$$\begin{aligned}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1) \geq 0 \implies & \sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)} \in \mathbb{R} \\ \implies & |\boldsymbol{\alpha}_1|^* = \left(\sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}\right)^* = \sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)} = |\boldsymbol{\alpha}_1|\end{aligned}$$

Then, if we define

$$\tilde{\boldsymbol{\alpha}}_1 := \frac{\boldsymbol{\alpha}_1}{|\boldsymbol{\alpha}_1|}$$

we clearly get

$$(\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_1) = \frac{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}{|\boldsymbol{\alpha}_1|^* |\boldsymbol{\alpha}_1|} = \frac{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}{|\boldsymbol{\alpha}_1|^2} = 1$$

because

$$|\boldsymbol{\alpha}_1|^2 = \sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}^2 = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)\,.$$

Therefore, we indeed verify that for any arbitrary vector $\boldsymbol{\alpha}_1$ dividing it by its norm $|\boldsymbol{\alpha}_1| = \sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}$ creates a normalised vector $\tilde{\boldsymbol{\alpha}}_1$ whose inner product with itself is equal to 1. If the vector was already normalised such that $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1) = 1$ then division by its norm $\sqrt{(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1)}$ is equal to division by

$\sqrt{1} = 1$, so it would simply be unchanged by this process and remain normalised.

Now that we have normalised the first vector in our basis, we move on to consider the second vector $\boldsymbol{\alpha}_2$. Define

$$\boldsymbol{\alpha}'_2 := \boldsymbol{\alpha}_2 - \tilde{\boldsymbol{\alpha}}_1 \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\right).$$

Is it orthogonal to $\tilde{\boldsymbol{\alpha}}_1$? We can verify, remembering fact IP 3 that the inner product distributes linearly:

$$\left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}'_2\right) = \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\right) - \left(\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_1\right)\left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\right)$$

and since $\tilde{\boldsymbol{\alpha}}_1$ is normalised, $\left(\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_1\right) = 1$, so

$$\left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}'_2\right) = \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\right) - \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_2\right) = 0$$

Therefore $\boldsymbol{\alpha}'_2$ is indeed orthogonal to $\tilde{\boldsymbol{\alpha}}_1$. Now we can run the normalisation procedure:

$$\tilde{\boldsymbol{\alpha}_2} := \frac{\boldsymbol{\alpha}'_2}{\sqrt{\left(\boldsymbol{\alpha}'_2, \boldsymbol{\alpha}'_2\right)}}$$

This is still clearly orthogonal to $\tilde{\boldsymbol{\alpha}}_1$ because of how we defined $\boldsymbol{\alpha}'_2$. This time, we just get instead:

$$\left(\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2\right) = \frac{\left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}'_2\right)}{\sqrt{\left(\boldsymbol{\alpha}'_2, \boldsymbol{\alpha}'_2\right)}} = \frac{0}{\sqrt{\left(\boldsymbol{\alpha}'_2, \boldsymbol{\alpha}'_2\right)}} = 0.$$

The process from now continues in the same way for as many basis vectors we need to convert into this orthonormal basis. Orthogonal vector $\boldsymbol{\alpha}'_3$ is introduced by

$$\begin{aligned} \boldsymbol{\alpha}'_3 &= \boldsymbol{\alpha}'_3 - \tilde{\boldsymbol{\alpha}}_1 \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_3\right) - \tilde{\boldsymbol{\alpha}}_2 \left(\tilde{\boldsymbol{\alpha}}_2, \boldsymbol{\alpha}_3\right) \\ \Rightarrow \quad \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}'_3\right) &= \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_3\right) - \left(\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_1\right)\left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_3\right) - \left(\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2\right)\left(\tilde{\boldsymbol{\alpha}}_2, \boldsymbol{\alpha}_3\right) \\ &= \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_3\right) - 1 \times \left(\tilde{\boldsymbol{\alpha}}_1, \boldsymbol{\alpha}_3\right) - 0 \times \left(\tilde{\boldsymbol{\alpha}}_2, \boldsymbol{\alpha}_3\right) \\ &= 0 \end{aligned}$$

After that, it can be normalised by dividing by its norm as already proven for all vectors. This orthonormalisation process can run forever, regardless

of which other normalised basis vector we are taking the inner product with. We prove this rigorously in general form by induction.

Claim: For a set of orthonormal vectors $\mathbb{O} = \{\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, ...\tilde{\boldsymbol{\alpha}}_{n-1}\}$, the vector

$$\boldsymbol{\alpha}'_n = \boldsymbol{\alpha}_n - \left(\sum_{i=0}^{n-1} \tilde{\boldsymbol{\alpha}}_i \left(\tilde{\boldsymbol{\alpha}}_i, \boldsymbol{\alpha}_n\right)\right)$$

is orthogonal to all the vectors in $\mathbb{O}$.

Proof: The base case is for $n = 3$, which we have already proved above. The inductive step is:

$$\text{Assume we have an orthonormal set } \{\tilde{\boldsymbol{\alpha}}_1, \tilde{\boldsymbol{\alpha}}_2, ...\tilde{\boldsymbol{\alpha}}_{n-1}\}. \;\; \forall\, \tilde{\boldsymbol{\alpha}}_{j<n},$$

$$(\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}'_n) = (\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_n) - \tilde{\boldsymbol{\alpha}}_j \left(\sum_{i=0}^{n-1} \tilde{\boldsymbol{\alpha}}_i \left(\tilde{\boldsymbol{\alpha}}_i, \boldsymbol{\alpha}_n\right)\right)$$

$$\Rightarrow (\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}'_n) = (\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_n) - (\tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_j)\left(\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_n\right) - \sum_{i=0\neq j}^{n-1} (\tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_i)\left(\tilde{\boldsymbol{\alpha}}_i, \boldsymbol{\alpha}_n\right)$$

$$\Rightarrow (\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}'_n) = (\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_n) - 1 \times (\tilde{\boldsymbol{\alpha}}_j, \boldsymbol{\alpha}_n) - \sum_{i=0\neq j}^{n-1} 0 \times (\tilde{\boldsymbol{\alpha}}_i, \boldsymbol{\alpha}_n) = 0$$

This is true since our assumption that all $\tilde{\boldsymbol{\alpha}}_j$ for $j < n$ are normalised and orthogonal to all other $\tilde{\boldsymbol{\alpha}}_{i\neq j}$ for $i < n$ (an assumption which is valid as it follows from the base case) necessitates that

$$\forall\; i, j < n, \;\; (\tilde{\boldsymbol{\alpha}}_j, \tilde{\boldsymbol{\alpha}}_i) = \delta_{ij}$$

Therefore, by induction, our orthogonalisation process works for as many basis vectors as we need to convert to our orthonormal basis.

Then, after we obtain $\boldsymbol{\alpha}'_n$,

$$\tilde{\boldsymbol{\alpha}_n} = \frac{\boldsymbol{\alpha}'_n(x)}{\sqrt{(\boldsymbol{\alpha}'_n, \boldsymbol{\alpha}'_n)}}$$

is normalised and does not affect the orthogonality condition:

$$\left(\tilde{\boldsymbol{\alpha}}_{j\neq n}, \frac{\boldsymbol{\alpha}'_n}{\sqrt{(\boldsymbol{\alpha}'_n, \boldsymbol{\alpha}'_n)}}\right) = \frac{(\tilde{\boldsymbol{\alpha}}_{j\neq n}, \boldsymbol{\alpha}'_n)}{\sqrt{(\boldsymbol{\alpha}'_n, \boldsymbol{\alpha}'_n)}} = \frac{0}{\sqrt{(\boldsymbol{\alpha}'_n, \boldsymbol{\alpha}'_n)}} = 0.$$

We have now fully mathematically proven a procedure which allows us to take a linearly independent basis of any size and convert it to an orthonormal basis. It obviously is a tedious process if you have a large number of basis vectors, but its theoretical importance is twofold: the first, that once you have converted to an orthonormal basis, the inner product will become much easier to evaluate; the second, that we now know there is a valid process which exists– no matter if we choose to perform it or not. Therefore, we can assume *without loss of generality* that all bases we use are orthonormal, since if they weren't we would easily theoretically be able to orthonormalise them. This procedure is the elegant **Gram-Schmidt Process**: a hugely powerful weapon in linear algebra and indeed therefore in quantum mechanics.

The original point of this section was to further define a vector space in terms of its dimensions. Now, having defined orthonormality, we will go full circle back to the first vector space we considered– the Cartesian x,y plane, or, $\mathbb{R}^2$. In our most basic study of this vector space, we associate the numbers of dimensions with the number of perpendicular directions. We will find that in fact the natural way to come back to this assertion is to look at it not in terms of linearly independent vectors, but, in a very subtle change, in terms of the number of mutually orthogonal vectors. Let's see why this is an appropriate replacement from our earlier definition of dimensions in terms of linearly independent vectors.

Theorem: All mutually orthogonal vectors are also linearly independent.

This is very easy to prove. Suppose we have a set of mutually orthogonal vectors $\mathbb{O} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_n\}$. They are linearly independent if no nontrivial combination of these vectors is equal to 0. Let us write a generalised form of a linear combination of these vectors, and set it equal to the null vector:

$$\sum_{i=1}^{n} c_i \boldsymbol{\alpha}_i = 0$$

Now we can manipulate the mutual orthogonality of these vectors.

$$\forall\, j \in [1,n],\ \left(\boldsymbol{\alpha}_j, \sum_{i=1}^{n} c_i \boldsymbol{\alpha}_i\right) = (\boldsymbol{\alpha}_j, 0)$$

$$\Rightarrow\ c_j\,(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) + \sum_{i=1\neq j}^{n} c_i\,(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_i) = 0$$

$$\Rightarrow\ c_j\,(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) + \sum_{i=1\neq j}^{n} 0 \times c_i = 0$$

$$\Rightarrow\ c_j(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) = 0 \Rightarrow\ c_j = 0$$

unless $(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j) = 0$ which is impossible unless it is the null vector, which is irrelevant in these discussions as it is never considered a basis vector. This proof clearly works without loss of generality for each of the constants c_i, so it implies that every constant $c_i = 0$. Therefore the only possible linear combination of a set of mutually orthogonal vectors equal to the null vector is the trivial combination– and so they are all linearly independent if they are all mutually orthogonal.

Our final definition of dimensionality is therefore the number of mutually orthogonal vectors which can reside in a vector space. This number cannot be exceeded, since it is also the maximum number of linearly independent vectors which can be accommodated, and every new mutually orthogonal vector is in itself a new linearly independent vector.

Finally, comes the inner product punchline, which really shows why the operation, combined with an orthonormal basis, is so powerful. If we recall, for a given basis $\{\boldsymbol{\alpha}_i\}$ of the state space it spans the space and therefore all state vectors can be expressed in the form:

$$\Psi = \sum_i c_i \boldsymbol{\alpha}_i.$$

The coefficients c_i are called the components of the vector in that basis $\{\boldsymbol{\alpha}_i\}$, and are seemingly difficult to determine for the given basis depending on which vector we choose. However, consider the case when the basis chosen is orthonormal, or was not orthonormal but has now undergone orthonormalisation under the Gram-Schmidt Process. Now, consider the inner product

$$(\boldsymbol{\alpha}_j, \Psi)$$

for some given basis vector $\boldsymbol{\alpha}_j$. This is, according to the above expansion,

$$(\boldsymbol{\alpha}_j, \Psi) = \left(\boldsymbol{\alpha}_j, \sum_{\{i\}} c_i \boldsymbol{\alpha}_i\right).$$

Then by the fact S2 of the constant multiple rule and IP 3 of linear distributivity again, this is:

$$\sum_{\{i\}} c_i \left(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_i\right)$$

which by the Kronecker delta results in just

$$(\boldsymbol{\alpha}_j, \Psi) = \left(\sum_{i \neq j} \left(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_i\right)\right) + c_i \left(\boldsymbol{\alpha}_j, \boldsymbol{\alpha}_j\right) = 0 + c_j \times 1 = c_j.$$

This is critical, as we see that these components c_i in an orthonormal basis are not mathematically random: the component corresponding to a basis vector can be obtained through the inner product of that basis vector with the state vector! The much more eye-opening form of a state vector is after you make this substitution for the components:

$$\Psi = \sum_{\{i\}} \left(\boldsymbol{\alpha}_i, \Psi\right) \boldsymbol{\alpha}_i.$$

If we find a basis of the state space, we can immediately orthonormalise it by the Gram-Schmidt process, and if we can find an orthonormal basis for the state space and understand the vectors which form it we can then theoretically define any vector in that basis exceptionally easily. We therefore always assume we are working with an orthonormal basis: such an assumption is valid since the Gram-Schmidt procedure exists, and is useful because orthonormality can often make algebraic manipulations like the above.

There is a final note to make. If

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_i) = c_i$$

and

$$(\boldsymbol{\alpha}_i, \Psi) = c_i$$

for an orthonormal basis, we therefore have

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_i) = \left(\boldsymbol{\alpha}_i, \Psi\right).$$

In the continuous case, we have

$$(x, \Psi) = \psi_{\boldsymbol{\alpha}}(x)$$

where x is the continuously variable orthonormal basis vector. This is dramatic, because it tells us exactly how to form the discrete and continuous wavefunctions given some basis– convert it into an orthonormal basis, and find the inner product of the state vector with the basis vectors to produce the wavefunction! So, we are done with everything we need to know about the state vector. We have an object which represents physical states, and we have a way to transform it into a tangible wavefunction by which we can obtain probabilities of measurements in a way shortly to follow.

3.3.4 Scaling State Vectors

Through careful discussion, we have begun to understand the key bijection which forms the state postulate:

$$\text{Physical States} \leftrightarrow \text{Hilbert Space Vectors.}$$

Technically, however, this is still not quite correct. The *actual* bijection is

$$\text{Physical States} \leftrightarrow \text{Hilbert Space Rays.}$$

Here, a 'ray' is a vector with norm 1. So in particular, taking any state vector Ψ and performing

$$\Psi \to \tilde{\Psi} := \frac{\Psi}{\sqrt{(\Psi, \Psi)}}$$

transforms it into a ray in "the same direction". This notion of a "direction" comes from the arrow visualisation of vectors, but with the caveat that we are in infinite dimensional space so we cannot actually visualise geometrically what these directions should look like. However, the notion of a direction is useful for our understanding because it helps us understand that we can take a unit ray and then multiply it by any scalar, creating a new Hilbert space vector, but which is equivalent to the original unit ray *up to some scaling*.

In practice, quantum mechanics regards these scalings as 'unphysical phase factors'. In other words, given some normalised state $\tilde{\Psi}$ and any constant $c \in \mathbb{C}$ then the state $\Psi := c\tilde{\Psi}$ is "the same state, up to some unphysical

scaling". Hilbert space vectors which do not have norm 1 are not physical states, but since we can always divide by their norm to create a unit ray in the Hilbert space, they can be considered to be physical states with some redundant scale factor attached. It may seem rather arbitrary to discard the multiplicative constants like this, but there are two logical justifications for this:

- Since we must accept *any* superposition of basis states as a new state, then
$$\forall c_i \in \mathbb{C}, \quad \sum_{\{i\}} c_i \boldsymbol{\alpha}_i \text{is a state}$$
but then
$$\forall c_i \in \mathbb{C}, \quad \sum_{\{i\}} (10000000) c_i \boldsymbol{\alpha}_i \text{is a state}$$
and it obviously would seem quite pointless if multiplying by arbitrary numbers in the above way created something completely new every time.

- More concretely, the components of the state vector are connected to probabilities of measuring the basis states in some way. We do not know how just yet, but intuitively, it should make sense that it the probabilistic distribution is about the *relative weight* that each basis state gets, not about the absolute weight. In other words, multiplying every component by 10000000 might change the absolute value of the components, but the component of some arbitrary basis vector $\boldsymbol{\alpha}_i$ will still be double the component of some arbitrary basis vector $\boldsymbol{\alpha}_j$ if this was originally the case before multiplying by 10000000. So the relative likelihood of $\boldsymbol{\alpha}_i$ over $\boldsymbol{\alpha}_j$ is in a sense still conserved.

- The reason why having norm 1 makes a state physical, we will see, is related to the fact that the probabilities for all possibilities of a probabilistic event must sum to 1: something happens with probability 1. This is why Hilbert space unit rays are the physically relevant state vectors, and any other scaled Hilbert space vectors should be reduced to the unit rays in the same directions.

  This discussion will become more concrete in the near future once we actually connect the components to measurement probabilities, but

hopefully it should intuitively make sense why scaling state vectors by arbitrary constants should not really have much meaning.

In practice, this will actually make our life much easier, because when we solve problems and algebraically manipulate the state vector we do not have to preserve normalisation through every step: we can work with these unphysical scalings, and only at the end of the problem we can normalise the final Hilbert space vector to become a physical state space vector. Since unphysical phase factors do not differ from the relevant unit ray state, we will always be transforming between states in the correct ways regardless of the norms of the state vectors in the intermediary steps.

3.3.5 Wavefunctions

We now understand that there is a bijection

$$\text{Physical States} \leftrightarrow \text{Hilbert Space Rays}$$

However, recall our lengthy discussion on how the state vector, without a basis, is *formless*. We must specify a coordinate basis to numerically specify the position of an object on a map, and similarly we must specify a vector basis for our state space in order to give the state vector an actual form. On the other hand, if we *are* given some fixed basis, say $\{\boldsymbol{\alpha}_i\}$, we can write

$$\Psi = \sum_{\{i\}} c_i \boldsymbol{\alpha}_i$$

for some set of coefficients $\{c_i\}$, and in particular this set of coefficients is unique, since we have already proved the expansion of any constituent vector in a spanning basis of a vector space is unique.

Then, however, we can create a new bijection. **In some fixed basis**,

$$\text{State Vectors} \leftrightarrow \text{Sets of Components}$$

using the fact every state vector has a unique expansion, and every expansion corresponds to a unique state vector. However, we can do better, by defining the mapping

$$\psi_{\boldsymbol{\alpha}}(x) : \mathbb{B} \to \mathbb{C}, \;\; \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_i) \mapsto c_i$$

which takes an input basis vector from the basis $\mathbb{B}$ and outputs its component with respect to that basis vector. Note here that the subscript α on ψ_α reflects that in practice we always use the same letter with varying indices to denote a set of basis vectors, and thus can use the same letter to indicate what basis our map is defined in. Now, what does that map look like? There are two cases:

1. **Discrete case:** The map will also clearly take a discrete set of values, one for each of the discrete basis vectors. A clue is that this map *resembles the probability mass function*, which also outputs a discrete set of scalars depending on a discrete set of outcomes.

2. **Continuous case:** The map should also clearly take a continuous set of values. The reason for this is that in for the concept of a continuous basis to arise we must have some notion of basis vectors being 'infinitely close' to each other (otherwise there would be gaps between basis vectors and we would have a discrete basis). But then, if the Hilbert space state vector is well defined then clearly for infinitesimally close basis vectors we must have infinitesimally close coefficients in its expansion, which in turn makes this map have a continuous output.

   A clue is that this map *resembles the probability density function*, which is also a continuous function, where the output scalars (which are probabilities) are infinitesimally close for infinitesimally close inputs (outcomes).

So we see why our notion of this 'component map' will be crucial. Given some meaningful basis, our component map should be able to transform into a probability distribution function! Now we do not know how it does so numerically yet, and neither do we know which bases are *meaningful*, but we will reach this imminently.

Remember, the form of this map, which we henceforth call the **wavefunction**, changes based on which basis we have picked to express it in (since the output components will be different). Thus, there are infinitely many different wavefunctions. We have thus a new rich set of bijections!

State Vectors $\leftrightarrow$ Basis Wavefunctions

but also, since

$$\text{Physical States} \leftrightarrow \text{State Vectors}$$

we therefore must also have

$$\text{Physical States} \leftrightarrow \text{Basis Wavefunctions}$$

So in a sense our state vector is the more general, intermediate object, which is in a *basis independent* bijection with physical states; the *basis-dependent* wavefunction is then our final object, which we can use to actually encode and deduce probabilities.

3.3.6 Summary

We have begun our study of the quantum state problem. We know that there is a state vector in bijection with physical states, and how to develop functional forms of it. One can now move onto the next chapter, where we will subsequently be expanding this theory, especially with regards to observables. Most importantly, we need to understand which bases are useful for us in quantum mechanics; to do this, we will need to fit observables into the picture and very fruitful results will follow.

3.4 Exercises from Chapter 3∗

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

Chapter 4

Chapter 4: State Space Operators and the Quantum Observables Problem

The previous chapter introduced the concept of Hilbert space vectors, called state vectors, representing physical states. The state space, like any infinite-dimensional vector space, has infinite bases, and this was claimed to be a motivating factor for the Hilbert space formulation in the first place: because it allows for the state represented by the state vector to be looked at from the perspective of different observables like momentum and position. To complete our understanding of how the state problem is approached in quantum mechanics we need to understand exactly how observables are incorporated into this linear algebraic vector space system, through very specific bases. This question will be answered by an investigation into **linear operators**.

4.1 Hilbert Space Operators

An operator is a function, but it is the lexicon used to describe a sort of function which acts on functions or vectors as inputs– here, in the state space, on state vectors. For a new student of quantum mechanics it is paramount to understand that the distinction between an operator and a function does not exist. The only reason the term operator is used in conjunction with vector spaces is to avoid exhausting the term function, since operators act on state

vectors which are themselves linked already to wavefunctions. If unused to the term, there is a harmful tendency to give a mystic image to operators, but this must not be done. An operator, just like a function, takes an input and maps it to an output, and that is the extent of it.

For example, one operator the reader will have seen already is the differential operator:

$$\frac{d}{dx}f(x) = f'(x).$$

This takes some input function $f(x)$, and maps it to an output function $f'(x)$. In that sense, it is a function because it has an input and output, but it is an operator because it acts on functions. If the input was $f(x) = x^2$ then the output of the differential operator would be $2x$; if it were e^x then its output would also be e^x. A scalar can also be an input of the operator: since an operator is a function of functions, we could just define the input function to be $f(x) \equiv k$ for that scalar k, and of course this would still be a function whilst sharing no differences to the scalar. Here, if such a scalar was an input of the differential operator, than the output would be the null function: or, 0, but for other operators, this may be different.

If this is understood, the technical requirement for an operator which is linear, here donated by Ω, is that they are linear maps:

$$\Omega : \mathscr{H} \mapsto \mathscr{H}, \quad \Omega(\Psi_1 + \Psi_2) \equiv \Omega\Psi_1 + \Omega\Psi_2.$$

There do exist non-linear operators, but in this book, and much of quantum mechanics, we will only ever encounter linear operators, and so we can take the latter linear distributive property to be a given for our work. We also have:

- Associativity of scalar multiplication:

$$(c\Omega)\Psi = c(\Omega\Psi).$$

- Distributivity:

$$(\Omega_1 + \Omega_2)\Psi = \Omega_1\Psi + \Omega_2\Psi.$$

- Associativity in operators:

$$\Omega_1(\Omega_2\Psi) = \Omega_1\Omega_2\Psi$$

We do not, however have commutativity:

$$\Omega_1\Omega_2\Psi \neq \Omega_2\Omega_1\Psi$$

for most pairs of operators Ω_1 and Ω_2 (some do by chance). Moreover, there is in general no useful relation between $\Omega_1\Omega_2$ and $\Omega_2\Omega_1$, unlike the inner product, which is not commutative, but which follows the much nicer relation

$$(\Psi_1, \Psi_2) = (\Psi_2, \Psi_1)^*.$$

The general irregularity of commutation relations between linear operators will turn out to be connected to an incredibly rich set of quantum phenomena!

The rest of the operator facts again should, like the rules of vector spaces, come mostly naturally. A useful point to remember is that for any operator we work with in quantum mechanics, it maps from the state space $\mathscr{H}$ to the state space $\mathscr{H}$: in other words, performing that operator on any input state space vector will directly create a new state space vector! Thus

$$\Omega\Psi$$

is a new state space vector, and so is

$$\Omega_1\Omega_2\Psi = \Omega_1(\Omega_2\Psi)$$

which is, the action of Ω_1 on the state space vector $\Omega_2\Psi$ obtained after operating with Ω_2 on an original state vector Ψ. We have a tendency, which is natural, to be suspicious when some arbitrary function acts on an immensely complex input vector. However, in the rules of quantum mechanics where linear operators dominate, we do not need this suspicion. Fluency should soon dictate that we see $\Omega\Psi$ as just another state space vector.

The caveat which must be noted is that general linear operators will not send a unit ray to another unit ray: i.e, if the input vector is normalised the transformation will usually give them a new unphysical phase factor. Yet we know the Hilbert space can contain these unphysically scaled state vectors, which are equivalent to the unit ray states, so we are fine, as we can renormalise them later.

The special class of operators which *do* preserve the normalisation of unit

ray states are called **unitary** operators. We will cover these later.

Now, the much more important qualities of operators in quantum mechanics are that they possess quantities called eigenvalues and eigenvectors, which is where the quantum mechanical formalism will really start to come together.

4.1.1 Eigenvalues and Eigenvectors

We have just studied operators, which are fundamentally crucial to all quantum mechanics, but of the moment seem to have vague physical interpretations. The other side of the same coin is **eigenvalues**. Flip to the middle chapter of any quantum mechanics text and you might find a large mixture of words with the prefix "eigen-": eigenvector, eigenfunction, eigenvalue, eigenenergy, eigenstate, eigenmomentum- et cetera. This will soon become natural, though it may be initially daunting. Understanding why these words are so commonplace will set us in a good position, so we do that now.

Consider a three dimensional example first. We will pick a random vector, which in this case can be a column vector,

$$V = \begin{pmatrix} \frac{47}{10} \\ 5.69 \\ 242 \end{pmatrix}$$

and a random operator,

$$\Omega \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_2 \\ x_3 \\ x_1 \end{pmatrix}$$

Upon applying this operator to the vector V, we are likely not going to get an output vector which is anything like the original vector in terms of how corresponding components are related.

We are however given the grace from mathematics of sets of vectors, belonging to specific operators, which behave much more stably. To each operator Ω there exists a set of vectors, called **eigenvectors**, such that:

$$\Omega\omega = \lambda\omega$$

for some constant λ and vector ω. In the above equation, which we call an eigenvalue equation, ω is an eigenvector of Ω, and λ is the corresponding eigenvalue: a complex constant. This equation corresponds to the operator Ω scaling the eigenvector ω by a scale factor λ: a relatively very trivial transformation compared to the nontrivial possibilities we expect for general vectors under the transformation. In the above example we just saw, for example, no such scaling exists at all, despite the transformation looking somewhat simple!

Let us consider a basic operator and try to solve for all possible eigenvectors and eigenvalues. We could use the identity operator, which is extremely trivial:

$$I\epsilon = \epsilon$$

means that any vector ϵ is an eigenvector of the identity operator, with corresponding eigenvalue 1. It is a rather uninteresting result, which pertains to a uniquely simple vector.

This eigenvalue equation was uniquely simple, but for more complicated operators, it is clear that trying to find eigenvectors by inspection will usually be futile. The more advanced methods of solving eigenvalue problems will come later in this book; for now, only the theory is important.

A final note on the above: the prefix eigen-, which is derived from German and means 'own' (hence, each operator has its 'own' set of eigenvectors), is always used in mathematics when we are dealing with the above cases. We therefore have eigenvalues, eigenvectors, eigenfunctions: but also, later on, eigenenergies, eigenmomenta, and so on, when the eigenvalues are energy and momentum values respectively. The context and this explicit note should demystify such eigen- words in the future.

4.1.2 Hermitian Operators

The next important definition central to quantum mechanics is of **Hermitian** operators. Operators are Hermitian if they possess the property:

$$(\boldsymbol{\alpha}, \Omega\boldsymbol{\beta}) = (\Omega\boldsymbol{\alpha}, \boldsymbol{\beta})\,.$$

for *any* vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$. This 'Hermiticity' property is deceptively simple, but here are some profound consequent facts for hermitian operators following

this definition.

H1. Hermitian operators must have real eigenvalues.

Proof:

For a hermitian operator the eigenvalue condition is the same

$$\Omega\omega = \lambda\omega.$$

We have, in taking the inner product with the eigenvector ω:

$$(\omega, \Omega\omega) = (\Omega\omega, \omega)$$

by the definition of Hermiticity, so

$$\begin{aligned}(\omega, \lambda\omega) = (\lambda\omega, \omega) \Rightarrow \;\; & \lambda(\omega, \omega) = \lambda^*(\omega, \omega) \\ \Rightarrow \;\; \lambda = \lambda^* \Rightarrow \;\; & \lambda \in \mathbb{R} \;\; \square\end{aligned}$$

H2. Eigenvectors ω_1 and ω_2 of the same hermitian operator corresponding to different eigenvalues are orthogonal to each other.

Proof:

To prove that ω_1 and ω_2 are orthogonal we need to prove that $(\omega_1, \omega_2) = 0$. We can do this by manipulating the hermitian property of the operator, here denoted Ω.

$$\begin{aligned}(\omega_1, \Omega\omega_2) &= (\Omega\omega_1, \omega_2) \\ \Rightarrow \;\; (\omega_1, \lambda_2\omega_2) &= (\lambda_1\omega_1, \omega_2) \\ \Rightarrow \;\; \lambda_2(\omega_1, \omega_2) &= \lambda_1^*(\omega_1, \omega_2)\end{aligned}$$

but λ_1 is an eigenvalue of a hermitian operator so it is real: i.e, $\lambda_1^* = \lambda_1$. So above we had

$$\lambda_2(\omega_1, \omega_2) = \lambda_1^*(\omega_1, \omega_2)$$

which is,

$$\lambda_2(\omega_1, \omega_2) = \lambda_1(\omega_1, \omega_2).$$

Therefore, if the eigenvectors do not have the same eigenvalue then $\lambda_2 \neq \lambda_1$ so the above implies that $(\omega_1, \omega_2) = 0$ and so these eigenvectors must be orthogonal. $\square$

There is a note for the above proof, however. The proof works on the assumption that for different eigenvectors their eigenvalues are also different. This is not always a correct assumption. Consider the identity operator I. It is clearly hermitian:

$$(\Psi_1, I\Psi_2) = (\Psi_1, \Psi_2) = (I\Psi_1, \Psi_2)\,.$$

However, it has infinite different eigenvectors but they all have the same eigenvalue: 1. Thus the proof above cannot apply to the identity operator. In this case, the collapse of the point should be obvious anyway– all vectors are eigenvectors of the identity operator, but certainly not all vectors are orthogonal to each other. In general, an operator where different eigenvectors can share the same eigenvalue is said to be **degenerate**. In particular, we say particular eigenvalues are degenerate if they can correspond to multiple different eigenvectors– but we do not say eigenvectors are degenerate because an eigenvector can never correspond to multiple different eigenvalues.

There are a few proofs of theorems in this book which involve the assumption that we are working with non-degenerate operators. These proofs, when we incorporate degeneracy, are usually different – and unfortunately, more difficult. For every step where we assume non-degeneracy, it would still be within the reaches of this book to prove an alternative proof in the case of degeneracy, but at the same time these would take labour and space. Therefore, I will not include them in this book because they will not alter anything in the fundamental understanding of a reader. Should the reader want to find such proofs, they may turn to a more advanced textbook which has the space and desire to cover this technicality. It should not make a massive difference either way whether the reader is aware of the degenerate case proof, so long as they understand when degeneracy makes a difference to actual consequent theorem or result (and indeed when it does not, which is quite common here). I will highlight these cases at when they occur.

H3. For an operator Ω with real eigenvalues λ_i and eigenvectors $\boldsymbol{\alpha}_i$, if the eigenvectors constitute an orthonormal basis in the Hilbert space then the operator is hermitian.

Proof:

Take the component expressions for the two arbitrary vectors Ψ_1 and Ψ_2. We know they can be both expressed as a linear combination of the eigenvectors $\boldsymbol{\alpha}_i$ since the set $\{\boldsymbol{\alpha}_i\}$ is stated in the conditions to be an orthonormal basis. So we have:

$$\Psi_1 = \sum_{\{i\}} c_i \boldsymbol{\alpha}_i, \;\; \Psi_2 = \sum_{\{j\}} \gamma_j \boldsymbol{\alpha}_j$$

where the components c_i and γ_j are found by $(\boldsymbol{\alpha}_i, \Psi_1)$ and $(\boldsymbol{\alpha}_j, \Psi_2)$ respectively. Then

$$\begin{aligned}(\Omega\Psi_1, \Psi_2) &= \left(\Omega \sum_{\{i\}} c_i \boldsymbol{\alpha}_i, \sum_{\{j\}} \gamma_j \boldsymbol{\alpha}_j\right) \\ &= \left(\sum_{\{i\}} c_i \Omega \boldsymbol{\alpha}_i, \sum_{\{j\}} \gamma_j \boldsymbol{\alpha}_j\right)\end{aligned}$$

where the operator can be incorporated into the sum term as it is a linear operator. Then this becomes

$$\begin{aligned}(\Omega\Psi_1, \Psi_2) &= \left(\sum_{\{i\}} c_i \lambda_i \boldsymbol{\alpha}_i, \sum_{\{j\}} \gamma_j \boldsymbol{\alpha}_j\right) \\ &= \sum_{i,j} c_i^* \lambda_i^* \gamma_j (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)\end{aligned}$$

but $\{\boldsymbol{\alpha}_i\}$ is an orthonormal basis so $(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = \delta_{ij}$. The above then becomes 0 except for when the two basis vectors are the same, so we are left with:

$$(\Omega\Psi_1, \Psi_2) = \sum_i c_i^* \lambda_i^* \gamma_i.$$

Now, considering $(\Psi_1, \Omega\Psi_2)$, we have very similarly:

$$\begin{aligned}(\Psi_1, \Omega\Psi_2) &= \left(\sum_{\{i\}} c_i \boldsymbol{\alpha}_i, \Omega \sum_{\{j\}} \gamma_j \boldsymbol{\alpha}_j\right) \\ &= \left(\sum_{\{i\}} c_i \boldsymbol{\alpha}_i, \sum_{\{j\}} \gamma_j \Omega \boldsymbol{\alpha}_j\right) \\ &= \left(\sum_{\{i\}} c_i \boldsymbol{\alpha}_i, \sum_{\{j\}} \gamma_j \lambda_j \boldsymbol{\alpha}_j\right) \\ &= \sum_{i,j} c_i^* \lambda_j \gamma_j (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j) = \sum_{i,j} c_i^* \lambda_j \gamma_j \delta_{ij} \\ &= \sum_i c_i^* \lambda_i \gamma_i.\end{aligned}$$

so we have

$$(\Omega\Psi_1, \Psi_2) = \sum_i c_i^* \lambda_i^* \gamma_i, \quad (\Psi_1, \Omega\Psi_2) = \sum_i c_i^* \lambda_i \gamma_i$$

but we have conditioned that the eigenvalues λ_i are real so we therefore see that $\lambda_i = \lambda_i^*$ and so

$$(\Psi_1, \Omega\Psi_2) = (\Omega\Psi_1, \Psi_2)$$

which is the definition of a Hermitian operator. This holds true for any arbitrary Ψ_1 and Ψ_2 so long as they are in the space spanned by the orthonormal basis and can subsequently be expressed as a linear combination of the orthonormal constituent vectors; therefore, any operator with real eigenvalues whose eigenvectors can form an orthonormal basis set is Hermitian. □

This proof in fact goes both ways: more significantly, any Hermitian operator possesses a set of eigenvectors which are an orthonormal basis set of the state space! The proof is rather technical, so it will be ignored– but the profound consequences are clear. If an operator in the state space is Hermitian it has a basis consisting eigenvectors, called its **eigenbasis**, spanning the space; if we take any basis of the state space we can take its inner product over all the basis eigenvectors with any state vector to produce a wavefunction. These representations will prove massively helpful.

H4. The action of all Hermitian operators whose eigenvectors form an orthonormal basis can be specified by their eigenvalues and eigenvectors.

Proof:

For a Hermitian operator Ω with eigenvalues λ_i and eigenvectors $\boldsymbol{\alpha}_i$, any vector can be expressed in the orthonormal basis:

$$\Psi = \sum_i (\boldsymbol{\alpha}_i, \Psi)\boldsymbol{\alpha}_i$$

so

$$\Omega\psi = \Omega \sum_i (\boldsymbol{\alpha}_i, \psi)\boldsymbol{\alpha}_i = \sum_i (\boldsymbol{\alpha}_i, \psi)\Omega\boldsymbol{\alpha}_i = \sum_i (\boldsymbol{\alpha}_i, \psi)\lambda_i\boldsymbol{\alpha}_i.$$

This clearly requires no external knowledge other than understanding the sets $\{\lambda_i\}$ and $\{\boldsymbol{\alpha}_i\}$. Conversely: if we completely understand these sets then we can completely specify the operator given an input vector ψ the operator is acting on.

With all this knowledge about operators in vector spaces, and clear signs that their eigenbases will be very useful, we now come to the Second Postulate of quantum mechanics.

4.2 Observables in Quantum Mechanics

The state problem is a question of information. The most relevant information to a physicist about a state is the value of its *observables*. The problem, we have seen, is that unlike with the classical state, the quantum state cannot just be said to possess a value for any observable, as it is instead a superposition of many different possible states and it is thus far unclear which state will emerge upon measurement. This problem was dealt with by placing states in correspondence with vectors in a vector space which meant that all possible states could be summed together in a superposition to create a new state without forming something out of the space of possible states, and we learnt how to create more 'tangible' wavefunctions from those state vectors. However, we still do not know why wavefunctions are so tangible. The genius of the formalism comes when we incorporate operators and their orthogonal eigenbases into the picture, where wavefunctions in eigenbases will answer our problem.

Postulate 2: Observables

To each physical observable there exists a corresponding hermitian operator. There exists an orthonormal eigenbasis of this operator which spans the state space: that is, for some observable operator $\hat{A}$ with eigenvalues $\{A_i\}$ corresponding to eigenvectors $\{\boldsymbol{\alpha}_i\}$,

$$\forall\, \Psi,\ \ \Psi = \sum_i c_i \boldsymbol{\alpha}_i$$

for some components $\{c_i\}$. The only possible values for the observable whose operator is $\hat{A}$ which can be measured are the eigenvalues $\{A_i\}$ and these correspond to the eigenstates $\{\boldsymbol{\alpha}_i\}$. If there comes a condition

$$\hat{A}\Psi = A_i \boldsymbol{\alpha}_i$$

then we say that the eigenvalue A_i is the measured value of the physical observable and the new state vector is the eigenvector $\boldsymbol{\alpha}_i$.

The most essential outcome to us here is that we want real values for the results of physical measurements, as imaginary position or imaginary energy for example would be nonsense; since the set of eigenvalues $\{A_i\}$ are the only possible results of measurements, we therefore require real eigenvalues. By the postulate, we have Hermitian operators representing observables, which therefore must have real eigenvalues. The postulate is that these eigenvalues are the only possible measurable results for that observable for any states, so that they are real is critical. It is from the fact we need the operator to have real eigenvalues if these are to be the measured values for a physical operator, combined with the fact it is postulated to have an orthonormal eigenbasis spanning the state space, which guarantees the operators representing observables must be Hermitian. That is precisely, we recall,

$$\left(\hat{A}\Psi_1, \Psi_2\right) = \left(\Psi_1, \hat{A}\Psi_2\right).$$

We start a fuller discussion by noting the assertions of the postulate in short-form:

P2A1. All physical observables are represented by hermitian operators whose eigenvectors form an eigenbasis which spans the state space. For brevity it is customary to call these operators which represent physical operators 'observable operators' throughout the course of this book.

P2A2. Each measurement of a physical observable must yield one of the real eigenvalues of the observable operator.

P2A3. As it is possible for an operator to have a finite/and or discrete number of eigenvalues, so can a physical observable have a finite number of possible results after being measured if their operator has a finite number of eigenvalues. Such a situation is rarer than one would assume in quantum mechanics, considering we are working in an infinite dimensional vector space where it is relatively unlikely there are not infinite eigenvectors and eigenvalues to an arbitrary operator, but it is far from a non-existent possibility. The physical phenomenon resulting from discretely distributed eigenvalues is called quantization; its implications, most famously perhaps in the discrete energy levels of electrons, are important and may well have been already known to the reader.

Now we must consider the importance of time. Crucial is that operators corresponding to physical observables never change with time, and there is only one operator corresponding to each physical observable. That does not mean that the measured values of the observable will remain constant across any time period. That is because clearly the measured value of the observable depends on the state vector Ψ representing the state of the system which is the input vector we the observable operator is operating on; we have already stated this state vector can evolve with time. The precise nature of all these temporal considerations will be covered in due course, but that observable operators do not evolve with time is surely a great relief, especially if we need to think about their eigenbases and eigenvalues and do not want to have to continually solve what are not trivial eigenvalue equations.

The second part of the postulate gives us an interesting and significant link between the state vectors which represent states, the state space operators representing physical observables, and the eigenvalues of those observables representing possible results after measurement.

To start off, note that we expect that most state space vectors will be linear combinations of the eigenvectors of any observable operator, since the set of (infinite) eigenvectors of a hermitian operator constitutes an orthonormal basis of its space. We do not expect all possible states to be pure scalar multiples of single eigenvectors since there can be infinite linear combinations of

the eigenvectors which are not pure scalar multiples of single eigenvectors. The postulate now states that the condition

$$\hat{A}\Psi = A_i \boldsymbol{\alpha}_i$$

means that A_i is the measured value of the physical observable represented by the operator $\hat{A}$. However, this condition is clearly very singular if we are working with a wavefunction which is a linear combination of the infinite eigenvectors of an observable operator– since the state Ψ does not naturally coincide with the eigenvector $\boldsymbol{\alpha}_i$ alone. What should be abundantly clear, however, is that, post-measurement, Ψ has changed from a linear combination of eigenvectors to a multiple of only one of them, the eigenvalue corresponding to which is the measured value of the observable. So the act of measurement is clearly very important; indeed, it forms one of the central pillars of quantum mechanics and especially the mathematics which formulates it. We have seen this, already, in the Stern Gerlach experiment! There, measuring the x spin led to irrevocable changes in the y spin even without physically affecting it in that dimension. This measurement problem is the final component of the quantum mechanical solution to the state problem, and pulls everything together in an understandable way. Thus to complement Postulate 2 on observables, we have Postulate 3, on Measurements.

4.2.1 Measurements

Postulate 3: Measurements

After a measurement of a physical observable, the state vector is forced into a specific eigenvector corresponding to the eigenvalue measured for that observable. The probability that the (normalised) state vector is forced into a state represented by a state vector $\boldsymbol{\alpha}_i$, which is called an eigenstate, is given by

$$P(\boldsymbol{\alpha}_i) = |\,(\boldsymbol{\alpha}_i, \Psi)\,|^2,$$

which is therefore also the probability of measuring the eigenvalue A_i as the final result of the measurement for the observable. This Postulate now provides great meaning to the discourse immediately preceding this section. By *forcing* a state vector into specific eigenvector of an observable operator after a measurement of that specific observable, we guarantee several things:

- We do not restrict the state vector to being a pure scalar multiple of one single eigenvector prior to measurement. This is important as by Postulate 1, all state space vectors represent physical states, and there certainly should be infinite Hilbert space vectors which are linear combinations of any orthonormal basis vectors which span it: including when that orthonormal basis is the eigenbasis of an observable operator. This is the superposition of different possible states which exists before a measurement.

- We guarantee that after measurement, Postulate 2 has meaning: since the measurement forces the wavefunction into a specific eigenstate, we will indeed achieve after measurement

$$\hat{A}\Psi \rightarrow \hat{A}\boldsymbol{\alpha}_i = A_i\boldsymbol{\alpha}_i$$

  and therefore we guarantee that a measurement will always yield one single value– the eigenvalue A_i: regardless of what superposition of states it was in previously.

This is all well and good, but what gives order to the chaos is the probabilistic link of the postulate. Without it, we would be wondering what to do in any arbitrary superposition in states, since intuition tells us that just because we have a superposition it does not mean that all the measurements must have equal probabilities. Fortunately, we have the postulate:

$$P(\boldsymbol{\alpha}_i) = |\,(\boldsymbol{\alpha}_i, \Psi)\,|^2$$

where $P(\boldsymbol{\alpha}_i)$ is the probability of measuring the state vector to be in the state $\boldsymbol{\alpha}_i$. Here we make a clarification of similar type as that regarding the distinction between state and wavefunction: the reader must understand that the state vector being in an eigenstate $\boldsymbol{\alpha}_i$ is not so interesting itself as is the fact that when it is in that eigenstate we know A_i is the eigenvalue is the measured result for the observable. Thus when we say the state is measured to be the eigenstate $\boldsymbol{\alpha}_i$ we really allude to the fact that a measurement will yield A_i as the value. The reason we do not write $P(A_i)$, the probability of measuring A_i, is due to the fact that in the face of degeneracy (say, eigenvalue A_1 corresponding to two different eigenstates $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$), we have the following problem:

$$P(A_1) \neq (\boldsymbol{\alpha}_1, \Psi) \neq (\boldsymbol{\alpha}_2, \Psi)$$

in fact, here it would be

$$P(A_1) = (\boldsymbol{\alpha}_1, \Psi) + (\boldsymbol{\alpha}_2, \Psi)\,.$$

So we see that defining the probability of a wavefunction being in an eigenstate is slightly easier and more consistent. Next, consider the state

$$\Psi = \boldsymbol{\alpha}_n.$$

In this state a measurement will yield value A_n with probability

$$P(A_n) := |\,(\boldsymbol{\alpha}_n, \Psi)\,|^2 = |\,(\boldsymbol{\alpha}_n, \boldsymbol{\alpha}_n)\,|^2 = 1$$

since the eigenvectors $\boldsymbol{\alpha}_n$ are assumed to be normalised. So if the state vector is a pure eigenstate then the eigenvalue corresponding to the eigenstate it is in will be measured with probability 1. When do pure eigenstates occur for state vectors? They may occur organically for some arbitrary physical state which happens to be a pure eigenstate of a physical observable, though we expect this to be comparatively rare. More importantly: they also occur after measurements, since by the first part of the postulate a measurement will force a state vector into an eigenvector – a pure eigenstate – of the observable operator. This now explains why instantaneous successive measurements must yield the same answer: the first measurement forces the state vector into a pure eigenstate $\boldsymbol{\alpha}_n$ corresponding to the eigenvalue A_n measured, and then the second measurement will give the same eigenvalue A_n with probability

$$P(A_n) = |\,(\boldsymbol{\alpha}_n, \boldsymbol{\alpha}_n)\,|^2 = 1$$

since the state vector is now the pure eigenstate $\boldsymbol{\alpha}_n$ after being forced into this eigenstate by the first measurement. We saw this intuitive consequence in the Stern Gerlach experiment, where successive magnetic fields in the same axis yielded the same spin results each time!

The disturbance to classical intuition comes when we make the same observation we have already made. By Postulate 1, all state space vectors represent physical states, and there are infinite state space vectors which are linear combinations of any orthonormal basis vectors which span it. Thus if the orthonormal basis is the eigenbasis of a physical observable operator, there are infinite state vectors which are not pure eigenvectors of the observable

operator, but rather linear combinations of the corresponding eigenvectors. Then,

$$\Psi = \sum_i (\boldsymbol{\alpha}_i, \Psi)\, \boldsymbol{\alpha}_i,$$

by the expansion of Hilbert Space vectors in an orthonormal basis. But then, for any A_n in a non-degenerate state (similar reasoning holds for degenerate states),

$$P(A_n) := |\, (\boldsymbol{\alpha}_n, \Psi)\, |^2 = |\left(\boldsymbol{\alpha}_n, \sum_i (\boldsymbol{\alpha}_i, \Psi)\, \boldsymbol{\alpha}_i\right)|^2$$

so if this probability was zero then that would imply the component $(\boldsymbol{\alpha}_i, \Psi) = 0$. In a nontrivial linear combination of eigenvectors, there will exist more than one eigenvector $\boldsymbol{\alpha}_i$ for which this is not true, and therefore more than one eigenvalue A_i which can be measured with nonzero probability.

This is the famous probabilistic nature of quantum mechanics encapsulated through our postulates. Do there exist such state vectors which are linear combinations of multiple eigenvectors of an observable operator? Certainly yes, by Postulate 1. Yet in such cases multiple eigenvalues can be measured with nonzero probability: that is, multiple values can be obtained for the same measurement. Thus the wording of the postulate– the state vector is forced into a specific eigenvector– is relevant: the state vector in these cases does not possess a single fixed value for an observable which must be revealed upon measurement so it cannot be said, technically, to have a position or momentum or any other observable value. We can only say that an eigenvalue is the measured value of this specific measurement: or, the eigenvector the state vector has been forced into was not necessarily the state vector before at all; in another scenario, with defined probability, the state vector may well have been forced into a different eigenstate and then yielded a different value for an observable.

We end this section with a summary on quantum states and inherent probability:

- In a pure state with respect to a physical observable the state vector is made up solely of one eigenvector of the observable operator (up to some phase factor). A measurement will therefore yield the eigenvalue corresponding to that eigenvector with probability 1. In such cases (a

rarity) a deterministic prediction can be made about the results of a measurement.

- *After* measurement a state vector is forced into one of the constituent pure eigenstates $\boldsymbol{\alpha}_i$ with probability $|(\boldsymbol{\alpha}_i, \Psi)|^2$. Thereon the above determinism of a pure state applies for successive measurements of the same observable unless the system experiences external perturbation which moves it out of the pure eigenstate.

- A state vector is in a mixed state with respect to a physical observable if the state vector is a non-trivial linear combination of more than one eigenvector of the observable operator. In those cases the strongest predictive statement about the result of a measurement is that a specific eigenstate $\boldsymbol{\alpha}_i$ has probability $|(\boldsymbol{\alpha}_i, \Psi)|^2$ of being measured. We cannot make any deterministic predictions at all, and we do not really think of a mixed state as having a value for that specific observable. Most naturally occurring states in quantum mechanics are indeed mixed states.

4.3 Probability Mass Functions

We conclude this chapter on measurements, and indeed conclude the fundamental postulates dealing with the quantum state problem, with a formalisation of how discrete wavefunctions encapsulate probabilities as probability mass functions.

We have already stated that the discrete wavefunction, which we denoted $\psi_{\boldsymbol{\alpha}}(x)$ with a domain of orthonormal eigenvectors, is exactly the function which stores the components corresponding to the eigenvectors we input. We therefore call it a **probability mass function**. This is a formal name for a very simple idea: it stores probabilities of discrete events- here, the event is the state vector being forced into a certain eigenstate by a measurement– and can be extracted as an output of the probability mass function when we input the event (eigenstate). We know that these components are probabilities, because of the measurement postulate and the common expansion we have already proved!

If the discrete wavefunction is a probability mass function, then necessarily the modulus squared of its outputs (the probabilities of the state taking

each pure eigenstate) must sum to exactly 1. Thus as we must get some result for a measurement, the sum of probabilities is the probability that some result will occur, and thus 1. There is an important clarification to make to prove that our formalism works.

Claim: The modulus squared components of a normalised state vector must sum to 1 in a discrete basis.

The importance of this claim is clear, since it is equivalent to the statement that the sums of the different probabilities for all the possible measurements of an observable must sum to 1, which must be true if they are to be considered probabilities in the first place.

Proof:

For some state vector

$$\Psi := \sum_{\{i\}} c_i \boldsymbol{\alpha}_i$$

in some orthonormal basis $\{\boldsymbol{\alpha}_i\}$, we need to prove that

$$\sum_{\{i\}} |(\boldsymbol{\alpha}_i, \Psi)|^2 = 1$$

given that the state vector is normalised. Well we know that

$$(\Psi, \Psi) = 1$$

so we know that

$$\left(\sum_{\{i\}} c_i \boldsymbol{\alpha}_i, \sum_{\{i\}} c_i \boldsymbol{\alpha}_i \right) = 1.$$

Then, by the rudimentary expansion this is

$$\left(\sum_{\{i\}} (\boldsymbol{\alpha}_i, \Psi) \boldsymbol{\alpha}_i, \sum_{\{j\}} (\boldsymbol{\alpha}_j, \Psi) \boldsymbol{\alpha}_j \right) = 1.$$

Due to linear distributivity this means that we get sum terms of the form

$$(\boldsymbol{\alpha}_i, \Psi)^* \, (\boldsymbol{\alpha}_j, \Psi) \, (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j)$$

for some i, j. However, due to the orthonormality of the basis, all terms when $i \neq j$ disappear, so we have

$$\left(\sum_{\{i\}}(\boldsymbol{\alpha}_i, \Psi)\boldsymbol{\alpha}_i, \sum_{\{j\}}(\boldsymbol{\alpha}_j, \Psi)\boldsymbol{\alpha}_j\right) = \sum_{\{i\}} (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i)^* (\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i) = 1.$$

But then this is simply

$$\sum_{\{i\}} |\, (\boldsymbol{\alpha}_i, \Psi)\, |^2 = 1.$$

and our proof is complete.

Thus indeed, we have the result that the square modulus components of the discrete state vector, which is, the square modulus of the outputs of its discrete wavefunctions, are valid probabilities of measurements. In this way, the encapsulation component of the state problem is much better understood. The full bijections are:

- In a basis-independent bijection (by the state postulate),

  Physical States $\leftrightarrow$ State Vector

- In a basis-dependent bijection (by the properties of expansions in vector spaces):

  State Vectors $\leftrightarrow$ Component Maps $\leftrightarrow$ Wavefunctions

- By reinterpreting the coefficients (according to the measurement postulate):

  Wavefunctions $\leftrightarrow$ Probability Mass Functions

4.4 Expecation Values and Ehrenfest's Theorem

With this probabilistic formalism complete, we can now determine physically relevant qualities of a state given some state vector. For example, one of the most important physically relevant values is the **expectation value** of an observable.

The expectation value is simple compared to what we have already discussed thus far. For a discrete random variable X, the expectation of X would be defined and denoted

$$\langle X \rangle := \sum_{i \in \mathcal{I}} x_i P(X = x_i)$$

if $\{x_i\}_{i \in \mathcal{I}}$ are the complete set of possible values for X. So there is no difference at all in the quantum mechanical case, where we treat the value of a measurement for an observable $\mathcal{A}$ on a state Ψ as the random variable in question:

$$\langle A \rangle = \sum_{i \in \mathcal{I}} A_i P(\boldsymbol{\alpha}_i)$$

since A_i is the (eigen)value for the observable $\mathcal{A}$ measured when the state vector is forced into the eigenstate $\boldsymbol{\alpha_i}$ in the event $\Psi \to \boldsymbol{\alpha}_i$. But this is then

$$\langle A \rangle = \sum_{i \in \mathcal{I}} A_i |(\Psi, \boldsymbol{\alpha}_i)|^2.$$

Why is this physical quantity relevant? Well, for one, while quantum states are strange on a quantum level (at small length scales), they may not appear strange on a macroscopic level, where one cannot distinguish between small differences between observable eigenstates.

That is, even though all superpositions are technically possible quantum states, in practice most superpositions will consist of heavy weights around a specific small interval of values. One will essentially never get an electron with 50% chance of being in a coordinate in California and 50% chance of being in some coordinate in Paris. Reality is strange, but not unboundedly so.

Does that mean that quantum phenomena are only valid on small scales? Not at all. Again, the only difference is we might not be able to *notice* them on larger scales. It is still anticlassical for an electron to be in a superposition of many different possible positions, but if these are within the 1/1000th of the radius of an atom, the human eye will detect no difference even if different measurements would yield different eigenstates.

Hence, we can use the expectation value as a substitute for the actual measured value if we are considering a quantum system at a higher level, whether that be within an ensemble of other states, or at a more macroscopic level.

We clearly do not *always* have to be concerned with the exact superposition a state is in.

Now, we introduce the connection between microscopic, quantum scales and macroscopic, classical scales. This is **Ehrenfest's Theorem**, at a very non-rigorous level.

Ehrenfest's Theorem:

The laws of classical physics relating observables to each other hold for their quantum expectation values.

For example, the kinetic energy, denoted here by T, satisfies

$$T = \frac{1}{2}mv^2$$

which is, substituting the momentum p:

$$T = \frac{p^2}{2m}.$$

Then, Ehrenfest's Theorem says that for the Kinetic Energy Operator $\hat{T}$ and the Momentum Operator $\hat{P}$ (whose exact forms we do not know yet),

$$\langle T \rangle = \frac{\langle P \rangle^2}{2m}.$$

And since the quantum expectation is defined regardless of the scale of the system and the initial superposition, this is a *deterministic* statement, not a probabilistic one. We may not know what value the measurement of $\mathcal{A}, \mathcal{B}$ will take, but we know what relation the expectation values $\langle A \rangle, \langle B \rangle$ should take.

This theorem is usually introduced rather late in undergraduate quantum textbooks, because it requires notions we have not reached yet to prove it. Since this is an expository primer, I will not prove it, as the reader can in the future look for a proof themselves, and the proof is not particularly instructive.

However, the theorem is clearly satisfying for us, because it provides a concrete bridge between quantum phenomena and classical phenomena. The

reason why Newton's Principles apply to macroscopic objects like balls and tables is because of Ehrenfest's Theorem: the balls may be in quantum superpositions, but their expected values will follow Newtonian mechanics. And since at a microscopic scale we cannot detect microscopic differences in eigenstates, the expectation value will nearly always be the value we observe.

This is why it took so long for quantum mechanics to be developed. Without developing the experimental sophistication to probe systems at a microscopic level, it is difficult to observe quantum phenomena. This is also why you do not need to wonder if your fridge is in the kitchen or not even though by quantum superposition it may not take a concrete value until you have observed it!

4.5 Basis Dependence of Observable Operators

When we refer to an **observable space**– for example, "position space" or "momentum space", we mean the state space with the eigenbasis of a physical observable set as the fixed basis we are expanding and considering our vectors in.

A large component of problem solving in quantum mechanics consists of being able to apply two concepts related to bases:

1. Staying flexible without committing to a specific base and considering the state vector without a basis first.

2. Switching into a basis which is useful for our specific needs and algebraic manipulations. Finally, calculating relevant physical values of the state in that basis (eg, probabilities, or expectation value).

It is clear, even from the very early section on orthonormality, that not a bases are created equal (orthonormal bases are far cleaner and thus superior to nonorthonormal bases). Some are simply better selected because algebraic manipulation becomes a lot easier when a problem is considered in those bases. Out of such bases, the eigenbases of physical observables are the most significant. Thus we reach this definition of 'observable spaces'. We understand that all these observable spaces are not actually new vector

spaces: they are all this *same* state space we have been studying, but simply with different eigenbases spanning the space and therefore where the same vectors will take different expansions.

All quantum mechanics problems are solved in the end in one of these spaces. Rarely, we might solve one part of a problem in one observable space and then switch bases (more in Chapter 7), after taking what we have learnt from that part, to solve the rest of the problem. Most of the time – especially overwhelmingly in this book – we work in position space where possible. This is not only because it makes sense to work with time and position as the two most important variables, as in our physical reality, but also because there are important variables to consider in the time evolution problem – in particular the potential $V(x)$, a function of position– which are usually easier to express in the position basis than any other. Momentum space is by far the other main counterpart in quantum mechanics, and usually we might work in momentum space in problems where the conditions or potential are easily expressed in terms of the momentum. Only rarely will we see energy space in action, and any other observable space will essentially be inferior to position and momentum space because the observables of position and momentum are far more important than any other.

Now, there is an important conceptual detail to understand. Recall when we first were introduced to hermitian operators; we proved that each operator's action can be uniquely specified in its own eigenbasis if we have its eigenvalues. Consider some vector Ψ in a the state space spanned by the eigenbasis $\{\omega_i\}$ with eigenvalues $\{\lambda_i\}$ of an operator Ω. We have, by the common inner product expansion:

$$\Psi = \sum_{i \in \mathcal{I}} (\omega_i, \Psi)\, \omega_i.$$

Then the action of the operator Ω is

$$\Omega\Psi = \Omega \sum_{i \in \mathcal{I}} (\omega_i, \Psi)\, \omega_i$$

which, as it is a linear operator,

$$\Omega\Psi = \sum_{i \in \mathcal{I}} (\omega_i, \Psi)\, \Omega\omega_i$$

which is,

$$\sum_{i\in\mathcal{I}} (\omega_i, \Psi)\, \lambda_i \omega_i.$$

So quite simply, we get

$$\Omega\Psi = \sum_{i\in\mathcal{I}} (\omega_i, \Psi)\, \lambda_i \omega_i.$$

Of course, it is very clear that these eigenvalues λ_i are not obtained if we are not working in the space spanned by the eigenbasis of Ω. This in turn shows us the very important fact: that the action of the operator– not just the forms of the state vectors– is **basis dependent**.

It is important to understand this idea, just as it is with different wavefunctions, which represent that expansions are different in different bases. So we will often talk about "the momentum operator in position space" or "the position operator in momentum space".

4.6 Summary

In this chapter, we discussed the action of Hilbert space linear operators, and the special property of Hermiticity. We used these to define operators associated with physical observables, and in turn, we used the eigenvectors and eigenvalues of these operators to define physical measurements and pure states. The measurement postulate then taught us how to calculate numerically probabilities associated with the superposition, and we also learnt the famous property of "the collapse of the wavefunction", which is that superpositions are collapsed into pure eigenstates when a measurement is made. Finally, we discussed the link between the microscopic quantum world and the macroscopic classical world, to provide a satisfactory explanation for why quantum phenomena do not occur in everyday life.

4.7 Exercises from Chapter 4∗

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

Chapter 5

Chapter 5: Commutation Relations and Simultaneous States

Another question still remains. We know that state vectors represent states– and that is, physical states. If we think to use our classical intuition, we know a physical state can be thought about from the perspective of one observable (such as when we consider the momentum of two colliding objects). However, that does not mean it does not possess any values for all other physical observables, as it still has energy, angular momentum, and so on: only that we are not currently considering the other observables. Similarly, we would not be particularly pleased if the physical states pertaining to a certain measurement of one specific observable– these so called pure states– suddenly contained absolutely no information on any other observables. This is a question of information encapsulation: how does an energy pure state store extractable information about momentum, for example? To understand this question, which is far more complicated than the classical one of "it's there, and just hasn't been measured yet", we treat the two seminal theorems on the matter.

It turns out that much of this information, on whether a state made hold certain values for both observables represented by a given pair of observable operators– in short, whether it can be a **simultaneous state** for these two observables– turns out to be related intimately with their *commutation*

relation. That is, if we define the **commutator**

$$[\hat{A}, \hat{B}] := \hat{A}\hat{B} - \hat{B}\hat{A}$$

then the specific value of this commutator relates to our question of "simultaneous states".

Here, we explore more consequences of the commutation relation between two arbitrary observable operators. Along the way, we will also be able to practise the mathematical operations we have introduced at a level and intensity we have not had the occasion to do so far.

5.1 Simultaneous states

The question of "simultaneous states" is to investigate which states ($\leftrightarrow$ state vectors) can contain information on multiple observables at a time, and which observables these are. We have seen in the Stern Gerlach experiment that for example x and y spin simultaneous states are impossible, so this is a relevant, fully quantum in nature, problem.

It turns out that the ability for different observables to have information represented in the same state vectors depends strongly on the commutator of their observable operators, as these in turn relate their orthonormal eigenvectors. To see this, there is one sweeping but simple theorem on operators for observables which can be measured simultaneously, and one dramatically anticlassical one for observables which cannot.

5.1.1 The Compatibility Theorem

Consider an unperturbed system, two physical observables, and three measurements ordered chronologically. The first and third measurements are for the first physical observable denoted $\mathcal{A}$, but the second measurement is for the second observable $\mathcal{B}$. We know from the Measurement Postulate that:

- The first measurement forces the wavefunction into some pure eigenstate $\boldsymbol{\alpha_i}$ of the first physical observable operator $\hat{A}$.

- The second measurement forces the wavefunction into some pure eigenstate $\boldsymbol{\beta}_i$ of the second physical observable operator $\hat{B}$.
- If the second measurement of the different observable did not exist, then we would have successive measurements of the same state (which is, the operator acting on the same pure state the starting state vector was forced into following the first measurement) and we would expect with certainty the same measured value of $\mathcal{A}$ which is the eigenvalue A_i associated with the pure state $\boldsymbol{\alpha_i}$.

The question is therefore whether or not this second measurement changes the result of the third. This is a profound question, because if it does, then we would conclude the simple act of measuring the second observable has moved the state vector out of the pure eigenstate it was in after the first measurement; that would then imply the second measurement is in itself a perturbation to the system: a confusing result. Indeed– the reader will recognise that this is exactly the class of behaviour which appeared so surpisingly in the Stern Gerlach experiment.

We define $\mathcal{A}$ and $\mathcal{B}$ to be **compatible observables** iff the first and third measurements yield the same value regardless of the starting state before measurement 1, and regardless of the value of the second observable measured in the second measurement. If we call the values measured $\mathcal{A}^{(1)}$, $\mathcal{B}^{(1)}$, $\mathcal{A}^{(2)}$, then observable $\mathcal{A}$ and $\mathcal{B}$ are compatible iff

$$\forall\, \Psi,\, \forall\, \mathcal{B}^{(1)},\ \ \mathcal{A}^{(1)} = \mathcal{A}^{(2)}.$$

Compatibility Theorem:

The following three conditions all equivalent to each other:

1. $\mathcal{A}$ and $\mathcal{B}$ are compatible observables.
2. $\hat{A}$ and $\hat{B}$ share a common eigenbasis.
3. $\hat{A}$ and $\hat{B}$ commute.

Proof:

First we prove that $\hat{A}$ commutes with $\hat{B}$ iff they possess a common eigenbasis.

Consider two observable operators which commute, and define their eigenbases to be $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{\beta}_i\}$. Now take an arbitrary eigenvector $\boldsymbol{\alpha}_i$ of $\hat{A}$ with eigenvalue A_i. We have

$$\hat{A}\hat{B} = \hat{B}\hat{A}$$

so we get

$$\hat{A}\hat{B}\boldsymbol{\alpha}_i = \hat{B}\hat{A}\boldsymbol{\alpha}_i = \hat{B}A_i\boldsymbol{\alpha}_i.$$

However, we can now pull the constant eigenvalue out:

$$\hat{A}(\hat{B}\boldsymbol{\alpha}_i) = A_i(\hat{B}\boldsymbol{\alpha}_i)$$

so clearly $\hat{B}\boldsymbol{\alpha}_i$ is an eigenvector of $\hat{A}$ corresponding to eigenvalue A_i. Assuming that the eigenvalues are nondegenerate this implies that $\hat{B}(\boldsymbol{\alpha}_i)$ coincides with $\boldsymbol{\alpha}_i$ as the eigenvalue A_i has only one distinguishable eigenvector. The fact that

$$\hat{B}\boldsymbol{\alpha}_i \equiv \boldsymbol{\alpha}_i$$

means we must have

$$\hat{B}\boldsymbol{\alpha}_i = c\boldsymbol{\alpha}_i$$

for some scalar multiple c. This means that $\boldsymbol{\alpha}_i$ is an eigenvector of $\hat{B}$ corresponding to eigenvalue c. So we can say that $\forall i, \boldsymbol{\alpha}_i$ is an eigenvector of $\hat{A}$ and $\hat{B}$: which means that they have the same eigenbasis. This isn't of course, to say, the eigenvalues are the same for $\hat{B}$ and $\hat{A}$ even though it may correspond to the same eigenvector (above, they are not the same unless $A_i = c$). Yet at the same time this is clearly helpful: if we know two physical observable operators commute and we have the eigenbasis of one then we automatically have an eigenbasis of the other.

Now, we prove it the other way around. Assume $\hat{A}$ and $\hat{B}$ both possess the eigenbasis $\{\gamma_i\}$: that is, the vectors $\boldsymbol{\gamma}_i$ are eigenvectors for both $\hat{A}$ and $\hat{B}$. We want to prove they commute. As they possess the same eigenbasis with eigenvalues $\{A_i\}$ and $\{B_i\}$ respectively, we can write

$$\hat{A}\hat{B}\gamma_i = \hat{A}B_i\gamma_i = B_i\hat{A}\gamma_i = B_iA_i\gamma_i$$

and the exact same applies for $\hat{B}\hat{A}\gamma_i$:

$$\hat{B}\hat{A}\gamma_i = \hat{B}A_i\gamma_i = A_i\hat{B}\gamma_i = A_iB_i\gamma_i.$$

Clearly, as A_i and B_i are constant eigenvalues,

$$A_i B_i \equiv B_i A_i.$$

So this easily proves that two observable operators possessing the same eigenbasis must commute. Thus the implication works both ways and therefore two observable operators commute iff they share a common eigenbasis.

Now to look at the practical definition: we are probably more interested in the concept of compatibility, as it concerns whether or not a measurement of a second observable in between measurements of a first observable will alter the measured results from the first measurement, effectively forcing the state vector out of the pure eigenstate it was forced into. Let's first prove that two observables having common operator eigenbases is necessary and sufficient for the above defined definition of compatibility to hold.

Start by considering two observables $\mathcal{A}$ and $\mathcal{B}$ represented by operators $\hat{A}$ and $\hat{B}$ respectively. Define the measurements to be $\mathcal{A}^{(1)}, \mathcal{B}^{(1)}, \mathcal{A}^{(2)}$. For the observables to be compatible we need $\mathcal{A}^{(1)}$ to be the same as $\mathcal{A}^{(2)}$ regardless of the starting state and $\mathcal{B}^{(2)}$. Assume to begin with that the two operators $\hat{A}$ and $\hat{B}$ have the common eigenbasis $\{\gamma_i\}$. By definition the first measurement of $\mathcal{A}$ must force the state vector into a single eigenvector in the eigenbasis of the operator $\hat{A}$: that is, some γ_i such that the measured value is for observable $\mathcal{A}$ the eigenvalue A_i. Next, measurement $B^{(1)}$ is the action of the operator $\hat{B}$ on the eigenvector γ_i. But by the Measurement Postulate of quantum mechanics,

$$P(\boldsymbol{\alpha}_i) = |\,(\boldsymbol{\alpha}_i, \Psi)\,|^2$$

That is, the probability that the arbitrary operator $\hat{A}$ forces the state vector into an arbitrary eigenvector $\boldsymbol{\alpha}_i$ from its eigenbasis. Here, then, since the state vector has been forced into the eigenstate γ_i by the first measurement, the probability the second measurement of the other observable $\mathcal{B}$ forces the state vector into the same eigenstate is:

$$P(\gamma_i) = |\,(\gamma_i, \gamma_i)\,|^2 = 1$$

where we assume as per usual that the eigenvectors γ_i have been normalised. So we can say that measurement B will not alter the eigenstate the state vector is in and therefore the third measurement will follow the same logic

to yield the exact same value, the eigenvalue A_i corresponding to γ_i. Thus, if two observable operators possess the same eigenbasis, they are compatible observables.

If the observables are compatible then this implies their operators have the same eigenbasis. The proof for this is simple. If the observables $\mathcal{A}$ and $\mathcal{B}$ are compatible then for the successive measurements $\mathcal{A}^{(1)}, \mathcal{B}^{(1)}, \mathcal{A}^{(2)}$ the measured values for $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ must be the same. The measurement $\mathcal{A}^{(1)}$ must have forced the wavefunction into an eigenvector of $\hat{A}$, some arbitrary $\boldsymbol{\alpha}_i$. Then, the measurement $\mathcal{B}^{(1)}$ must force the wavefunction into some arbitrary eigenvector $\boldsymbol{\beta}_i$ of the operator $\hat{B}$. However, the final measurement must yield the same result as the first if the observables are compatible, which is, the same eigenvalue corresponding to the same eigenvector $\boldsymbol{\alpha}_i$ of operator $\hat{A}$ as it originally was in. The probability that the measurement forces the wavefunction, currently in the eigenstate $\boldsymbol{\beta}_i$ of $\hat{B}$ as the measurement $\mathcal{B}^{(1)}$ has just been performed, into the same eigenstate $\boldsymbol{\alpha}_i$ as originally measured is:

$$P(\boldsymbol{\alpha}_i) = |\,(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)\,|^2.$$

However, if these observables are to be compatible, the final measurement must with certainty yield the eigenvalue A_i again and therefore the above probability of measurement $\mathcal{A}^{(2)}$ forcing it back into the original eigenstate must be 1. So

$$|\,(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i)\,|^2 = 1 \Rightarrow \;\; \boldsymbol{\alpha}_i \equiv \boldsymbol{\beta}_i$$

and therefore their eigenbases must be the same as the above holds true for any arbitrary $\boldsymbol{\alpha}_i$ and corresponding $\boldsymbol{\beta}_i$ from the measurements.

The Compatibility Theorem is now complete. We have shown that:

- Two operators commuting is necessary and sufficient for them to possess a common eigenbasis.
- Two operators possessing a common eigenbasis is necessary and sufficient for the two observables they represent to be compatible.
- Therefore, two observable operators commuting is also necessary and sufficient for them to represent compatible observables.

The logical implications of these facts all run three ways.

While we have now seen facts about compatible observables, an example of incompatible observables sticks in our mind– that of the Stern Gerlach experiment. We saw exactly that x and y spins were incompatible, because measuring the x spin in between two y measurements stopped the second y measurement from being the same as the first with certainty– which is to say, we now know, that the measurement of x spin forced it out of the eigenstate of y spin it had been previously forced into. All the questions about the quantum state raised by the Stern Gerlach experiment will finally come to an end with this section. We would like to formalise our understanding of how incompatibility affected the experiment. To explain it all, we witness– and prove ourselves– one of Physics' most groundbreaking and shocking theorems, the famous **Heisenberg Uncertainty Principle**.

5.2 The Heisenberg Uncertainty Principle

The idea of commuting observable operators being necessary and sufficient for the two observables they represent to be compatible is a very important one for the question of simultaneous states, and has been shown above. Now we must surely consider when two observable operators do not commute: in other words, when they represent **incompatible** observables. One of the most important and dramatic results of all quantum mechanics, the Heisenberg Uncertainty Principle, results when we carry out some elegant mathematics to investigate this problem. Before we begin the statement and proof, let us define the commutator between two operators to be

$$[\hat{A}, \hat{B}] := \hat{A}\hat{B} - \hat{B}\hat{A}$$

so that if we have two commuting operators $\hat{A}$ and $\hat{B}$, then

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = 0$$

since $\hat{A}\hat{B} = \hat{B}\hat{A}$ iff they commute. For operators which do not commute, their commutator may take a wide variety of forms: which is why it is useful under universal convention to have this shorthand.

Heisenberg Uncertainty Principle

For any state Ψ_t,

$$\Delta A_t \Delta B_t \geq \frac{1}{2} | \left(\Psi_t, [\hat{A}, \hat{B}] \Psi_t \right) |$$

where ΔA_t is the standard deviation of measurable values of observable $\mathcal{A}$ at time t: which is therefore a measure of uncertainty for these variables.

Proof:

We will continue to refer to arbitrary observables $\mathcal{A}$ and $\mathcal{B}$ for the proof; all the proof is relevant at any instant of time and so time subscripts will be eschewed. The notation ΔA refers to the standard deviation of the measurements of observable $\mathcal{A}$; this standard deviation is no different from the statistical definition:

$$\Delta A = \sqrt{\langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2}$$

where the symbol $\langle X \rangle$ is the expected value of the variable X, as seen in the probability preliminary. First we note that this principle is valid for compatible observables: as compatible observables, their operators must commute. Thus

$$[\hat{A}, \hat{B}] = 0 \Rightarrow \ \Delta A_t \Delta B_t \geq \frac{1}{2} | \left(\Psi_t, [\hat{A}, \hat{B}] \Psi_t \right) | = \frac{1}{2} | \left(\Psi_t, 0 \right) | = 0.$$

So for compatible observables,

$$\Delta A_t \Delta B_t \geq 0$$

which is neither interesting nor invalid at all since the standard deviation of any measurement can never be negative. Now, we will prove this for all physical operators, regardless of whether they commute.

Lemma 1:
Any operator $\hat{X}' := \hat{X} - \langle \hat{X} \rangle$ where $\hat{X}$ is a Hermitian physical operator is also Hermitian.

Proof:

Recall that the definition for an expected value of a variable is the sum of its possible values multiplied by the probabilities of the variable taking those values. Therefore, we can say that, over the eigenbasis $\{\xi_i\}$ of $\hat{X}$ with eigenvalues $\{X_i\}$,

$$\langle \hat{X} \rangle = \sum_{\{i\}} P(\xi_i) X_i,$$

but by our knowledge of the previous postulates we can describe the probability more precisely: the measurement postulate defines this to be

$$\langle \hat{X} \rangle = \sum_{\{i\}} X_i |\, (\xi_i, \Psi)\, |^2.$$

Our job is to prove that the operator $\hat{X}' := \hat{X} - \langle \hat{X} \rangle$ is hermitian if $\hat{X}$ is hermitian for all quantum operators. That is, we need to prove that:

$$\left(\Psi_1, \hat{X}'\Psi_2\right) = \left(\hat{X}'\Psi_1, \Psi_2\right)$$

for all Hilbert space functions Ψ_1 and Ψ_2. The operator $\hat{X}$ must be hermitian as $\hat{X}$ is defined to be a quantum operator corresponding to a physical observable. Meanwhile, the expectation value

$$\langle \hat{X} \rangle = \sum_{\{i\}} X_i |\, (\xi_i, \Psi)\, |^2.$$

is clearly a real scalar, as the probabilities, which are square moduli, will all be real numbers and so will each eigenvalue of the hermitian operators. Therefore,

$$\left(\hat{X}\Psi_1, \Psi_2\right) \equiv \left(\Psi_1, \hat{X}\Psi_2\right)$$

and

$$\left(\langle \hat{X} \rangle \Psi_1, \Psi_2\right) \equiv \left(\Psi_1, \langle \hat{X} \rangle \Psi_2\right) \equiv \langle \hat{X} \rangle \, (\Psi_1, \Psi_2)$$

so for any physical operator $\hat{X}$ the defined operator $\hat{X}'$ is the sum of two hermitian operators. So

$$\begin{aligned} \left(\hat{X}'\Psi_1, \Psi_2\right) &= \left([\hat{X} - \langle \hat{X} \rangle]\Psi_1, \Psi_2\right) = \left(\hat{X}\Psi_1, \Psi_2\right) - \left(\langle \hat{X} \rangle \Psi_1, \Psi_2\right) \\ &= \left(\Psi_1, \hat{X}\Psi_2\right) - \left(\Psi_1, \langle \hat{X} \rangle \Psi_2\right) \\ &= \left(\Psi_1, \hat{X}'\Psi_2\right) \end{aligned}$$

using the linear properties of the inner product. Thus, the operator $\hat{X}'$ is Hermitian for any physical operator. Therefore, defining $\hat{A}' := \hat{A} - \langle\hat{A}\rangle$ and $\hat{B}' := \hat{B} - \langle\hat{B}\rangle$ for the purpose of the problem also gives us two hermitian operators. $\square$

The commutator in the generalised principle might give pause with regards to the development of these new operators, but, importantly,

$$[\hat{A}', \hat{B}'] = [\hat{A}, \hat{B}].$$

This fact can be proved quite simply:

$$\begin{aligned}[\hat{A}', \hat{B}'] &= \hat{A}'\hat{B}' - \hat{A}'\hat{B}' \\ &= (\hat{A} - \langle\hat{A}\rangle)(\hat{B} - \langle\hat{B}\rangle) - (\hat{B} - \langle\hat{B}\rangle)(\hat{A} - \langle\hat{A}\rangle) \\ &= (\hat{A}\hat{B} - \hat{A}\langle\hat{B}\rangle - \langle\hat{A}\rangle\hat{B} - \langle\hat{A}\rangle\langle\hat{B}\rangle) - (\hat{B}\hat{A} - \hat{B}\langle\hat{A}\rangle - \langle\hat{B}\rangle\hat{A} - \langle\hat{B}\rangle\langle\hat{A}\rangle)\end{aligned}$$

but as the expectation values $\langle\hat{A}\rangle$ and $\langle\hat{B}\rangle$ are real scalars it is clear that $\langle\hat{A}\rangle\langle\hat{B}\rangle = \langle\hat{B}\rangle\langle\hat{A}\rangle$, and $\langle\hat{A}\rangle\hat{B} = \hat{B}\langle\hat{A}\rangle$ and vice versa swapping the A and B around. So the terms cancel out and we are left with

$$[\hat{A}', \hat{B}'] = \hat{A}\hat{B} - \hat{B}\hat{A} := [\hat{A}, \hat{B}]. \ \square$$

Now, one last important lemma:

Lemma 2:

$$\left(\hat{A}'\Psi, \hat{A}'\Psi\right) = (\Delta\hat{A})^2$$

Proof:
By the Hermiticity of $\hat{A}'$,

$$\left(\hat{A}'\Psi, \hat{A}'\Psi\right) = \left(\Psi, ([\hat{A}']^2\Psi\right).$$

Expanding the definition,

$$\begin{aligned}
\left(\hat{A}'\Psi, \hat{A}'\Psi\right) &= \left(\Psi, \hat{A}'^2\Psi\right) \\
&= \left(\Psi, [\hat{A} - \langle\hat{A}\rangle][\hat{A} - \langle\hat{A}\rangle]\Psi\right) \\
&= \left(\Psi, [\hat{A}^2]\Psi - 2\langle\hat{A}\rangle\hat{A}\Psi + \langle\hat{A}\rangle^2\Psi\right) \\
&= \left(\Psi, [\hat{A}^2]\Psi\right) - 2\langle\hat{A}\rangle\left(\Psi, \hat{A}\Psi\right) + \langle\hat{A}\rangle^2\left(\Psi, \Psi\right) \\
&= \langle\hat{A}^2\rangle\left(\Psi, \Psi\right) - 2\langle\hat{A}\rangle\langle\hat{A}\rangle\left(\Psi, \Psi\right) + \langle\hat{A}\rangle^2\left(\Psi, \Psi\right) \\
&= \langle\hat{A}^2\rangle - 2\langle\hat{A}\rangle\langle\hat{A}\rangle + \langle\hat{A}\rangle^2 \\
&= \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2 \\
&= (\Delta\hat{A})^2 \quad \square
\end{aligned}$$

Now we can use these lemmas to prove the problem. We want to prove that

$$\Delta A \Delta B \geq \frac{1}{2}|\left(\Psi, [\hat{A}, \hat{B}]\Psi\right)|$$

at all times t. We start by replacing $[\hat{A}, \hat{B}]$ with $[\hat{A}', \hat{B}']$. Then, we have,

$$\left(\Psi, [\hat{A}, \hat{B}]\Psi\right) = \left(\Psi, [\hat{A}', \hat{B}']\Psi\right) = \left(\Psi, [\hat{A}'\hat{B}' - \hat{B}'\hat{A}']\Psi\right).$$

This is,

$$\left(\Psi, [\hat{A}, \hat{B}]\Psi\right) = \left(\Psi, \hat{A}'\hat{B}'\Psi\right) - \left(\Psi, \hat{B}'\hat{A}'\Psi\right).$$

We can rearrange this by the hermiticity of $\hat{A}'$ and $\hat{B}'$:

$$\left(\Psi, [\hat{A}, \hat{B}]\Psi\right) = \left(\hat{A}'\Psi, \hat{B}'\Psi\right) - \left(\hat{B}'\Psi, \hat{A}'\Psi\right) = \left(\hat{A}'\Psi, \hat{B}'\Psi\right) - \left(\hat{A}'\Psi, \hat{B}'\Psi\right)^*$$

so this is

$$\left(\Psi, [\hat{A}, \hat{B}]\Psi\right) = 2i\mathrm{Im}\left(\left(\hat{A}'\Psi, \hat{B}'\Psi\right)\right)$$

according to rudimentary arithmetic of complex numbers. Then, the expression we need is

$$\frac{1}{2}|\left(\Psi, [\hat{A}, \hat{B}]\Psi\right)| \leq \frac{1}{2} \times 2|\left(\hat{A}'\Psi, \hat{B}'\Psi\right)| = |\left(\hat{A}'\Psi, \hat{B}'\Psi\right)|.$$

This is because of the above expression for $\left(\Psi, [\hat{A}, \hat{B}]\Psi\right)$ and the fact that the modulus of the imaginary part of a scalar cannot be greater than the modulus of the scalar (Exercise 1.3.2a). Then, by Lemma 2

$$\left(\hat{A}'\Psi, \hat{A}'\Psi\right) = (\Delta\hat{A})^2 \Rightarrow \ \Delta\hat{A} = \sqrt{\left(\hat{A}'\Psi, \hat{A}'\Psi\right)}.$$

So

$$\Delta\hat{A}\Delta\hat{B} = \sqrt{\left(\hat{A}'\Psi, \hat{A}'\Psi\right)}\sqrt{\left(\hat{B}'\Psi, \hat{B}'\Psi\right)}.$$

By Cauchy-Schwartz,

$$\sqrt{\left(\hat{A}'\Psi, \hat{A}'\Psi\right)}\sqrt{\left(\hat{B}'\Psi, \hat{B}'\Psi\right)} \geq |\left(\hat{A}'\Psi, \hat{B}'\Psi\right)|$$

and so, conclusively,

$$\Delta\hat{A}\Delta\hat{B} = \sqrt{(\hat{A}'\Psi, \hat{A}'\Psi)}\sqrt{(\hat{B}'\Psi, \hat{B}'\Psi)} \geq |\left(\hat{A}'\Psi, \hat{B}'\Psi\right)| \geq \frac{1}{2}|\left(\Psi, [\hat{A}, \hat{B}]\Psi\right)|$$

so

$$\Delta\hat{A}\Delta\hat{B} \geq \frac{1}{2}|\left(\Psi, [\hat{A}, \hat{B}]\Psi\right)|.$$

This proves Heisenberg's Uncertainty Principle. □

This general form we have above is still difficult to interpret, but if we consider a few examples we will realise this is a very important result. One of the most famous iterations comes with considering simply the two central operators of quantum mechanics: the position and momentum operators, which we have not yet introduced but will for now just use for calculation purposes. We can calculate the commutator:

$$\begin{aligned}
& [\hat{X}, \hat{P}] = \hat{X}\hat{P} - \hat{P}\hat{X} \\
\Rightarrow\ & [\hat{X}, \hat{P}]\Psi(x) = -xi\hbar\frac{\partial}{\partial x}\Psi(x) - -i\hbar\frac{\partial}{\partial x}\left(x\Psi(x)\right) \\
\Rightarrow\ & [\hat{X}, \hat{P}]\Psi(x) = -i\hbar x\frac{\partial\Psi}{\partial x} - -i\hbar\left[\frac{dx}{dx}\Psi(x) + x\frac{\partial\Psi}{\partial x}\right] \\
\Rightarrow\ & [\hat{X}, \hat{P}]\Psi(x) = -i\hbar\left[x\frac{\partial\Psi}{\partial x} - \Psi(x) - x\frac{\partial\Psi}{\partial x}\right] \\
\Rightarrow\ & [\hat{X}, \hat{P}]\Psi(x) = i\hbar\Psi(x) \\
\Rightarrow\ & [\hat{X}, \hat{P}] \equiv i\hbar.
\end{aligned}$$

After this, if we plug this into the Generalised Uncertainty principle and assume that the wavefunction is normalised,

$$\Delta\hat{X}\Delta\hat{P} \geq \frac{1}{2}|\left(\Psi, i\hbar\Psi\right)| \Rightarrow \Delta\hat{X}\Delta\hat{P} \geq \frac{1}{2}|i\hbar\left(\Psi, \Psi\right)| = \frac{1}{2}|i\hbar| = \frac{1}{2}\sqrt{-i\hbar \times i\hbar} = \frac{\hbar}{2}.$$

The key end result is that

$$\Delta\hat{X}\Delta\hat{P} \geq \frac{\hbar}{2}.$$

This is the most well known case of the Uncertainty Principle in action, but we can see that the Generalised Uncertainty Principle can be applied more broadly than just to the two observables of position and momentum.

Returning to our considerations of the physical results of trying to measure two incompatible observables, it is clear how bizarre this result is. Consider if we have just made a measurement for the position of a particle. Then we have forced its wavefunction into a position eigenstate and therefore we can say that the uncertainty in the position is now 0: we know the successive measurement must yield the same position value with probability 1. However, if we plug in $\Delta\hat{X}$ into the Uncertainty Principle we get

$$0 \times \Delta\hat{P} \geq \frac{\hbar}{2}$$

which implies somehow that the uncertainty in momentum must be infinite!

Thus, Heisenberg's Uncertainty Principle tells us that if we know the value of the position with certainty we are completely have infinite uncertainty in momentum. The relationship works both ways so the same applies for the momentum: if we know the momentum of a particle then we necessarily have infinite uncertainty in the position of the particle and we have not a clue where it is. This is undoubtedly one of the most anti-classical results in quantum mechanics, and yet it results beautifully from the mathematics we have defined (and has never been experimentally refuted). If nothing else, it should now be clear that the mathematical manipulations of quantum mechanics are rich and impactful.

The same is manifested, of course, in the Stern-Gerlach experiment. By knowing x spin, we had infinite uncertainty in y spin- with absolutely no way to tell if an electron would be up or down spin. By knowing the y spin, we had infinite uncertainty in the x spin. This is one example of an experimental verification of the Heisenberg Uncertainty Principle.

5.3 Formulation of Observable Operators

We have discussed in Chapter 4 the fact that the specific functional form of an operator, just like the specific expansion of a state vector, depends on the basis we are working on. Nevertheless, we would like to ask the question of how to form the observable operator for a given physical observable. Otherwise, our discussions above on Compatibility and Incompatibility, specifically via the study of commutation relations, would be completely useless, as we would never be able to specify the commutator at all!

5.3.1 Dirac's Canonical Commutation Relation

Very early on in classical mechanics, it was already established that two observables, position, x and momentum, p, are enough to completely specify the state of a single body. We will not concern ourselves with why that is true, but the same logic carries over to quantum mechanics. Position and momentum are, it turns out, the only two operators we need to generate any new observable operator, at least for observables which have classical counterparts.

The form of the position and momentum operators, in position space, are:

$$\hat{X}\psi(x) = x\psi(x)$$

and

$$\hat{P}\psi(x) = -i\hbar\frac{\partial}{\partial x}\psi(x)$$

There are a few things to note with these forms:

1. Note that we use $\psi(x)$ instead of the usual notation for the general state vector, Ψ. This is because we are working in position space, so the position/momentum operators in the position space must act on the position space expansion $\psi(x)$ of Ψ.

2. The position space expansion, of course, can be written as $\psi(x)$ since we explicitly defined it as the map from basis vectors to the components of the state vector in those basis directions. Thus, everything depends on an input eigenstate of $\hat{X}$; hence, we use x to denote this.

3. These forms would not be the same in, for example, momentum space! The reason why we use the position space form is because, as aforementioned, the majority of quantum problems are solved in position space.

How do we know these forms? Are they postulates? The answer is that they are *derived* from a postulate. The postulate which really generates these forms is Dirac's Canonical Commutation Relation:

Dirac's Canonical Commutation Relation

$$[\hat{X}, \hat{P}] = i\hbar$$

where $i\hbar$ is the notation for the scalar $i\hbar$ multiplying the identity operator.

We already quoted this commutation relation, when we plugged it into the Heisenberg Uncertainty Principle. Here, the reader understands that it is a postulate. However, this simple postulate remarkably generates the forms of nearly all observable operators!

Indeed, from this seemingly small postulate, we can *deduce* the position space (and momentum space) forms of the position operator $\hat{X}$ and $\hat{P}$ shown above. The full derivation will come in Chapter 8, since we require the tools developed there to handle continuous observables, but for now we can take the position space form of the operators for granted.

With this in mind, we can introduce the next Postulate, which tells us how to form the observable operator of any observable, within a given basis.

5.3.2 Formation of Observable Operators

Postulate 4: The Formation of Observable Operators

The two canonical operators in quantum mechanics, expressed in position space, are the position operator:

$$\hat{X}\psi(x) = x$$

and the momentum operator:

$$\hat{P}\psi(x) = -i\hbar\frac{\partial}{\partial x}\psi(x).$$

which can be both shown to be Hermitian. To form any other physical observable operator, express the classical observable in terms of the classical variables of x (position) and p (momentum), and replace the x terms with the operator $\hat{X}$ and the p terms with the operator $\hat{P}$.

Note that the position operator and momentum operator are here one dimensional: if instead we were considering position and momentum in three directions then all we would do would be to replace the x with the y and z variables for the y and z directional operators instead. Generalising up physical dimensions (as in, moving from space in one dimension to space in three dimensions) is only marginally more complicated because we have more terms to consider, and otherwise this rationale stays the same and everything is straightforward. Let us, to punch this in, consider the kinetic energy operator, which will become very useful. Note again that everything is happening in position space, or things would look quite different in another space with a different eigenbasis.

Kinetic Energy Operator

The classical formula for kinetic energy is of course

$$KE = \frac{1}{2}mv^2.$$

We want this in terms of position and momentum. It is important that mass is treated as a constant so does not interfere in this as a separate observable.

The kinetic energy formula expressed in momentum and position is

$$KE = \frac{1}{2}mv^2 = \left(\frac{m}{2m}\right)mv^2 = \frac{m^2v^2}{2m}$$

which is,

$$KE = \frac{p^2}{2m}$$

where position does not appear in this particular observable expression, which is completely fine. Then, by the postulate, all we need to do is to replace the classical variable p with the momentum operator:

$$KE = \frac{p^2}{2m} \leftrightarrow \hat{KE} = \frac{\hat{P}^2}{2m}.$$

The notation $\hat{KE}$ looks foolish and so convention uses $\hat{T}$ instead. The expression $\hat{P}^2$, we know, means applying the same momentum operator twice.

Therefore, in position space, where the momentum operator is

$$\hat{P}\psi(x) = -i\hbar\frac{\partial}{\partial x}\psi(x),$$

we get

$$\hat{T}\psi(x) = \frac{\hat{P}^2}{2m}\psi(x) = -\frac{i\hbar}{2m}\left(\frac{\partial}{\partial x}\left(-i\hbar\frac{\partial\psi(x)}{\partial x}\right)\right)$$

which is,

$$\hat{T}\psi(x) = \frac{(i\hbar)^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2}.$$

Thus we are done, and have the position space form of the kinetic operator:

$$\hat{T}\psi(x) = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2}.$$

5.3.3 Mass and Time

There is a final note to make, on mass and time. Both of these are not treated as observables in quantum mechanics. This makes intuitive sense for time: in experiments we would measure the duration of an event, but this is first of all not something which pertains to the state problem but more

importantly not something we expect to involve probability. The time at which we are measuring the time obviously should not be probabilistic; it is therefore treated more like an underlying variable just like the space of the universe.

With mass, the complete justification is more nuanced. It turns out that particle physics shows the existence of so-called "fundamental particles", and that everything is made up of these particles. Then, we create a notion of mass simply based on the "quantum number" of those particles: which turns out to be impossible to change, hence the idea of being 'fundamental'. Since mass is built up concretely from this bedrock of fundamental properties, then, it is not subject to the probabilistic superpositions and uncertainties of other observables. Hence, it may also be treated as a constant in a given stationary equation.

The best thing to do, then, is to treat m as a constant and t as a varying but non-probabilistic parameter. In both cases, they are not really observables and thus do not require an observable operator to be linked to them.

5.4 Summary

In this chapter, we finished addressing all the uniquely quantum behaviour the Stern-Gerlach experiment threw at us. We proved the incredible Heisenberg Uncertainty Principle, as well as the less explosive but equally insightful Compatibility Theorem, and showed that the mathematical manipulations of quantum mechanics, specifically herein with respect to the commutation relations between two observable operators, can really yield very rich conceptual and algebraic connections. Finally, we also explicitly showed how to construct observable operators without needing to postulate the form of every single one, even though this story is still to be completed more carefully in Chapter 7 and Chapter 8.

Chapter 6

Chapter 6: Time Evolution and Problem Solving Techniques

We will complete our study of the postulates of quantum mechanics in this chapter, which will be less cumbersome than the preceding two chapters where we had to deal with abstract mathematical theory in vector spaces and operators. This will provide quantum mechanics' answer to the other major problem of a Physical Model: the Time-Evolution Problem.

The quantum time-evolution problem will prove to be a double-edged challenge. On one hand, the time evolution postulate is completely trivial conceptually, especially compared to the state vector, observable and measurement postulates, so the reader can breathe a sigh of relief and know they won't have to deal with learning a new field of mathematics just to tackle this problem. All it will include, indeed, is learning one equation: the all-famous Schrödinger Equation postulated by Erwin Schrödinger, and this equation will not be difficult to understand.

On the other hand, the time evolution problem has by far the most potential to be complex, and in many advanced cases, *impossible* to solve without new intensely demanding mathematical techniques for manipulation and indeed often (though still very accurate) approximation. This should not be disheartening, however. The great physicists of the 20th century were not immortalised for only reaching the depth of an introductory book like this one. The Schrödinger Equation is still always valid, even when it is difficult, or impossible, for us to practically solve the actual complete solutions of it.

6.1 Time Evolution and Schrödinger's Equation

Now that we have established the stationary properties of quantum states, observables and measurements, we are done with the quantum mechanical state problem. The second paramount question of Physics is the question of time evolution. One might be relieved to find that, theoretically speaking, the time evolution problem is much simpler and will not require us to do so much complex postulating as the state problem did. In fact, we only need one more postulate to introduce time evolution; this is the famous Schrödinger Equation (which is a postulate, not a derivation!), which will be important to quantum mechanics much similar to the way $F = ma$ is ubiquitous in classical mechanics.

Postulate 5: Schrödinger Equation and the Hamiltonian

In quantum mechanics, there exists the Hamiltonian operator, written $\hat{H}$, which corresponds to the total energy of the system. It is also hermitian, and it plays an integral role in the time-evolution equation in quantum mechanics, the Schödinger Equation:

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

with the Hamiltonian Operator $\hat{H}$ defined as

$$H = \frac{\hat{P}^2}{2m} + V(\hat{X})$$

for some function V of the position operator $\hat{X}$ which is called the potential. This equation determines how the wavefunction will evolve in time provided there are no perturbations to the system.

Note that we have moved from Ψ_t to $\Psi(t)$ which is a better shorthand notation now that we are not discussing stationary states. This function notation doesn't clash with the fact that the state vector is a state vector: it just means that the input is a time value and the output is the state vector corresponding to the state at that time.

Now, we once more start by listing the assertions of this postulate for clarity.

P5A1. This equation is *basis-independent*, because the derivative which appears is with respect to time, which is a parameter of the wavefunction regardless of which basis we are working in.

For example, in position space, the Schrödinger Equation is:

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

after substituting the position space forms

$$\hat{X} = x,\ \ \hat{P} = -i\hbar\frac{\partial}{\partial x},\ \ \Psi(t) = \psi(x,t)$$

Note that

P5A2. There is a total energy, regardless of the specific qualities of the potential, which is integral to quantum mechanics. Clearly it is a physical observable since it is represented by a hermitian operator, the Hamiltonian.

P5A3. The Hamiltonian operator is the operator form of the classical Hamiltonian- that is, the quantum mechanical version of

$$\mathscr{H}(x,p) = \frac{p^2}{2m} + V(x).$$

which is obtained merely by replacing the classical p with $\hat{P}$ and the function $V(x)$ of the classical x, called the potential, with the same function V but of $\hat{X}$.

P5A4. The potential $V(x)$ may not be familiar to a reader who has not studied classical mechanics. However, note that the kinetic energy operator

$$\hat{T} = \hat{H} - V(\hat{X}) = \frac{\hat{P}^2}{2m}$$

is a very rigid term. Therefore, it is really the potential which makes $\hat{H}$ a complicated operator.

The potential make take a variety of forms, depending on the setting of our quantum system which we are trying to solve. Some examples might be:

(a) The free potential (no potential at all):

$$V(x) \equiv 0 \implies V(\hat{X}) \equiv 0 \implies \hat{H} = \frac{\hat{P}^2}{2m}$$

(b) A linear potential,

$$V(x) = -\gamma x \implies V(\hat{X}) = -\gamma \hat{X} \implies \hat{H} = \frac{\hat{P}^2}{2m} - \gamma \hat{X}$$

for some real number γ.

(c) The oscillator potential:

$$V(x) = \frac{1}{2} m\omega^2 x^2 \implies V(\hat{X}) = \frac{1}{2} m\omega^2 \hat{X}^2$$

Note that we are showing explicitly that these potentials come from classical mechanics by transforming them into quantum potentials with the substitution $x \to \hat{X}$.

It is hard to provide in short words an intuition for what the potential is, but for our purposes it can be thought of as defining the "backdrop" of our problem– almost a "shape" our solution must fit into– given its spacial dependence, which stretches through all values of position x.

Importantly, it is the form of this potential, which can be very exotic, which determines whether our Hamiltonian will be *solvable*, that is, if we can determine its eigenstates analytically. It turns out that most complicated potentials such as

$$V(\hat{X}) = X^4$$

make the Hamiltonian unsolvable.

P5A5. Since we say that this Hamiltonian operator exists, there must be eigenvalues (also called eigenenergies for obvious reasons) which are the possible measured values of energy and corresponding eigenvectors of the Hamiltonian- or, energy eigenstates. In fact we often get that the eigenvalues are discretely distributed for the Hamiltonian operator: which means energy is quantised. Bohr's famous electron model results from this energy quantisation.

Finding the eigenstates of the Hamiltonian is sufficient to solve Schrödinger's Equation. We will see why below.

P5A6. Given a state vector at time 0 it evolves in a completely deterministic way. This is surely a great relief. The state may not be deterministic– in which case it is a mixed state for which the strongest predictive statements which can be made are those detailed in Postulate 3. However, it will evolve into a new state vector in a predictable way. That is not to say at all that after the evolution it will not still be in a mixed state, as the new state it has evolved over time into may well still be a superposition of eigenstates. It is just to say that we can determine future state vectors (not necessarily measurements) well given the starting one and the Hamiltonian for the given system.

For all quantum mechanics problems, solving the Schrödinger equation is the most difficult part of the problem. The reason for this is that the Hamiltonian operator is the only major operator which will change depending on the conditions of the problem. The position operator, momentum operator, spin operators all remain the same, but the Hamiltonian is subject to great variation and even variation over time if the potential of the system is varying over time. This means that for different physical problems we always have to go through the considerable difficulty of finding the form of the Hamiltonian given the different conditions, and then solving the eigenvalue equation for that Hamiltonian, which is rarely not incredibly difficult. With all that said, this next section details why such painstaking effort is worth it: with the energy eigenvalue, or Hamiltonian eigenvalue, equation solved, Schrödinger's Equation is also immediately solved.

6.1.1 Solving with Energy Eigenstates

This section will require a difficult mixing of the rules we have thus far learnt and to be able to follow the developments we make will be tantamount to truly consolidating our grasps on the postulates and mathematics up to this point. It also gives many procedural insights into how one should approach problems (not just subproblems, but full problems) in quantum mechanics, by giving the most elementary way to solve the Schrödinger equation. This method is through considering **energy eigenstates**: that is, the eigenstates of the Hamiltonian operator and the state vectors in the orthonormal basis they form which spans the Hilbert space.

To start recall that for any state vector in the state space and orthonormal basis $\{\boldsymbol{\alpha}_i\}$ the state vector can be expressed in terms of how it acts on those eigenvectors:

$$\Psi_t = \sum_{i=1}^{k} (\boldsymbol{\alpha}_i, \Psi_t)\, \boldsymbol{\alpha}_i.$$

where k is the dimensionality of the space: in the state space, we would sum to infinity. There are infinite orthornormal bases which can we can choose to span the state space, but we have already seen that, for the action of an operator on the state vector, considering that state vector as a combination in the form above but with eigenvectors from the eigenbasis of that operator is natural and fruitful because we get simplifications involving eigenvalues, which also have clearer physical meaning. Now the Schrödinger Equation clearly puts the Hamiltonian to the forefront of our focus, and therefore we might like to consider the state vector Ψ when it is expressed in the energy eigenbasis. The Schrödinger Equation states that

$$i\hbar \frac{\partial \Psi(t)}{\partial t} = \hat{H}\Psi(t).$$

If we take the eigenbasis of the Hamiltonian to be $\{\boldsymbol{\varepsilon}_n\}$ and the corresponding energy eigenvalues to be $\{E_n\}$ then the state vector can be expressed as

$$\Psi_t(x) = \sum_{n} (\boldsymbol{\varepsilon}_n, \Psi_t)\, \boldsymbol{\varepsilon}_n$$

and the Schrödinger Equation now takes the form

$$i\hbar \frac{\partial}{\partial t} \sum_{n} (\boldsymbol{\varepsilon}_n, \Psi_t)\, \boldsymbol{\varepsilon}_n = \hat{H} \sum_{n} (\boldsymbol{\varepsilon}_n, \Psi_t)\, \boldsymbol{\varepsilon}_n.$$

Continue by using the linear distributivity of Hermitian operators (fact H4). This tells us that the right hand side can be written

$$\hat{H}\sum_n (\boldsymbol{\varepsilon}_n, \Psi_t)\,\boldsymbol{\varepsilon}_n = \sum_n (\boldsymbol{\varepsilon}_n, \Psi_t)\,\hat{H}\boldsymbol{\varepsilon}_n = \sum_n E_n\,(\boldsymbol{\varepsilon}_n, \Psi_t)\,\boldsymbol{\varepsilon}_n.$$

Then consider the time derivative of the quantities $(\boldsymbol{\varepsilon}_n, \Psi_t)$. The eigenvectors of the Hamiltonian in the state space must be independent of time, presuming the Hamiltonian itself isn't varying over time (cases with time-varying Hamiltonians are far trickier to solve and thus will not be considered in this book). Therefore, they have 0 time derivative, which means we have

$$\frac{\partial}{\partial t}(\boldsymbol{\varepsilon}_n, \Psi_t) = \left(\boldsymbol{\varepsilon}_n, \frac{\partial}{\partial t}\Psi_t\right).$$

One can be assured of this fact by explicitly writing out the summation form of the inner product. Next, by rearrangement of Schrödinger's Equation,

$$\frac{\partial}{\partial t}\Psi_t = -\frac{i}{\hbar}\hat{H}\Psi_t$$

so we can substitute this into the above expression:

$$\frac{\partial}{\partial t}(\boldsymbol{\varepsilon}_n, \Psi_t) = \left(\boldsymbol{\varepsilon}_n, -\frac{i}{\hbar}\hat{H}\Psi_t\right) = -\frac{i}{\hbar}\left(\boldsymbol{\varepsilon}_n, \hat{H}\Psi_t\right).$$

Then, substituting the energy eigenbasis expression of Ψ_t,

$$\begin{aligned}\frac{\partial}{\partial t}(\boldsymbol{\varepsilon}_n, \Psi_t) &= -\frac{i}{\hbar}\left(\boldsymbol{\varepsilon}_n, \hat{H}\sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)\,\boldsymbol{\varepsilon}_m\right) = -\frac{i}{\hbar}\left(\hat{H}\boldsymbol{\varepsilon}_n, \sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)\,\boldsymbol{\varepsilon}_m\right)\\ &= -\frac{i}{\hbar}\left(E_n\boldsymbol{\varepsilon}_n, \sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)\,\boldsymbol{\varepsilon}_m\right) = -\frac{i}{\hbar}E_n^*\left(\boldsymbol{\varepsilon}_n, \sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)\,\boldsymbol{\varepsilon}_m\right)\\ &= -\frac{i}{\hbar}E_n\left(\boldsymbol{\varepsilon}_n, \sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)\,\boldsymbol{\varepsilon}_m\right)\end{aligned}$$

where the fact that $\hat{H}$ is hermitian is used in the algebraic manipulations. Then, as the inner product is linearly distributive across the sum term, this becomes

$$\frac{\partial}{\partial t}(\boldsymbol{\varepsilon}_n, \Psi_t) = -\frac{i}{\hbar}E_n\sum_m (\boldsymbol{\varepsilon}_n, (\boldsymbol{\varepsilon}_m, \Psi_t)\,\boldsymbol{\varepsilon}_m)$$

and as the inner product $(\boldsymbol{\varepsilon}_m, \Psi_t)$ is some constant for fixed m and t, we can pull it out to write the expression as

$$\frac{\partial}{\partial t}(\boldsymbol{\varepsilon}_n, \Psi_t) = -\frac{i}{\hbar} E_n \sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_m) = -\frac{i}{\hbar} E_n \sum_m (\boldsymbol{\varepsilon}_m, \Psi_t)\delta_{nm},$$

which is,

$$\frac{\partial}{\partial t}(\boldsymbol{\varepsilon}_n, \Psi_t) = -\frac{i}{\hbar} E_n (\boldsymbol{\varepsilon}_n, \Psi_t)$$

as the Kronecker delta resulting from the orthogonality of the eigenvectors cancels out all other sum terms except for when the index m matches up with n. This is clearly equivalent to the differential equation

$$\frac{\partial y}{\partial t} = ky$$

which has general solution $x = Ce^{kt}$ for some constant of integration C. One might consider that the inner product is a constant and therefore not a traditional function one might find in differential equations of this form, but we recall that constants can be seen as functions of the form $f(x) \equiv c$. Substituting $x := (\boldsymbol{\varepsilon}_n, \Psi_t)$ and $k := -\frac{i}{\hbar} E_n$ analogously leaves us with the solution

$$(\boldsymbol{\varepsilon}_n, \Psi_t) = Ce^{-\frac{iE_n t}{\hbar}}.$$

The final step is to realise the constant C is not random: at $t = 0$ we should have $(\boldsymbol{\varepsilon}_n, \Psi_0) = C$, which implies that the constant is $C = (\boldsymbol{\varepsilon}_n, \Psi_0)$. Thus we conclude that the rule for time-evolution is

$$(\boldsymbol{\varepsilon}_n, \Psi_t) = (\boldsymbol{\varepsilon}_n, \Psi_0)\, e^{-\frac{iE_n t}{\hbar}} \quad \square$$

All the above suggest the critical importance of the energy eigenstates. Finding these energy eigenstates essentially amounts to solving the SE! So we therefore call the eigenvalue equation of the Hamiltonian its own name: the **time-independent Schrödinger Equation (TISE)**]:

$$\hat{H}\Psi_n = \left(\frac{\hat{P}^2}{2m} + V(\hat{X})\right)\Psi_n = E_n \Psi_n$$

for the energy eigenstates Ψ_n. We will continue to refer to this as the TISE, and we will refer to the original Schrödinger Equation

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Upsilon$$

as the time-dependent Schrödinger Equation, as it is time-dependent.

Many consequences for the solution of Schrödinger's Equation derive themselves promptly. We list them in the taxonomical format again.

SE1. We first note what the fact above actually means. By the rule S8, the term $(\boldsymbol{\varepsilon}_n, \Psi_t)$ is the component of Ψ_t in the eigenbasis $\{\boldsymbol{\varepsilon}_i\}$ corresponding to eigenvector $\boldsymbol{\varepsilon}_n$:

$$\Psi_t = \sum_n (\boldsymbol{\varepsilon}_n, \Psi_t)\, \boldsymbol{\varepsilon}_n.$$

The eigenvector $\boldsymbol{\varepsilon}_n$ itself does not evolve with time. Therefore, by determining how the component $(\boldsymbol{\varepsilon}_n, \Psi_t)$ evolves with time, we get:

$$\Psi_t = \sum_n (\boldsymbol{\varepsilon}_n, \Psi_t)\, \boldsymbol{\varepsilon}_n = \sum_n (\boldsymbol{\varepsilon}_n, \Psi_0)\, e^{-\frac{iE_n t}{\hbar}} \boldsymbol{\varepsilon}_n.$$

We cannot fall for the common deception of thinking we can pull out the term $e^{-iE_n t/\hbar}$ from the sum, since the eigenvalues $\{E_n\}$ corresponding to the eigenvectors $\{\boldsymbol{\varepsilon}_n\}$ change for each n so it is not a constant. However, we do know now that, if we can determine the eigenvectors of the Hamiltonian for a given system, and the corresponding eigenvalues, and know the initial state (and therefore the components $(\boldsymbol{\varepsilon}_n, \Psi_0)$ of the initial state vector) then we have a fully defined state Ψ_t for any t as we can track how each of its components evolve very easily. This amounts, of course, to solving the Schrödinger Equation: or, more satisfyingly put, solving the problem of time evolution and solving both the central problems of quantum mechanics.

SE2. An important clarification which has been alluded to but not explicitly stated thus far must now be emphasised. There is no physical operator which changes form over time. Most operators, like the position and momentum operators, do not change even if there is a perturbation to

the system. However, the Hamiltonian is not the same for all systems. Like all other observables operators, its formulation does not change- but, unlike most other operators, this formulation consists of something which can vary through perturbations. Specifically, the Hamiltonian is expressed as

$$\hat{H} = \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(\hat{X})$$

where x here is the position variable and the function $V(\hat{X})$ is the potential of the system. It is the potential which changes the form of the operator, as different systems have different potentials; moreover, these potentials can shift over time so in that way the form of the Hamiltonian changes over time. Note the difference: the formula to construct the Hamiltonian doesn't change, but the Hamiltonian itself does– when $V(\hat{X})$ changes.

The role of energy eigenbases as the easiest way to solve the Schrödinger Equation is now however fully clear to us. For a given system, if we can formulate the Hamiltonian accurately, then we can attempt to find its eigenvalues and corresponding eigenvectors using the characteristic equation, and then by SE1 we can theoretically determine the answer to the time evolution problem for the state with that Hamiltonian. This is why almost invariably, for a given quantum system of interest for which we are trying to understand the dynamics (that is, the time evolution) the goal of the physicists are:

(a) Formulate the Hamiltonian. This can be exceptionally difficult, for example if you have a group of electrons interacting with each other for which the form of the potential is very difficult to derive. Many problems in ongoing research are to find the Hamiltonian of relevant systems.

(b) Try to solve, either analytically or by numerical approximation, what the eigenstates of the given Hamiltonian are. This can also be essentially impossible for some problems, which is why considerably ingenuity in algebraic manipulation and improved approximation schemes is needed. Looking for such techniques is also a very important part of modern-day quantum research.

SE3. By Postulate 3, the value $|(\boldsymbol{\varepsilon}_n, \Psi_t)|^2$ is the probability energy value

E_n is measured to be the value of energy for the system at time t. In computing this amplitude we achieve very interesting results:

$$\begin{aligned} |\,(\boldsymbol{\varepsilon}_n, \Psi_t)\,|^2 &= |\,(\boldsymbol{\varepsilon}_n, \Psi_0)\, e^{-\frac{iE_n t}{\hbar}}|^2 \\ &= |\,(\boldsymbol{\varepsilon}_n, \Psi_0)\,|^2 e^{-\frac{iE_n t}{\hbar}} e^{\frac{iE_n t}{\hbar}} \\ &= |\,(\boldsymbol{\varepsilon}_n, \Psi_0)\,|^2 \times 1 \\ &= |\,(\boldsymbol{\varepsilon}_n, \Psi_0)\,|^2 \end{aligned}$$

In other words, the probability of measuring the energy E_n at time t, represented by $|\,(\boldsymbol{\varepsilon}_n, \Psi_t)\,|^2$, is the exact same as the probability of measuring that energy at time 0, which is $|\,(\boldsymbol{\varepsilon}_n, \Psi_0)\,|^2$. In another sense this means that, unless there is some perturbation to our system, there is no change in the probability of a certain energy value being measured; this rule of time evolution essentially amounts to the energy conservation law for a closed unperturbed system as whatever energy we measure it to have at time 0 stays the same for all time t.

This is not the last of this important result! We once more see that the above holds for any observables with the eigenbasis $\{\boldsymbol{\varepsilon}_i\}$ which is the same as the energy eigenbasis, as the energy values themselves cancel out, leaving the component amplitudes purely in terms of the eigenvectors. This in fact means that for other observables compatible with the energy- for other observables whose operators commute with the Hamiltonian– the probability of making a measurement of the eigenvalue corresponding to a certain eigenvector is also constant over time. This is significant, as we have just proven the quantum mechanical requirement for some observable to be a **constant of motion**.

SE4. Consider the case when the state vector is in a pure energy eigenstate- when

$$\Psi_t = \boldsymbol{\varepsilon}_k$$

for some k. Then the probabilities of measuring the eigenvalues E_n are

$$|\,(\boldsymbol{\varepsilon}_n, \Psi_t)\,|^2 = |\,(\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_k)\,|^2 = \delta_{nk},$$

which means that the probability of measuring the energy E_k is 1 and the probability of measuring all other energies $E_{n\neq k}$ is 0. This is the deterministic energy pure state, whose relevance is clear due to the

heavy discussion following Postulate 3 on measurement and quantum states. Yet there is more to be said:

$$(\boldsymbol{\varepsilon}_n, \Psi_t) = (\boldsymbol{\varepsilon}_n, \Psi_0)\, e^{-iE_n t/\hbar}$$

so the state vector would be

$$\Psi_t = \sum_n (\boldsymbol{\varepsilon}_n, \boldsymbol{\varepsilon}_k)\, e^{-iE_n t/\hbar} \boldsymbol{\varepsilon}_n = \sum_n \delta_{nk} e^{-iE_n t/\hbar} \boldsymbol{\varepsilon}_n = e^{-iE_k t/\hbar} \boldsymbol{\varepsilon}_k.$$

However, this is the same as the system at time 0 because the exponential has modulus 1 and therefore does not change the eigenvector in the Hilbert space from the ray $\boldsymbol{\varepsilon}_k$ which represents the initial state. Thus the whole system does not change at all if it starts in a pure energy eigenstate; therefore, all observables remain constant under time evolution so long as the Hamiltonian remains the same.

The consequence of this is that we can define the vectors:

$$\forall n \in \mathbb{Z}^+, \quad \Psi_t^{(n)} := e^{-iE_n t/\hbar} \boldsymbol{\varepsilon}_n.$$

and these vectors $\Psi_t^{(n)}$ are the deterministic forms of the state vector if at initial state for $t = 0$ the state vector is coincident with the eigenvector $\boldsymbol{\varepsilon}_n$. We call these **stationary states**, which belong to the specific system, as their value is contingent on the eigenbasis of the Hamiltonian which describes the system. Now, if at time 0 the system is in a pure energy eigenstate $\boldsymbol{\varepsilon}_n$, then:

$$\Psi_t = \Psi_t^{(n)}.$$

Otherwise, if it is in a mixed state at time 0, then

$$\Psi_t(x) = \sum_n e^{-iE_n t/\hbar}\, (\boldsymbol{\varepsilon}_n, \Psi_0)\, \boldsymbol{\varepsilon}_n$$

by SE1, which can then be written as

$$\Psi_t = \sum_n (\boldsymbol{\varepsilon}_n, \Psi_0)\, \Psi_t^{(n)}$$

since $\Psi_0^{(n)}$ (the stationary state at time 0) is the initial pure state corresponding to stationary state n, which is the eigenvector $\boldsymbol{\varepsilon}_n$. This gives

us a physically meaningful method of describing the solution to the Schrödinger Equation: finding its stationary states along with knowledge of the initial state is sufficient to solve the Schrödinger Equation.

Note that the solution now does not give us a deterministic state if the initial state was not a pure state; we get a mixed state in terms of the stationary states. Thus we might be confused as to how it amounts to a solution: but we recall that a fully defined mixed state is most of the time tantamount to achieving the highest level of understanding physically possible, by the measurement postulates– so it does count as a solution since it represents the precise superposition of possible states we get in reality. Finally, the solution of finding the stationary states is absolutely the same as the solution detailed in SE2 and SE3– we've just labelled vectors to be stationary states, but the underlying procedure is still simple. This is:

(a) Formulate the Hamiltonian for the system

(b) Solve the time-independent Schrödinger, or energy eigenvalue equation.

(c) Formulate the state vector as a function of time and the initial state as shown above.

6.1.2 Free Particle

The first problem we can easily solve is the solution of the time evolution of a free particle. A free particle is a particle in a zero potential: so $V(\hat{X}) = 0$. We begin by listing out our procedure:

1. List the boundary conditions.

2. Formulate the Hamiltonian and solve its eigenvalue equation.

3. Use the eigenvectors and their eigenvalues to formulate the stationary states and therefore the state vector.

An important thing to note is that this procedure only applies for systems where the Hamiltonian remains constant. This occurs if and only if the potential of the system does not vary with time– we recall that the Hamiltonian

consists of the kinetic energy operator and therefore is not affected by the kinetic energy of the particle changing! The energy of the state will be affected, but the operator representing all possible energy states will not because the kinetic energy operator is constant. For the questions in these chapters, as well as for most introductory quantum mechanical problems, time-varying potentials will not be considered, as they are extremely complex and do not have any place in the rudimentary foundations of quantum mechanics.

Let us now solve this simplest problem in quantum mechanics with this procedure.

List the Boundary Conditions

The "boundary conditions" of a problem are simply all the conditions the problem sets up for a solution. Listing them at this stage can be very useful because there are many times, for example when evaluating integrals or probabilities, that the boundary conditions can provide insights to shortcuts where there seem like there are too many obstacles to a solution. For now, though, this as a separate step will not be vindicated considering there is one boundary condition specified here only.

1. The potential energy is 0.

Formulate the Hamiltonian

This is also quite easy, as we do not have to consider the potential.

$$\hat{H} := -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(\hat{X}),\ \ V(\hat{X}) = 0 \Rightarrow \hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}$$

Solving the eigenvalue equation will take more work since this is the first time we have done it for such a complicated problem. We want to find eigenstates and eigenvalues which satisfy the eigenvalue equation

$$\hat{H}\boldsymbol{\varepsilon}_n = E_n\boldsymbol{\varepsilon}_n$$

for real E_n. This equation is, as aforementioned, also known as the time-independent Schrödinger Equation in literature. We can solve it for the free particle relatively easily. The Hamiltonian is

$$\hat{H}\Psi(x) = \frac{\hat{P}^2}{2m}\Psi(x)$$

and so the eigenvalue equation (time-independent Schrödinger) is to find eigenvectors $\{\boldsymbol{\varepsilon}_n\}$ which satisfy:

$$\frac{\hat{P}^2}{2m}\boldsymbol{\varepsilon}_n = E_n\boldsymbol{\varepsilon}_n$$

for some real constant eigenvalues E_n. The crucial insight for this eigenvalue equation is that the Hamiltonian operator of the system commutes with the momentum operator! This is intuitively obvious, since the Hamiltonian operator is the momentum operator squared divided by a constant, and the momentum operator is just the momentum operator, so the commutator

$$\left[\frac{\hat{P}^2}{2m}, \hat{P}\right]$$

consists of only the operator $\hat{P}$, somewhat modified on one side. Constants clearly do not affect commutators and operators commute with themselves, so we would expect the commutator to be 0. It can be verified by the reader as well, if we put in $\hat{P} = -i\hbar\frac{d}{dx}$, but this is not greatly necessary.

From our vigilance, however, in checking the commutation relation between the Hamiltonian and momentum operators (over time, one gains an intuitive feeling for this vigilance) we achieve something much greater. As their operators commute, energy and position are compatible observables. And as they are compatible observables, they must possess a common eigenbasis! Therefore, to any energy eigenvector $\boldsymbol{\varepsilon}_n$ there also exists a momentum eigenvector ϕ_n which is the same function! Let us therefore try to put that in instead, with this information. For all energy eigenvectors $\boldsymbol{\varepsilon}_n$ such that

$$\hat{H}\boldsymbol{\varepsilon}_n = \frac{\hat{P}^2}{2m}\boldsymbol{\varepsilon}_n = E_n\boldsymbol{\varepsilon}_n,$$

there must exist a momentum eigenvector such that

$$\phi_n \equiv \boldsymbol{\varepsilon}_n \implies \frac{\hat{P}^2}{2m}\phi_n = E_n\phi_n.$$

Now, if the eigenvalue of momentum corresponding to the eigenvector ϕ_n is P_n, then the eigenvalue corresponding to $\hat{P}^2$ on ϕ_n is P_n^2. So the above can be written

$$\frac{\hat{P}^2}{2m}\phi_n = E_n\phi_n \iff \frac{P_n^2}{2m}\phi_n = E_n\phi_n.$$

The eigenfunction ϕ_n is clearly not the null vector, and therefore the above implies that the n'th eigenmomenta and eigenenergies are related by:

$$\frac{P_n^2}{2m} = E_n \implies P_n = \pm\sqrt{2mE}.$$

We can now try to find the free particle wavefunction. Replacing the momentum operator with its algebraic formulation (in position space, as we have been working thus far), we have

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} = E\Psi \Rightarrow \frac{d^2\Psi}{dx^2} = -\frac{2mE}{h^2}\Psi.$$

We could also express this as

$$\frac{d^2\Psi}{dx^2} = -\frac{p^2}{h^2}\Psi$$

since we know that for the free particle $E = p^2/2m$ as just shown. We have already seen that momentum has extra importance in the free particle problem as it is compatible with energy (though that is not to say this is the only such question where this may be true); therefore, the momentum eigenstates are also the energy eigenstates, which we know are very important due to their intimate relationship with time evolution. It is also very clear that the free particle problem is a fundamental conceptual problem. De Broglie therefore defined a relationship

$$k = p/\hbar \implies p = \hbar k$$

(where the latter form is far more commonly seen) for a constant k which we will now plug into the equation we have above:

$$\frac{d^2\Psi}{dx^2} = -\frac{p^2}{h^2}\Psi, \ p = \hbar k \implies \frac{d^2\Psi}{dx^2} = -k^2\Psi.$$

The physical meaning of the constant k is somewhat tangential for this discussion, though it must be remarked that the De Broglie relations concern all the important aspects of a classical wave and are not simply random definitions. Nevertheless, for our purposes working with k instead of $p/\hbar$ will be cleaner for the algebra to follow. A reader should clearly see that

the equation we have just reached is perfectly analogous to the rudimentary differential equation

$$\frac{d^2y}{dx^2} = -k^2y.$$

Which has the general solution

$$y = Ae^{-ikx} + Be^{ikx} := \Psi.$$

This is therefore the general solution to the free particle wavefunction where constants A and B are to be determined based on further boundary conditions of the physical problem! It looks generic, but when we do set bounds: whether these be positional bounds or any other bounds, we will see that we can suddenly get quite interesting and specific forms for the wavefunction. Such is for example shown in part 8.2, where we see the free particle confined to an ellipse and get a very clear reult.

6.2 A Holistic Summary

The meaning of these last three chapters has always been to provide a robust defence against the many conceptual pitfalls which one can fall into if they try to educate themselves on quantum mechanics purely mathematically without grasping the underlying ideas fully. Certainly, the rest of the book will be far more dense mathematically- though effort will still be made to be extremely clear with that discourse. It is therefore important, given that I have offered such a verbose discourse on the relationship between quantum mechanics and physical reality, to summarise everything we have learnt in a concise way; by revising this section, the reader should be able to answer all the physical questions they have about quantum mechanics.

The problem of a physical model consists of two components, the state problem and the time-evolution problem. The state problem consists of two main problems: the first, how we represent physical information; the second, how we extract physical information.

The representation of physical information

- The representation of physical information is carried out by the state vector and its wavefunctions in quantum mechanics. It is represented

by a normed Hilbert space vector; all unphysically scaled multiples are considered the same ray so represent the same state.

- In the book we have referred to the state vector at times as "storing information", which is somewhat of a shorthand and can be confusing. What really is true is that the state vector is given meaning by the relation

$$P(\boldsymbol{\alpha}_i) = |\,(\boldsymbol{\alpha}_i, \Psi)\,|^2$$

for any arbitrary eigenstate $\boldsymbol{\alpha}_i$, which means we can convert it to wavefunctions which are probability distribution functions. Subsequently, the wavefunction does store information because it is defined by its components, which must correspond to specific eigenvectors and therefore are probability amplitudes. As every arbitrary state vector must have a unique expansion in every basis spanning the space, every state vector therefore inherently can be converted to its expression in any space, and therefore every state vector possesses components in observable eigenspaces- which are, probability amplitudes of measurements with respect to those physical observables whose eigenspace it is in. That is how a state vector encapsulates information.

- Physical observables are represented by hermitian operators, which do not change over time. The distinct eigenvectors of these observable operators and the corresponding eigenvalues are also therefore unchanging over time. The Hamiltonian is an exception, but not because its formulation changes over time, but because its form changes over time as the potential $V(\hat{X})$ may change over time.

- The vector space expressed as being spanned by an observable's eigenbasis is called its eigenspace. It is especially useful when considering measurements of that observable, since the components of the state vector in that eigenbasis are the probability amplitudes corresponding to measurements of the observable.

6.3 Exercises from Chapter 5*

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

Chapter 7

Chapter 7: Dirac Notation and Matrix Mechanics

The second part of this book, starting now, will be far more mathematically demanding. It seeks to complete the original aim of this pragmatic book: to allow a reader to segue onto mathematically advanced quantum mechanics texts without chronic confusions of what the mathematics *means*. Particularly, we will now tackle the challenging case of extending our quantum mechanics to cover continuously distributed observables.

We have covered now all of the quantum mechanical postulates we need to cover. On the other hand, our problem solving skills, which are very different in quantum mechanics to just standard algebra, are still very limited in the wider scope of things. There are many manipulative techniques we have not covered which are fundamental, but far from trivial. Furthermore, learning these techniques will require a reformulation of almost everything we have done so far into a new notation. This switch will be strange and unwelcome at first, but sooner than later the reader will find that the new notation is far more useful than confusing, and at that point they will be on a very good path indeed.

The benefit of this somewhat counterintuitive approach of learning one way and then reformulating into a new notation is not only that we don't have several factors affecting our understanding at once and causing trouble, but that the reader gets training to an undoubtedly undergraduate level without wrecking their ability to appreciate nuances as they grapple with difficult

problems. Without a wider context of measurements and probabilities and state evolution within which to frame the mathematics in this book, it would seem like an endless barrage of symbols for which we have no use and no appreciation. On the other hand, after reading the previous chapters, we are capable of understanding why components are important– they are probability amplitudes; how the state vector is both a label and can be used as a substitute for a physical state; how operators are central to the measurements we achieve for different operators. With all this conceptual learning done, in many ways our task from hereon is relatively simple– change notation and crunch algebra without any other concerns any more! Remember what started things off. That (↑) notation, which was designed to show how unimportant symbols themselves are, was where our discussion started, and we now come full circle to this idea that notation is far from sacred and must be evaluated simply based on pragmatic convenience. This new Dirac notation we are switching too will certainly make things more convenient, even though the state vector is still a state vector and the wavefunctions are still wavefunctions and the operators are still operators. Subsequently, looking in greater detail at the mathematical processes of quantum mechanics will become easier with our increased ability to engage in more efficient algebra. Nothing has changed. We are just switching from more intuitive symbols to the proper quantum mechanical convention of Dirac notation, which one might not have seen before but which needn't be mystified.

In addition, it is simply necessary to cover the conventional notation, called Dirac notation for its founder, and the pertinent techniques which follow it, because there does not exist a textbook whose level is a step up from this one and which does not use such conventions and even in many cases assume it is prior mastered. We will also see that it would have been difficult to understand what it means outright before we understood what a state vector is, for example, but on the other hand now that we do understand what the state vector is, it will make problem solving more powerful without any danger of confusions.

In this chapter, we will set forwards some new limitations so we can focus on the algebraic content (given that we understand the concepts of the postulates quite well now). The primary limitation will be the one maintained in previous chapters: that we are working in discrete cases. The next chapter will be all dealing with continuity, which contains some extremely difficult

mathematics resulting from the mathematics in this chapter, so this limitation will be soon loosened. We will also be avoiding any thoughts about the wavefunctions being the primary 'tangible' representations of abstract state vectors, because we will see that with Dirac's notation comes another natural and powerful idea of matrix representations! We can also feel free to ignore measurements and related probabilities so that we can focus on developing algebraic techniques which will make problem solving more efficient and workable, because we already understand the postulates we need to therein. So for now, the job is mainly to forget the wavefunction, and focus on the more explicitly state vector (and matrix) focused approach we will start below. We appreciate throughout that wavefunctions (and therefore measurement probabilities) are still obtainable, by taking the inner product of a state vector with the relevant input basis vectors in that component function transformation we have seen, so its absence is far from its demise.

7.1 Dirac Bra-Ket Notation

Our new notation begins here. Like any new notation, our job will be changing the way we label the objects we want to consider. The starting point will be the state vector, and its notation. We will not change any properties of the state vector! Things may look different because the notation looks nothing like we have ever seen before, but that does not mean at all that the characteristics are different.

The state vector will now be written as $|\Psi\rangle$. This called **ket** psi. We simply call it a ket instead of vector because Dirac notation is so ubiquitous in quantum mechanics that its authority over the discourse of quantum mechanics has turned the word ket into a common noun in itself. The ket and the state vector are one and the same! The only difference is that we now use the notation $|\Psi\rangle$ instead of the notation Ψ.

The space of possible physical states will now be called the ket space instead of the state space. Obviously, since kets are the same as state vectors, kets are still in bijection with physical states. The vectors which span the ket space are also kets themselves. Therefore, an arbitrary ket can be expressed

with components $\{c_i\}$ as

$$|X\rangle := \sum_{i=1}^{n} c_i |\alpha_i\rangle$$

for a basis set $\{|\alpha_i\rangle\}$. We can define scalar multiplication exactly as we have before:

$$k \times |X\rangle = k \sum_{i=1}^{n} c_i |\alpha_i\rangle = \sum_{i=1}^{n} (kc_i)|\alpha_i\rangle$$

where the scalar multiplies each component of the vector to give it new components. Addition is defined by summing corresponding components as well:

$$|X\rangle := \sum_{i=1}^{n} c_i |\alpha_i\rangle, |Y\rangle := \sum_{i=1}^{n} c_i' |\alpha_i\rangle \implies |X\rangle + |Y\rangle = \sum_{i=1}^{n} (c + c_i')|\alpha_i\rangle$$

An extremely important point, again, is that the ket is an abstract object. Recall that the state vector could not be given a form until we chose a specific basis. Since the ket is the same, we will here need to chose a basis for the ket to take a form. The key novelty of our new way of approaching the same ideas is that, in our bid to make our state vector, now ket, more 'tangible', we will this time try to divorce the ket from the idea of a wavefunction representation and replace this with yet another representation in bijection to the state vector– a matrix representation! This is because ideas of matrices will be very useful to us in our work and manipulation. To do this, recall that in any basis a constituent of any vector space has unique components in that basis, most often referred to as its unique expansion in that basis. This allows us to create a unique matrix expression of any ket: for any

$$|X\rangle := \sum_{i=1}^{n} c_i |\alpha_i\rangle$$

in the basis $\{|\alpha_i\rangle\}$, we can express it as a column vector of its components

$$|X\rangle = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

and this is a unique way to express the ket in that basis. As all kets have the same dimensions in this matrix expression (since the dimension of the ket space is the same so the number of components specifying them are the same), we can form any linear combinations of them since we are just summing matrices of the same dimensions, which is acceptable, and this matrix expression will not be meaningless.

Another important point is for the inner labelling of kets: i.e-

$$|\textbf{Inner Label}\rangle.$$

The inner label of a ket never has any mathematical meaning. We can label kets

$$|1\rangle, |2\rangle, ...|5\rangle$$

if we want to. They are absolutely nothing to do with the numbers 1 to 5. Instead, the inner label serves to **organise** different kets: above, we have ordered ket 1, ket 2, ket 3, and so on in some arbitrary way which is meaningful to us. We could also label them

$$|\text{Dirac}\rangle, |\text{Heisenberg}\rangle, |\text{Feynman}\rangle$$

except this provides no logical order for us and is long to write. Sometimes one will see some long-winded labels, such as

$$|P = \sqrt{2mE/\hbar}\,\rangle$$

which is an example taken from Chapter 8, where we needed to label a ket by its momentum value $\sqrt{2mE/\hbar}$ to distinguish it from another ket with another momentum value (but otherwise identical characteristics). One should not be confused if there are numerical numbers as inner labels of kets: they are part of a taxonomy we have chosen ourselves. For example, in this book and ubiquitously in all books, convention is to denote eigenkets by the eigenvalue they represent. This is not to say that the label is physically meaningful– we could easily label each eigenket by the square of its eigenvalue, for example- but it is useful to remember and use because it makes things clear for us when we have several lines of algebra simultaneously. Next, we need to complete the new notation for the inner product.

7.1.1 Bra space and Inner Products

One operation we cannot perform with a ket alone is the all-crucial inner product. To define this, we first define a new vector space. This vector space is the bra space, and is made by a simple transformation: for every ket $|X\rangle$, the corresponding bra is written $\langle X|$ and is defined by

$$\langle X| = |X\rangle^{\dagger}.$$

The "dagger" notation means the **Hermitian adjoint** of the ket $|X\rangle$. This is the formal way to refer to the transpose conjugate of any matrix (see the matrix preliminary) – that is, complex conjugate all the matrix element values and then transpose the matrix. For the column vector kets, this therefore turns them to row vectors with the complex conjugate entries:

$$|X\rangle = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \Rightarrow \ \langle X| = \begin{bmatrix} c_1^* & c_2^* \dots c_n^*. \end{bmatrix}$$

One sees that the relationship exhibited between the two vector spaces of the ket space of kets and the bra space of bras is bijection. This is because every ket is unique as a column vector– due to its unique expansion in any basis set– and therefore taking the Hermitian adjoint will give us one single unique bra. For every element of the ket space there is therefore a single element in the bra space, and vice versa. This exists precisely because of the way we have defined the bra space.

Why define such a bra space? Bras never represent physical states directly- even though they are in bijection with the state vectors (kets) which are in bijection with physical states so each bra is in bijection with physical states. The answer is simple– because we need the inner product to be a real scalar given two input states: as we have already seen for the last few chapters. We cannot simply multiply two kets because we cannot multiply two column vectors by the rules of matrix multiplication! Therefore, we define the bra, which is a row vector with the complex conjugate entries, and now if we multiply them:

$$\langle X| \times |X\rangle$$

we should get a $1 \times n$ row vector multiplying an $n \times 1$ column vector, and therefore a 1×1 result: eg, a scalar, as desired.

We should be convinced that this new inner product is exactly the same as the old inner product. Take any two state vectors

$$\forall\ |\beta\rangle := \sum_{i=1}^{n} b_i |\alpha_i\rangle,\ \ |\gamma\rangle := \sum_{i=1}^{n} c_i |\alpha_i\rangle,$$

By the rules of matrix multiplication, the i'th column component of the row vector multiplies the i'th row component of the column vector. Thus we have

$$\langle\beta| \times |\gamma\rangle := \langle\beta|\gamma\rangle = \sum_{i=1}^{n} b_i^* c_i.$$

as the components of the ket column vector, here $|\beta\rangle$ are replaced by their complex conjugates in the bra row vector. There is absolutely no difference to the inner product operation: only, we consider it from a new perspective due to the matrix representations we have introduced.

Several properties we have established previously can now be written in the new notation:

- $\langle X|Y\rangle = \langle Y|X\rangle^*$
- $\langle X|X\rangle \geq 0$
- A normalised ket $|\tilde{\alpha}\rangle$ is such that $\langle\tilde{\alpha}|\tilde{\alpha}\rangle = 1$.
- Two kets $|\alpha\rangle$ and $|\beta\rangle$ are orthogonal if $\langle\alpha|\beta\rangle = 0$.

To further the bijection between the ket space and bra space, as bras will prove more useful than only to appear in inner products, we have, using the conventional $\leftrightarrow$ symbol to mean "corresponds to":

$$\begin{aligned} |X\rangle &\leftrightarrow \langle X| \\ c|X\rangle &\leftrightarrow c^*\langle X| \\ |X\rangle + |Y\rangle &\leftrightarrow \langle X| + \langle Y|. \end{aligned}$$

The way we obtain this is an important algebraic point: by performing the operation "dagger" on any linear combination of kets, we immediately create a new set of bras in bijection with linear combinations of kets of that type, in much of a similar argument to the one we used originally to show the bra

space is in bijection to the ket space.

The expansion of bras in their basis should be clear following the above facts. For

$$|X\rangle := \sum_{i=1}^{n} c_i |\alpha_i\rangle$$

we have

$$\langle X| = \sum_{i=1}^{n} c_i^* \langle \alpha_i|.$$

Thus completes our investigation of the bra space, and now we can continue to reformulate old ideas in this new notation to familiarise ourselves with it.

Further with inner label conventions: in general, for scalars we write $|aX\rangle$ to mean $a|X\rangle$, and $\langle aX|$ to mean $a^*\langle X|$. It is up to the reader to distinguish which are the scalars and which the placeholder letters for kets and bras in question, though the context should never make this in any real doubt if one follows the steps through. Similarly, we want to ensure brevity with this notation as one (but not the only) of the benefits of employing bra ket notation, so we also often write $|aX + bY\rangle$ instead of $|aX\rangle + |bY\rangle$ or $a|X\rangle + b|Y\rangle$. Using the ket $|aX + bY\rangle := |aX\rangle + |bY\rangle$ is somewhat easier than for example defining

$$|Z\rangle := |aX\rangle + |bY\rangle$$

every time we sum two kets, where a new letter in the inner label would just be confusing.

The inner product short-form facts are always helpful; indeed, we have *linearity in ket*:

$$\langle V|aX + bY\rangle \equiv \langle V| \times (|aX\rangle + |bY\rangle) \equiv \langle V|aX\rangle + \langle V|bY\rangle \equiv a\langle V|X\rangle + b\langle V|Y\rangle.$$

This fact is one we have seen already (fact S7 in Chapter 3) for inner products, but expressed in Dirac notation. The comparable idea exists in bra:

$$\langle aX + bY|V\rangle = (\langle aX| + \langle bY|) \times |X\rangle = \langle aX|V\rangle + \langle bY|V\rangle = a^*\langle X|V\rangle + b^*\langle Y|V\rangle.$$

Next, we can also recall the ubiquitous expansion. Take an arbitrary ket $|X\rangle$

in the orthonormal basis $\{|\tilde{\alpha}_i\rangle\}$ and the following will always hold:

$$|X\rangle := \sum_{i=1}^{n} c_i|\tilde{\alpha}_i\rangle \implies \forall j,\ \langle\tilde{\alpha}_j|X\rangle = \langle\tilde{\alpha}_j| \times \sum_{i=1}^{n} c_i|\tilde{\alpha}_i\rangle.$$

By linearity in ket, the outside bra is absorbed into the sum notation so we get

$$\langle\tilde{\alpha}_j|X\rangle = \sum_{i=1}^{n} c_i\langle\tilde{\alpha}_j|\tilde{\alpha}_i\rangle = \sum_{i=1}^{n} c_i\delta_{ij}$$

by the orthonormality of the basis. Then, this is simply

$$\langle\tilde{\alpha}_j|X\rangle = c_j,$$

which should certainly be familiar now, though perhaps not before in Dirac notation. It tells us that, given a basis, if we orthonormalise it (with the Gram-Schmidt process), the expansion coefficients of any ket are not random: they are the inner products of the ket being expanded with the corresponding basis kets. So for any ket $|X\rangle$,

$$|X\rangle = \sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|X\rangle.$$

It is important to represent it explicitly here because the above will be used universally for problems to follow, and without explicitly stating that this is the expansion we already know it could otherwise be confusing. The reader should find it simple to produce a bra expansion in an orthonormal bra basis, though this is seldom as useful or commonly used.

7.1.2 Operators and Eigenkets

The action of an operator Ω on a ket $|X\rangle$ is written $\Omega|X\rangle$. This is perhaps why we incorporate scalars into the ket inner label, as noted above– so that there is no mixup between scalar multiplication and the action of an operator on a ket. The action of an operator on a bra is conversely written $\langle X|\Omega$.

We always focus on linear operators which map one ket onto another ket in the vector space. This makes sense: the application of observable operators for example we do not expect to make an impossible state, and any possible state is represented by some superposition of the state vectors which

constitute the ket space. It is good to keep in mind very explicitly the notion that an operator acting on a ket always produces a new ket. More properties of linear operators can be expressed in bra-ket notation:

- $a\Omega|X\rangle = \Omega a|X\rangle$
- $\langle X|a\Omega = \langle X|\Omega a$
- $\Omega_1|X\rangle = \Omega_2|X\rangle \forall |X\rangle \implies \Omega_1 = \Omega_2$
- $\Omega\{a|X\rangle + b|Y\rangle\} = a\Omega|X\rangle + b\Omega|Y\rangle$
- $\Omega\{a\langle X| + b\langle Y|\} = a\langle X|\Omega + b\langle Y|\Omega$

The product of two operators means to apply the "closer" operator to the ket or bra respectively and then apply the second, "further" operator to the resulting ket or bra. We already know from our study of compatibility in particular that the assumption that two given operators will commute is false. The commutator is denoted the same way:

$$[\Omega, \Lambda] := \Omega\Lambda - \Lambda\Omega$$

and so is the anticommutator:

$$\{\Omega, \Lambda\} := \Omega\Lambda + \Lambda\Omega.$$

The eigenvectors of the operators will now often be called eigenkets, for obvious reasons, and the name eigenbras follows for bras in bijection with eigenkets.

We must now introduce the unique way to specify an operator as a matrix. This will allow us to perform some important operations in the future.

We start by noting that the action of an operator on the basis vectors of a ket space (any basis, not just its own eigenbasis!) is sufficient knowledge to specify its actions on all kets in that basis:

$$\Omega|\alpha_i\rangle = |\alpha_i'\rangle \Rightarrow \ \forall \ |X\rangle := \sum_{i=1}^{n} c_i|\alpha_i\rangle, \quad \Omega|X\rangle = \Omega\sum_{i=1}^{n} c_i|\alpha_i\rangle$$

and by the linearity of the operator, this is just

$$\Omega|X\rangle = \sum_{i=1}^{n} \Omega c_i|\alpha_i\rangle = \sum_{i=1}^{n} c_i\Omega|\alpha_i\rangle = \sum_{i=1}^{n} c_i|\alpha_i'\rangle.$$

However, things are not over from the perspective of a matrix formulation. Not only does this specification not give any clue as to how to express the operator as a matrix, but the resulting ket after the transformation is made is also not in any way representable as a column vector. This is because we must remember that a linear transformation on the basis vectors of a space by no means produces another basis which still spans the space at all (in which case a vector of components makes no sense anymore). The best example of this, of course, would be if $\Omega = \Omega_0$, the null operator, in which case all $|\alpha_i'\rangle$ would be null kets and certainly span no space at all.

So the above informs us that to complete the task we need to return all the above to the original basis, which we know is serviceable for matrix representations. We thus start by expressing the components of the transformed kets $\{|\alpha_i'\rangle\}$ in the original basis. If

$$|\alpha_j'\rangle := \sum_{i=1}^{n} c_i^{(j)} |\alpha_i\rangle$$

and we assume the starting basis $\{|\alpha_i\rangle\}$ is orthonormal, since otherwise it could be orthonormalised, then we clearly know the components $c_i^{(j)}$ are simply the inner products $\langle\alpha_i|\alpha_j'\rangle$. This component is the component of the j'th transformed ket corresponding to the i'th basis ket. We can then define the entities

$$\Omega_{ij} := \langle\alpha_i|\alpha_j'\rangle = \langle\alpha_i|\Omega|\alpha_j\rangle$$

and these are the components of the transformed kets in the original basis before they were transformed. The original question can be reposed. If we define

$$\Omega|X\rangle := |x_0\rangle := \sum_{i=1}^{n} c_i'|\alpha_i\rangle$$

for some components c_i', then these components are

$$c_i' = \langle\alpha_i|x_0\rangle = \langle\alpha_i|\Omega|X\rangle = \langle\alpha_i|\Omega\left(\sum_{i=1}^{n} c_i|\alpha_i\rangle\right).$$

By the linearity of the operator, this becomes

$$c_i' = \langle\alpha_i| \times \left(\sum_{j=1}^{n} c_j \Omega|\alpha_j\rangle\right),$$

and then by linearity in ket this becomes

$$c_i' = \sum_{j=1}^{n} c_j \langle \alpha_i | \Omega | \alpha_j \rangle.$$

Thus using the same notations defined above this is simply

$$c_i' = \sum_{i=1}^{n} c_i \Omega_{ij}.$$

The notation Ω_{ij} is clearly meant to hint that these can be placed in some matrix where each value Ω_{ij} (note these are inner products, so they are indeed a scalar values) is the entry in the i'th row and j'th column of the matrix. And as each Ω_{ij} is the component of the j'th transformed ket corresponding to the i'th basis ket, we say that the upper limit of i and j are both n since there are n original basis kets, and therefore n transformed kets as well. Thus we can create an $n \times n$ matrix for all the entries Ω_{ij}. The relationship we get is that

$$\begin{bmatrix} c_1' \\ c_2' \\ \vdots \\ c_n' \end{bmatrix} = \begin{bmatrix} \langle \alpha_1 | \Omega | \alpha_1 \rangle & \langle \alpha_1 | \Omega | \alpha_2 \rangle & \dots & \langle \alpha_1 | \Omega | \alpha_n \rangle \\ \langle \alpha_2 | \Omega | \alpha_1 \rangle & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle \alpha_n | \Omega | \alpha_1 \rangle & \dots & \dots & \langle \alpha_n | \Omega | \alpha_n \rangle \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

as a way to relate the components of the transformed vector $\Omega|X\rangle := |x_0\rangle$ to the original components of $|X\rangle$ before it was transformed. We see that the left hand side is the matrix representation of $|x_0\rangle$ in the basis we have been using, since it specifies the components of $|x_0\rangle$ as a column vector. Meanwhile, the right column vector clearly does the same for the original $|X\rangle$. And therefore this whole matrix equation is clearly the matrix form of the definition $|x_0\rangle = \Omega|X\rangle$, which then means that the $n \times n$ matrix in the middle is the matrix representation of Ω. So to conclude this discussion we restate the fact that

$$\begin{bmatrix} \langle \alpha_1 | \Omega | \alpha_1 \rangle & \langle \alpha_1 | \Omega | \alpha_2 \rangle & \dots & \langle \alpha_1 | \Omega | \alpha_n \rangle \\ \langle \alpha_2 | \Omega | \alpha_1 \rangle & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \langle \alpha_n | \Omega | \alpha_1 \rangle & \dots & \dots & \langle \alpha_n | \Omega | \alpha_n \rangle \end{bmatrix}$$

is the way to represent the operator Ω in the n-dimensional ket space spanned by the orthonormal basis vectors $\{|\alpha_i\rangle\}$.

7.2 A Rudimentary Manipulation Toolbox

The introduction of Dirac notation is surprisingly powerful because its unorthodox aesthetic form allows for the formulation of some identities and operators which under normal circumstances would look like a long amalgamation of random algebraic entities, but look simple and orderly (and therefore easy to memorise and use) in Dirac notation. This section will introduce essential identities and axioms which will be used without second thought for advanced problem solving, and in doing so, I hope the reader can also continue the process of becoming fluent in Bra Ket.

7.2.1 General Solution of the Eigenvalue Problem

The first tool we will see is a general solution of the eigenvalue problem, for which we have been waiting. The eigenvalue condition is always

$$\Omega|\omega\rangle = \lambda|\omega\rangle.$$

for some eigenvalue λ and eigenvector $|\omega\rangle$ We know that all operators in the space can be represented by an $n \times n$ matrix (where n is the dimensionality of the space). Just to convert the eigenvalue λ into matrix form as well, we will multiply both sides by the $n \times n$ identity operator I:

$$\begin{aligned} \Omega|\omega\rangle &= \lambda I|\omega\rangle \\ \Rightarrow \quad (\Omega - \lambda I)|\omega\rangle &= 0 \end{aligned}$$

where 0 is the null matrix. Considering the whole in matrix form, we write:

$$\begin{bmatrix} \Omega_{1,1} - \lambda & \Omega_{1,2} & \dots & \Omega_{1,n} \\ \vdots & \Omega_{2,2} - \lambda & \vdots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ \Omega_{n,1} & \dots & \dots & \Omega_{n,n} - \lambda \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Looking at the matrix representation of the equation above, we can deduce that

$$\forall i, \ \sum_j c_j(\Omega_{ij} - \lambda\delta_{ij}) = 0$$

by the rules of matrix multiplication. This is a linear system of equations, with coefficients $\Omega_{ij} - \lambda\delta_{ij}$ and unknowns c_i. Therefore the logic behind

Cramer's Rules apply: and since the c_i are the components of $|\omega\rangle$ in the basis they are not all zero since $|\omega\rangle \neq 0$, and thus we need nontrivial solutions and therefore the determinant of the leftmost term must be zero. In other words,

$$\det(\Omega - \lambda I) = 0 \iff \forall i,\ \sum_j c_j(\Omega_{ij} - \lambda\delta_{ij}) = 0$$

This is practically helpful, especially in problems with fewer dimensional spaces, and as a theoretical concept which shows us we can always solve the eigenvalue problem. There will be other methods to solve eigenvalue equations too, but these often rely somewhat more on inspection and then proof that one has all the solutions by inspection; these inspection methods only arise as shortcuts based on specific conditions we get.

7.2.2 The Associative Axiom

Dirac introduced in his paper on Dirac notation a critical axiom called the Associative Axiom. This axiom will save a huge amount of confusion for those unfamiliar with Bra-Ket notation, and is possibly the most powerful idea of them all. This axiom states that all (legal) "multiplication" operations in Dirac notation are always associative.

Put in scalar terms, this word associative is usually seen as

$$(a \times b) \times c = a \times (b \times c).$$

In other words, so long as the order of the terms a, b, c are kept (here, a before b before c), the two multiplication operations can be performed in either order. With numbers this is completely natural (as also is additive associativity, but not subtractive or divisive associativity), but with Dirac notation this is much less trivial and also much more powerful.

The first example we have seen but not explained explicitly is the inner product

$$\langle X|\Omega|Y\rangle.$$

This is an inner product since it is the product of a bra, $\langle X|$ with a ket $\Omega|Y\rangle$ (which is another ket in the ket space the operator Ω has mapped the original ket $|Y\rangle$ to). However, it can also be seen as a bra, $\langle X|\Omega$, multiplied by a ket $|Y\rangle$. This is the associative axiom: in fact it does not matter which

way the multiplication goes and only because of this can we write the concise $\langle X|\Omega|Y\rangle$.

The ramifications of this fact are profound, as beginners with Dirac notation often get confused about which products are "bunched together" and must be performed first, leading to a great state of bewilderment when a term might have several bras, operators and kets next to each other. The answer is very simply that, so long as operations are being performed left to right, the order does not matter and the reader may read the multiplications in any way they like. It also immediately leads to theorems for entities in Dirac notation, by defining them through the perspective of other known products. The first example of this we will see is the following completeness relation.

The Outer Product

We discussed above how the associative axiom holds for all legal multiplications between bras, kets and operators. What constitutes an illegal multiplication? Well, to start we can give the example of the orthonormal expansion of any ket $|X\rangle$,

$$|X\rangle = \sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|X\rangle.$$

Contrast this to the naive expansion

$$|X\rangle = \sum_{i=1}^{n} c_i|\tilde{\alpha}_i\rangle$$

which we might use if we were not sure the basis was orthonormal, for example. We see that the position of the component shifted from the leftmost sum term to the rightmost sum term when c_i was replaced with $\langle\tilde{\alpha}_i|X\rangle$. One might ask why this is. The best answer for this is simply the associative axiom: if we had written

$$|X\rangle = \sum_{i=1}^{n} \langle\tilde{\alpha}_i|X\rangle|\tilde{\alpha}_i\rangle,$$

then we might think this is okay as we have a scalar $\langle\tilde{\alpha}_i|X\rangle$ multiplying a ket $|\tilde{\alpha}_i\rangle$. However, by the associative axiom we also expect the above to be

equally well expressed as

$$|X\rangle = \sum_{i=1}^{n} \langle\tilde{\alpha}_i| \times \Big(|X\rangle \times |\tilde{\alpha}_i\rangle \Big).$$

This is where our problem is– two kets cannot be multiplied together, as they are both $n \times 1$ matrices! Similarly, two bras cannot be multiplied together as they are both $1 \times n$ matrices. So this is why we cannot write

$$|X\rangle = \sum_{i=1}^{n} \langle\tilde{\alpha}_i|X\rangle|\tilde{\alpha}_i\rangle.$$

This however implies that for the expansion

$$|X\rangle = \sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|X\rangle$$

it is correct and therefore not only is the operation $|\tilde{\alpha}_i\rangle \times \langle\tilde{\alpha}_i|X\rangle$ legal (it is a scalar multiplying a ket, so it should be)- but also that the operation

$$(|\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|) \times |X\rangle$$

should be possible. This product $|\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|$ is called the outer product between the bra $\langle\tilde{\alpha}_i|$ and ket $|\tilde{\alpha}_i\rangle$. We can verify it should be possible, as it is an $n \times 1$ matrix multiplied by a $1 \times n$ matrix, which should give an $n \times n$ matrix. As it produces an $n \times n$ matrix it clearly does not result in a scalar like the inner product; rather, we might posit that it is an operator, as operators are represented by $n \times n$ matrices. This is in fact a true assumption: an outer product is fundamentally meant to be treated as an operator. However, before we discuss that we should establish and prove a fundamental theorem in quantum mechanics which will be surprisingly useful for manipulation. This is the completeness relation, which results from the associative axiom, and deserves its own section for its importance even though the proof requires no considerable thought if we have the associative axiom in mind.

The Completeness Relation

We have seen that the representation

$$|X\rangle = \sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|X\rangle$$

is allowed because the sum term $|\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i| \times |X\rangle$ is just the action of the outer product on the ket X. However, by the rule of linear operators that

$$(\Omega_1 + \Omega_2)|X\rangle = \Omega_1|X\rangle + \Omega_2|X\rangle$$

and the fact that the outer products are operators, we can use the same principle to write

$$|X\rangle = \left(\sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|\right)|X\rangle.$$

And this is very revealing, of course, because this is true for any arbitrary ket $|X\rangle$, and therefore we come to the conclusion that

$$\sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i| = 1.$$

This is the most commonly written form of the completeness relation, but one should be aware that the 1 here represents the identity operator, rather than a scalar– since the sum of $n \times n$ matrices will give us an $n \times n$ matrix rather than a scalar.

On the other hand, the more we foray into problem solving the more we will see of this seemingly completely useless relation. One example can be in the proof that the sum of the modulus squared components of a normalised ket is also equal to 1:

$$\langle X|X\rangle = \langle X| \times |X\rangle = \langle X| \times (1|X\rangle) = \langle X| \times \sum_{i=1}^{n} |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i| \times |X\rangle$$

where the number 1 is the identity operator again and the kets $\{\tilde{\alpha}_i\}$ are the orthonormal basis vectors. Then, by linearity in ket, we can write this as

$$\langle X|X\rangle = \sum_{i=1}^{n} \langle X| \times |\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i| \times |X\rangle$$

and by the associative axiom this is just

$$\langle X|X\rangle = \sum_{i=1}^{n} \langle X|\tilde{\alpha}_i\rangle\langle\tilde{\alpha}_i|X\rangle = \sum_{i=1}^{n} \langle\tilde{\alpha}_i|X\rangle \times (\langle\tilde{\alpha}_i|X\rangle)^* = \sum_{i=1}^{n} |\langle\tilde{\alpha}_i|X\rangle|^2.$$

However, as the inner products $\langle\tilde{\alpha}_i|X\rangle$ are the components of $|X\rangle$ in the basis $\{\tilde{\alpha}_i\}$, if the ket X is normalised then the left side is 1 so we get

$$1 = \sum_{i=1}^{n} |c_i|^2.$$

So the modulus squared of the components of a normalised ket in a basis sum to 1. We know these modulus squared components are probabilities if we are working in an observable space by the measurement postulate! So what we have just proved, in a flick of the pen when working with Dirac notation, is that the reason we like working with discrete wavefunctions is because their components in this way perfectly represent probabilities summing in total to 1! We have already proved this fact, at the end of Chapter 4 on normalised discrete wavefunctions being interpreted as probability mass functions, but, if one takes a quite glance at that, the proof is neither so short, nor so elegant: it is positively clunky in comparision. The purpose of showing this is to exhibit that, just like in normal algebra, the world of Dirac notation can belie some extremely unusual and ingenious manipulations which without the notation would not be so easy to express, but with the fluency in the notations, can be done in just a few lines or seen in a matter of seconds.

7.3 Change of Basis

One of the most important concepts in matrix mechanics is that of diagonal matrices. A square matrix is diagonal if all its elements are 0 except those in the major diagonal: that is, elements M_{ii}, which can be anything. One example of a diagonal matrix is the identity matrix, which has all zero entries except the major diagonal, which is filled with 1's.

Theorem: Every Hermitian operator has at least one orthonormal
eigenbasis in which its matrix representation is diagonal. The diagonal entries are the eigenvalues of the operator.

This proof is quite tough and not particularly illustrative for our current needs, and therefore will not be shown for now. However, we will find that the implications of this theorem are hugely powerful, because it means that if

we can diagonalise the operator matrix then we immediately solve the eigenvalue problem as the diagonal matrix values give us the eigenvalues from which deriving the eigenvectors should be easy: clearly, a great problem-solving step.

So we want to diagonalise the operator. By the theorem, this means we want to take one matrix in any basis to the eigenbasis which diagonalises the operator. Thus we will often be interested in a **change of basis**.

A change of basis is performed by applying an operator to the original basis kets and mapping them onto the new kets. Such is a common idea in quantum mechanics: to transform one ket to another we also try to see if there is a way we can formulate it as an operator equation. Here, we certainly can; such operators are usually denoted U. So we guess that for some original basis $\{|\alpha_i\rangle\}$ and operator Ω with eigenvalues $\{\beta_i\}$ and eigenvectors $\{|\beta_i\rangle\}$,

$$\forall i,\ U|\alpha_i\rangle = |\beta_i\rangle.$$

Assuming without loss of generality that these basis sets are both orthonormal, the operator which works is

$$U := \sum_j |\beta_j\rangle\langle\alpha_j|.$$

We can prove this. For any original basis ket $|\alpha_i\rangle$,

$$\left(\sum_j |\beta_j\rangle\langle\alpha_j|\right)|\alpha_i\rangle = \sum_j |\beta_j\rangle\langle\alpha_j|\alpha_i\rangle = \sum_j |\beta_j\rangle\delta_{ij} = |\beta_i\rangle$$

and this holds for any i since we have orthonormal sets. What is much more important, however, is that the operator U satisfies a very interesting condition. Consider, for the operator defined, the product

$$U^\dagger U = \left(\sum_k |\alpha_k\rangle\langle\beta_k|\right) \times \left(\sum_j |\beta_j\rangle\langle\alpha_j|\right)$$

As the basis sets are orthonormal, we see that all the terms disappear on account of the inner product $\langle\beta_k|\beta_j\rangle$, except when $k = j$. In those cases, we get

$$\sum_k |\alpha_k\rangle\langle\alpha_k| = 1.$$

So for this "transformation operator", we have $U^\dagger U = 1$. Similarly, $UU^\dagger = 1$. Such an operator is called **unitary**. Now consider some arbitrary operator which is unitary and its action on a state ket which is normalised. We know that for the ket $|\Psi\rangle$

7.4 Exercises from Chapter 6∗

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

Chapter 8

Chapter 8: Continuous Spectra

In this section, we take a huge leap forwards towards full physical reality by finally loosening the discrete spectrum restraint which we have been working under for the entire book so far. If we recall, eigenvalues of observable operators represent physically measurable valuables for those observables represented by those operators. Now, quantum mechanics does show us that some observables can be discretely distributed: most notably, energy as in the free particle solution. However, we also know that some observables must never be able to be discretely distributed. In particular, position being discretely distributed would be a huge issue in any physical model, because it would mean that us and any object moves by teleporting between discrete positions with increments between them which we somehow cannot move into. This is clearly nonsense; position must be continuous. Thus, we must also be prepared to deal with continuous spectra in quantum mechanics because, if nothing else, position eigenvalues must always be in a continuous spectrum. In the end, position is not the only observable which exhibits continuous spectra at all, but the illustration is already there.

8.1 A Conceptual Definition of Continuity

Clearly, as just established, some observables like position must take infinite continuously distributed eigenvalues. Now, if we take the eigenvalue definition:

$$\Omega|\omega\rangle = \lambda|\omega\rangle$$

then we can come to a clear realisation that we must also have infinite eigenkets as well. This is a result of the fact that the number of eigenvectors (eigenkets) must be greater than or equal to the number of eigenvalues. Otherwise, if there were fewer eigenvectors than eigenvalues, at least one eigenvector would correspond to multiple different eigenvalues. Then this would be:

$$\Omega|\omega\rangle = \lambda|\omega\rangle$$

and

$$\Omega|\omega\rangle = \lambda'|\omega\rangle$$

but

$$\lambda \neq \lambda' \implies \lambda|\omega\rangle \neq \lambda'|\omega\rangle$$

so

$$\Omega|\omega\rangle$$

would simultaneously be two different kets, even though it is the same operator acting on the same input ket. Then, Ω simply could not be a linear operator, as it would have two outputs for a single input.

So now that we have established the number of eigenvectors is greater than (the degenerate case) or equal to (the non-degenerate case) the number of eigenvalues, and we know that some observable operators take infinite eigenvalues, we must have indeed infinite eigenvectors as well in such cases.

Now, consider what practical function a eigenket has. It is an abstraction, so the answer is nothing in itself! The reason why an eigenket is important is because it is interpreted as the eigenstate where a corresponding eigenvalue has probability 1 of being measured. Thus, as already described in our section on inner labelling when we started with Dirac notation, we often just label eigenkets by that eigenvalue they are connected to.

Considering position eigenkets now, we would label them by

$$|x_0\rangle$$

being the eigenstate where a measurement of position would yield position x_0 with certainty. We might also have for example

$$|0\rangle, |L\rangle$$

if we were working bounded in some length L where a position could not be further than distance L away from the other end we labelled position 0. Now, we would never write:

$$|L\rangle > |L/2\rangle > |0\rangle$$

because kets are abstract vectors and it makes no sense to compare their magnitudes with a 'greater than' sign. Yet at the same time, that $|L\rangle > |L/2\rangle > |0\rangle$ would be easy to mistake as somewhat valid, because L is a positive distance so indeed the values these kets represent themselves do satisfy $L > L/2 > 0$. We can see that inner label confusion can occur when it comes to magnitudes, because if we a label a ket by a specific value: usually an eigenket by an eigenvalue, we can be tempted to compare the magnitudes of those inner labels.

What this pedantic clarification is to say is that

$$|L\rangle > |0\rangle$$

is mathematically void, but there is some concept of comparing the values labelling kets when it is a situation where values are labelling kets. Specifically, this means that when we write

$$|x_0 + dx\rangle,$$

we can claim it is 'infinitesimally close' to $|x_0\rangle$ as $dx \to 0$. Now, that technically would not be correct syntax, considering kets cannot be infinitesimally close if they do not have magnitudes and therefore the definitions of 'close' and 'infinitesimal' do not exist. However, the idea is that $|x_0 + dx\rangle$ and $|x_0\rangle$ may be infinitesimally close in that they represent eigenstates of positions which are infinitesimally close to each other! This is another case of the unfortunate semantic clarifications which result when we need to postulate so many abstractions in bijections, but at the same time, now that this is said, the reader should no longer scratch their heads when I or other texts say "$|x_0 + dx\rangle$ is infinitesimally close to $|x_0\rangle$" rather than "$|x_0 + dx\rangle$ represents a position which is infinitesimally close to the position represented by $|x_0\rangle$". Most compendia would not make such clarifications, but I think it is useful here.

Now, then we can finally understand how continuity occurs in quantum mechanics and in the abstract state spaces we are working in. While kets cannot exhibit continuity, they can represent values exhibiting continuity, and this can be seen as one and the same.

The mathematical definitions of continuity are:

- A variable Z is continuous over an interval $[a, b]$ if:

$$\forall\ z = Z \in [a, b],\ \ dz \to 0, \quad z + dz \in [a, b].$$

  This is probably already intuitive and not much more needs to be said: simply, there is a continuum of values which exist such that no matter how infinitesimal a scale we go down to, we will not be able to distinguish values by a discernible increment.

- A function $f(x)$ is continuous over x if x is a continuous variable and

$$\forall x, \lim_{dx \to 0} f(x + dx) = f(x).$$

  This means that there are no sudden jumps over infinitesimal intervals, and the graph of the function is *smooth*.

8.1.1 Infinities

A discussion on continuity necessarily involves a thought about infinities. The issue herein is that the mathematics of infinities is rather inaccessible; in fact, all we primarily need to understand is this extremely simplistic idea of $\infty + 1 \not> \infty$. I want to stray as far from these sums including infinities and arithmetic numbers as I can, and this is achievable to us in quantum mechanics. Quite simply, the less thought here, the better.

The one clarification I would like to make is on the comparison between discrete and continuous spectra in an infinite dimensional vector space.

If a vector space is infinite dimensional, then, we know that the basis of that space– that is, the set of linearly independent vectors in that space– must have infinite cardinality (there are infinite basis vectors), by definition of dimensionality. The state space, we know, is an example of a vector space

of infinite dimensions, and is critical to our study. Now, if we compare discretely distributed and continuously distributed eigenbases, we must make the clarification of how an infinite dimensional space can be spanned by both discrete bases and continuous bases. A reader might be surprised by this. If dimensionality is to do with the number of linearly independent basis vectors we can accommodate in the space, how could it be that one basis could be discrete and still have all the maximum number of basis vectors, while another basis is continuous? There is a sense, which is not unfounded, that the very point of continuous sets is that they have infinite elements between discrete points, and so they must have more elements than any discrete set.

This idea is countered by the question "How many integers are there?". Of course, we know there are infinite integers– that is, after all, the original point of having the entity infinity in the first place: if we keep on going through the consecutive positive integers we reach infinity. Can we say that there are more decimals than integers? In feeling, perhaps, but this is not mathematically rigorous. The idea is in fact of **countable and uncountable infinities**.

A countable infinity is an infinity where one can hypothetically number each of those infinite elements. The positive integers are natural for demonstrate this, because they are already numbered- the integer 1 is the first element, the integer 2 the second, and so on, but yet they are definitely infinite in number. On the other hand, all decimal numbers constitute an uncountable infinity- we cannot number them since we could by definition always produce at least 1 decimal between the decimal we have numbered to be the first and the decimal we have numbered to be the second.

This answers our question of how discrete and continuous basis can both span the same vector spaces. If a continuous basis spans a space, there must be infinite basis vectors, which is why the space is infinite dimensional. However, we could also span the space with a discrete basis, which has a *countably* infinite number of basis vectors rather than an uncountably infinite one. A continuous basis cannot span a finite dimensional vector space, but a discrete basis can span both finite and infinite dimensional vector spaces depending on what bases we are considering. Discrete and finite are not synonyms, even though we might consider them to be. Thus what might appear confusing in regards to these discrete versus continuous bases is in fact answered by the

idea that uncountable infinities are not larger in magnitude than countable infinities since both of them are infinite! Perhaps this section may be simplified, but quite frankly I am not interested in entering such fringe discussions which are extremely unimportant to our learning of quantum mechanics, and doubt the reader is either, so we can close this discussion.

8.2 Continuous Wavefunctions

Having a continuous case in our work will necessitate that some of our old summation and matrix notation seems defunct, because thinking of summing infinite terms or infinitely large matrices is not natural. However, the generalisation to infinite dimensions is in fact quite fine. Starting with infinitely sized matrices, We can consider the position basis for example to be column kets represented by

$$|x_i\rangle \leftrightarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

with the unity in the i'th row, 0 elsewhere, and infinite rows. Clearly, they constitute an orthonormal basis and the actual number of rows of the vector is already not of much concern.

Meanwhile, in continuous cases, we can simply replace the sum terms we have been using with integrals– infinite sums over continuous variables– instead! This is an idea which generalises both in continuous cases and to infinities: the idea of an integral is that we are taking the sum of values separated by increments when that increment approaches 0: thereby an infinite sum over infinitesimally different (continuous) values.

For the inner product, we have thus far used sums for our multiplications in Dirac notation: we set up a column entity, which we called a ket, and a row entity, a bra, and we simply multiplied them together, but matrix multiplication in finite dimensions is an algorithm which is essentially a sum of n

values:

$$\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i b_i.$$

The inner product we have been so far using in the Dirac notation sections is therefore the exact same idea, but with a conjugate transpose of column vector kets rather than just a transpose as the bras. That is,

$$\langle \alpha | \beta \rangle = \sum_{i=1}^{n} a_i^* b_i$$

with components $\{a_i\}$ and $\{b_i\}$ for $|\alpha\rangle$ and $|\beta\rangle$ respectively in an n-dimensional space.

The problem we now face is that, while we wish to integrate kets in some way to create a continuous inner product, we cannot do this currently as kets are in no way related to any functions, which are our natural inputs for integrals– kets are abstract vectors. However we have seen this problem of getting by an abstract ket: or, state vector, already! All we need to generate a function which exists in a bijection with state kets. We require this function to be continuous so we can integrate it.

The function which makes sense, then is the same component function we worked with before

$$|\Psi\rangle \to \psi(x).$$

This should be governed by:

- Input: Basis ket.
- Output: Component corresponding to the input basis ket.
- Domain: All kets in the basis.
- Range: Complex numbers (the components).

which changes with the input position x and gives as an output the component corresponding to basis eigenket $|x\rangle$.

Well, this is the wavefunction, we know, because we have done exactly this already in the discrete case. Here, the argument (x) represents that we are inputting position values, so this is the position wavefunction, which is only possible now because the position observable exhibits a continuous spectrum. One problem we do have with this is that, if we are trying to denote the value of the function evaluated at position 0, for example, the form becomes

$$\psi(0)$$

where it is now unclear whether or not it is a position wavefunction with position value 0 or momentum wavefunction with momentum value 0 or so on. Unfortunately, if we want to leave space for a subscript identifying separate wavefunctions apart from each other, for example in

$$|\Psi_1\rangle \leftrightarrow \psi_1(x), \;\; |\Psi_2\rangle \leftrightarrow \psi_2(x)$$

we have to sacrifice the ability to tell when the argument is absent which basis a wavefunction is expressed in. The other side of the coin is that first of all, context should always make it abundantly clear which basis everything is expressed in anyway, and secondly, reserving the same letter ψ representing wavefunctions is more valuable than alternative possibilities such as identifying different observable wavefunctions with different letters, which would be just confusing.

Now we can see easily that the position wavefunction is a continuous function since,

$$\forall x_0, \;\; \lim_{dx \to 0} \psi(x_0 + dx) = \psi(x_0).$$

This intuitively makes sense, since the component corresponding to the position $x + dx$ should approach the component as we rotate to the direction corresponding to position x. A final convincing of this should be completed by the fact that the component relates the probability of measuring a certain position, by the measurement postulate: thus as we approach one position the component (probability) of measuring positions approaching that specific position should become progressively closer to the probability of measuring that specific position. Note also that we could have

$$\psi(p)$$

which changes depending on input momentum– assuming we are working in instances of continuous momentum, which do exist– to give the corresponding component of the eigenket which is the eigenstate with that input momentum. This would simply be used if we were working in momentum space.

Now, the components of a ket $|\Psi\rangle$ in a given basis $\{|x\rangle\}$, which we will use since it is most natural, are given by

$$\langle x_i|\Psi\rangle = c_i.$$

Thus the position wavefunction is

$$\psi(x) := \langle x|\Psi\rangle.$$

which is, the component of the ket $|\Psi\rangle$ in the direction of the basis vector $|x\rangle$.

Now, we pause, because the above definition in ket form is prone to confusion, especially on passing glances, so it is important the next clarification is understood thoroughly before we move on.

The reason the function definition we have just reached looks so absolutely wrong is because on the right hand side we have what appears exactly as an inner product, which we instantly associate (rightly) with a scalar constant, but on the left we have the position wavefunction we have just learnt about as a continuous function. What is very important here is that the inner label of the bra $\langle x|$ in the inner product is meant to represent a varying index, rather than just an inner tag on a static bra obtained from a static ket. So we would have

$$\psi(0) = \langle 0|\Psi\rangle$$

which is, the component of the state $|\Psi\rangle$ corresponding to position 0 which is represented by the ket $|0\rangle$ the corresponding bra of which is $\langle 0|$. We could also have

$$\psi(L) = \langle L|\Psi\rangle$$

which is, the component according to position L.

It is not wrong for the reader to find the above somewhat unusual, since a single inner product is always a scalar constant. It therefore might seem like a bit of an abuse of notation to write

$$\langle x|\Psi\rangle$$

to be a varying quantity, since before our inner labels were stationary orders for kets and nothing more. Unfortunately, this is one point in quantum mechanics which convention has not seen fit to bother with changing notation to accommodate. The best way to deal with it is that in this book I will always use $|x\rangle$ to represent a variable position ket, $|p\rangle$ a variable momentum ket and $|E\rangle$ a variable energy ket. Otherwise, for static single position kets I will use notation such as $|x_0\rangle$ or $|x_1\rangle$ to denote these fixed points – usually, just x_0 if we only require one fixed point to compare to, as this notation is quite common in quantum mechanics. Now we can fully enunciate the action of this wavefunction is through the following steps:

1. Take the input observable value.
2. Find the eigenket corresponding to that observable value, which is a pure eigenstate with probability 1 of obtaining that value.
3. Find the inner product product of that eigenket with the state ket.
4. Output that inner product value.

8.3 Dirac Delta Function

That notational liberty of variable kets is aesthetically cumbersome, so we should demand it to be at least useful to us. It is useful, in fact, since we can finally define the integral for our abstract kets and bras:

$$\langle \Psi_1 | \Psi_2 \rangle = \int_{-\infty}^{\infty} \psi_1^*(x) \psi_2(x)\, dx.$$

Finally, we have the continuous inner product! We indeed see that the above is very much related to the discrete inner product, since instead of discrete components we now have a continuous wavefunction storing the components to deal with the fact there are infinitely many continuous eigenkets, and finally we integrate just as an integral is a continuous sum. The same varying input variable x ensures that the coefficients are matched up for both states $|\Psi_1\rangle$ and $|\Psi_2\rangle$ since each time the functions $\psi_1(x)$ and $\psi_2(x)$ give the components corresponding to the eigenket representing the inputted position, by their definition.

The integral bounds can become definite, and often do, in physical problems. An example of this fact might be if we are considering position functions on a string– in which case the variable x might be considered, and the bounds might be 0 to L– the other end of the string of length L from the end we defined to be position 0. It is clear that in this case it does not make sense to take an integral from 0 to 70000km displaced from 0, for example, if a string is say 50cm long, as we do not expect any particles we consider to be any further than 50cm (in the one dimension we are considering) from the point on the other end which we defined to be position 0. This is some sort of way to saying that we can ignore the bounds of the integrals in continuous quantum mechanics until a physical problem sets them for us.

With this given, the orthogonal condition is the same: two kets $|x_0\rangle$ and $|x_1\rangle$ (which, by the aforementioned convention we are using, represent static kets corresponding to positions x_0 and x_1 respectively) are orthogonal if

$$\langle x_0|x_1\rangle = 0.$$

However, the normalisation condition is different. To understand this, consider the completeness relation, which we expect to be:

$$\int_{-\infty}^{\infty} |x\rangle\langle x|\, dx = 1$$

where we have continued the convention of x without any subscripts representing a variable inner label. Left multiplying by $\langle x_0|$ and right multiplying by any non-basis ket $|\alpha\rangle$ gives us

$$\int_{-\infty}^{\infty} dx\, \langle x_0|x\rangle\langle x|\alpha\rangle = \langle x_0|\alpha\rangle$$

which is, the component $\psi_\alpha(x_0)$ of the ket $|\alpha\rangle$in the direction of eigenket $|x_0\rangle$ corresponding to position value x_0.

Denote now the inner product between basis vectors $|x\rangle$ and $|x_0\rangle$ as

$$\langle x|x_0\rangle := \delta(x - x_0).$$

The way of presenting the two arguments x and x_0, which are, the inputs, of the function should not throw one off: it is convention, which has a not so significant but still very much logical reason. We know that

$$\forall x \neq x_0, \quad \delta(x - x_0) = 0$$

since the position kets are all orthogonal to each other as they are non-degenerate eigenkets of a hermitian operator. Now, we can look at the above expression again:

$$\int_{-\infty}^{\infty} \langle x_0|x\rangle\langle x|\alpha\rangle \, dx = \langle x_0|\alpha\rangle \implies \int_{-\infty}^{\infty} \delta(x_0 - x)\psi_\alpha(x) \, dx = \psi_\alpha(x_0)$$

The range over negative infinity to positive infinity seems expansive, but in fact, an infinite part of it is redundant; this is because the inner product $\delta(x - x_0)$ is 0 for all $x \neq x_0$! Now it is clear we need to continue to consider infinities since we are working with continuous variables rather than discrete ones. Consider an infinitely small region $[x_0 - \Delta x, x_0 + \Delta x]$ for an infinitesimal difference Δx, which centres at x_0. In this region we can consider the integral because it is only here where we can consider $\delta(x - x_0)$ to not definitively be 0 since if x is infinitely close to x_0 we cannot just easily say $x \neq x_0$. With these new bounds,

$$\int_{x_0-\Delta x}^{x_0+\Delta x} \delta(x - x_0)\psi_\alpha(x) = \psi_\alpha(x_0)$$

since this region is the only area where we cannot say the inner product $\delta(x - x_0)$ is 0. As all values of x in the integral lie in this infinitesimal region, we can as we assume in the limit that the components $\psi_\alpha(x) = \psi_\alpha(x_0)$ since the input direction $|x\rangle$ is infinitesimally different from the fixed direction $|x_0\rangle$. This then is a specific value (not a function, despite the notation), so we can pull it out as a constant:

$$\int_{x_0-\Delta x}^{x_0+\Delta x} \delta(x - x_0)\psi_\alpha(x) \, dx = \psi_\alpha(x_0) \int_{x_0-\Delta x}^{x_0+\Delta x} \delta(x - x_0) \, dx = \psi_\alpha(x_0).$$

That is,

$$\int_{x_0-\Delta x}^{x_0+\Delta x} \delta(x - x_0) \, dx_0 = 1.$$

What implications does this have for the inner product $\delta(x - x_0)$, which we call the (Dirac) delta function? The most intuitive answer comes from using the conventional visualisation of an integral as a way to measure the area under a smooth function. We know the delta function $\delta(x - x_0)$ is 0 until it reaches an infinitely small interval around the value x_0. Yet the integral of the whole function with respect to x is 1. So if we draw a horizontal axis for varying x and a vertical axis for the value of $\delta(x - x_0)$ (with x_0 fixed), we will

get a flat line for all infinity until we get infinitely close to x_0. Yet, this as a whole must have area 1! So the infinitely small width interval close to x is the only region which contributes any area to the integral, and this whole area is 1. So, in the picture we have created, the value $\delta(x - x_0)$ in this region is the height which contributes 1 to the area despite having an infinitely small width. The only explanation therefore is that at the infinitesimally small region, the delta function has infinite value. Otherwise, the infinitely small width interval could not have any area which is not infinitely small! Then the infinitesimally small domain for x around x_0 can simply be reduced to $x = x_0$, so $\delta(x_0 - x_0) := \delta(0)$ is infinity!

We summarise with the following:

$$\delta(x - x_0) = \begin{cases} 0 \text{ if } x \neq x_0 \\ \infty \text{ if } x = x_0 \end{cases}$$

Of course, we do not like writing the infinity symbol as if it is a value very often in mathematics, so this is better put

$$\delta(x - x_0) = 0 \text{ if } x \neq x_0, \qquad \int \delta(x - x_0)\, dx = 1$$

where the bounds of integration do not matter since the function is 0 anyway until we get infinitesimally close to x_0. This will be commonplace any time we consider a continuous basis. Another way to summarise this is also in the framework of viewing the delta function as a *sampling function*, which means that

$$\int \delta(x - x_0) f(x)\, dx = f(x_0)$$

That is- because the range where x and x_0 are completely distinct vanishes, the integral only selects the value of $f(x_0)$, which changes over a continuously varying x_0, which is the same as that at point x. Finally, it is crucial to know that the delta function is real! This means that

$$\delta(x - x_0) = \langle x|x_0\rangle = \langle x_0|x\rangle^* = \delta(x_0 - x)^*$$

but as the delta function is real,

$$\delta(x_0 - x)^* = \delta(x_0 - x)$$

which altogether means that

$$\delta(x - x_0) = \delta(x_0 - x)$$

-an important point, of course. It is also very important to know that the minus sign in the function is meant to be taken literally (technically, the Delta function is a function of the difference between its two arguments, but this point is not important at all). So we will often see expressions like

$$\delta(x) = \delta(x - 0) = \delta(0 - x)$$

or

$$\delta(0) = \delta(x - x)$$

and even

$$\delta(k) = \delta(x + k - x)$$

in shorthand. A reader should not get confused by this, and should also remember that

$$\delta(k) = \delta(-k)$$

since $\delta(x + k - x) = \delta(x - (x + k))$.

8.4 Position and Momentum

In order to understand the importance of orthonormalising to the Dirac delta function as the continuous analogue of orthonormalising to the Kronecker delta in the discrete case we have beforehand been working on, we will undergo the necessary algebraic steps to work with the two canonical quantum mechanical operators. Through this, one can practice both the algebra and rationale of the Dirac delta function.

8.4.1 Position and Momentum Operators

The first step we can take using the Dirac Delta function is to confirm the action of the position operator– the operator in quantum mechanics which represents the physical variable of position– in **position space**, the space

spanned by the continuous position eigenkets. Take some function $f(x) := \langle x|f\rangle$ represented by the ket $|f\rangle$ and define

$$\hat{X}|f\rangle := |F\rangle$$

with

$$F(x_0) := \langle x_0|F\rangle.$$

Now in the usual way we can determine the action of an operator on a ket by left multiplying it by a basis bra:

$$\langle x_0|\hat{X}|f\rangle = \langle x_0|F\rangle$$

We can then insert the continuous completeness relation:

$$\int |x\rangle\langle x|\,dx = I \implies \langle x_0|\hat{X}|f\rangle \equiv \int \langle x_0|\hat{X}|x\rangle\langle x|f\rangle\,dx = \int x\langle x_0|x\rangle\langle x|f\rangle\,dx.$$

In Delta notation, this is

$$\int \langle x_0|\hat{X}|x\rangle\langle x|f\rangle\,dx = \int x\delta(x-x_0)f(x)\,dx.$$

As per usual, the delta function vanishes all the integrated terms except for when $x = x_0$. Therefore we can assume the x variable is only relevant when $x = x_0$ and thus pull it out as x_0. To finalise the result, we simply get the sampling property of the delta function.

$$\int x\delta(x-x_0)f(x)\,dx \equiv x_0\int \delta(x-x_0)f(x)\,dx = x_0f(x_0).$$

So we have

$$\langle x_0|\hat{X}|f\rangle = \langle x_0|F\rangle = F(x_0) = x_0f(x_0)$$

So

$$\hat{X}f(x_0) = F(x_0) = x_0f(x_0).$$

Since this holds true for all x_0, we can generalise this for any input of the position variable x as simply:

$$\hat{X}f(x) = xf(x)$$

in the position basis; indeed, this is the position operator we are familiar with. The matrix elements of $\hat{X}$ in the position basis are trivial:

$$\langle x_1|\hat{X}|x_0\rangle = x_0\langle x_1|x_0\rangle = x_0\delta(x_1-x_0).$$

The position and momentum operators in any basis are related solely by the commutation relation

$$[\hat{X}, \hat{P}] = i\hbar.$$

The choice one makes for the momentum operator in position space is therefore

$$\hat{P} := -i\hbar \frac{d}{dx}.$$

Solving its eigenvalue problem in the continuous position space will prove a more challenging task, as it contains a derivative. We start by considering its action on a function in the position basis:

$$\langle x_0|\hat{P}|f\rangle \equiv -i\hbar \left\langle x_0 \middle| \frac{df}{dx} \right\rangle = -i\hbar f'(x)$$

where $f'(x)$ is the component in the x direction of the derivative of f with respect to x. Expanding again with the completeness relation, we also have the alternate equations

$$\langle x_0|\hat{P}|f\rangle = \int \langle x_0|\hat{P}|x\rangle\langle x|f\rangle \, dx = \int \langle x_0|\hat{P}|x\rangle f(x) \, dx$$

So this means that

$$-i\hbar f'(x) = \langle x_0|\hat{P}|f\rangle = \int \langle x_0|\hat{P}|x\rangle f(x) \, dx.$$

The form we expect is that

$$\langle x_0|\hat{P}|x\rangle = -i\hbar\delta(x - x_0)\frac{d}{dx}.$$

Indeed, this form works with the delta function. We get:

$$\langle x_0|\hat{P}|f\rangle = \int -i\hbar\delta(x - x_0)\frac{d}{dx}f(x) \, dx = -i\hbar \int \delta(x - x_0)f'(x) \, dx$$

and this is, by the sampling property,

$$-i\hbar f'(x)\Big|_{x=x_0} = -i\hbar f'(x_0)$$

so we get

$$\langle x_0|\hat{P}|f\rangle = \hat{P}f(x_0) = -i\hbar f'(x_0)$$

as required. So we have the matrix element

$$\langle x_0|\hat{P}|x\rangle = -i\hbar\delta(x - x_0)\frac{d}{dx}.$$

Interestingly, this is in fact

$$\langle x_0|\hat{P}|x\rangle = -i\hbar\delta'(x - x_0),$$

that is, $-i\hbar$ multiplied by the derivative of $\delta(x - x_0)$ with respect to its first argument (which here is x)! The rule for the derivatives of Delta functions, which are too mysterious and challenging to warrant their own proofs for now but which must be stated, is that

$$\frac{d^{(n)}\delta(x - x_0)}{dx^{(n)}} = \delta'^{(n)}(x - x_0) = \delta(x - x_0)\frac{d^{(n)}}{dx^{(n)}}$$

where the exponent n represents the order of the derivative. Do note that the derivative is with respect to x here because here x is the first argument of the delta function and the argument which is variable.

It is equally important to know that a very different definition for the operators and their matrix elements occurs if we are in a different basis- for example, the momentum space. Our original definition in Postulate 6 was that the position operator was $\hat{X} = x$ and the momentum operator was $\hat{P} = -i\hbar\frac{d}{dx}$, but this only holds true in position space. Position space is generally the space used, as aforementioned in chapter 5, as it is in physical reality, and it usually is easier to express keep the potential $V(x)$, a function of position, in the position basis than it is to find its form in momentum space. Nevertheless, momentum space can also for example be considered.

Now observe the way in which we derived the action of the position operator in position space. It should be convincing to the reader that if we were to replace the position operator with the momentum operator, and multiply it by a momentum eigenbra, and use the completeness relation but with momentum eigenkets, we will observe the exact same result! Nothing at all is contributed by the fact that it was the position operator we were discussing,

other than that it was the position operator whose eigenbasis was spanning the space above while we were trying to express its action on constituent kets. Therefore, if we consider any other continuous operator in a space spanned by its orthonormal basis, we should obtain the exact same result– but simply expressed for a different physical variable!

If one is convinced that we *derived* the action of the position operator in its own space simply by using continuous relations and manipulation with its own eigenkets, they should then be convinced that in momentum space we have:

$$\hat{P}f(p) = pf(p).$$

Its matrix elements in this basis are easy as well:

$$\langle p'|\hat{P}|p\rangle = p\langle p'|p\rangle = p\delta(p' - p).$$

8.4.2 Position and Momentum eigenfunctions

Fruitful (but difficult) discussions come out of considering the eigenfunctions of position and momentum. We will work in the position space, as that is sufficient to provide the necessary discourse and is the most common space to work in.

One might assume that the eigenvalues of $\hat{X}$ in its own space are natural. We can denote an eigenvector as $\xi(x)$ and the corresponding eigenvalue as x_0. Then we have

$$\hat{X}\xi(x) = x\xi(x) = x_0\xi(x).$$

Consider this now carefully. We have

$$x\xi(x) = x_0\xi(x)$$

where x is a continuous variable and x_0 is a single constant eigenvalue! Clearly something is wrong. It is impossible for a single eigenvalue to be multiple values of x at the same time; if it was a function instead, it would not be viable as an eigenvalue and therefore $\xi(x)$ would also not be an eigenvector.

It turns out this problem is fixed if we make the eigenvector $\xi(x)$ a special type of function. Of course, by now one should anticipate this is the

Dirac delta function (we can see how useful it is), with arguments x_0 and x. That way, if we have

$$\hat{X}\delta(x_0 - x) = x\delta(x_0 - x) = x_0\delta(x_0 - x)$$

we will see that the function does the work for us! We get a crude $0 = 0$ for all the values of continuous variable x which are not x_0, and then for $x = x_0$ clearly the eigenvalue and eigenvector conditions are satisfied. We will rewrite this as

$$\hat{X}\delta(x - x_0) = x_0\delta(x - x_0)$$

as it is more conventional to write $\delta(x-x_0)$ than $\delta(x_0-x)$, even though they are the same thing.

The momentum eigenvalue problem in position space is more important, and difficult. The problem is

$$\hat{P}\phi(x) = -i\hbar\frac{d}{dx}\phi(x) = p\phi(x)$$

for an arbitrary eigenvector $\phi(x)$ corresponding to eigenvalue p. We can tell that immediately the problem we had with the position eigenvectors in position space isn't going to be prominent here simply due to the form of the operator; indeed, we will not need the delta function. Instead, what one finds is that this is very simply a recognisable first order differential equation.

$$\begin{aligned} p\phi(x) &= -i\hbar\frac{d}{dx}\phi(x) \\ \Rightarrow \quad \frac{d\phi}{dx} &= ip/\hbar\phi(x) \end{aligned}$$

and this is an equation of the form

$$\frac{dy}{dx} = ky$$

with general solution $y = Ae^{kx}$. So matching these two equations we have the comparisons $y = \phi(x)$, $x = x$ and $k = ip/\hbar$. Therefore, the solution to the eigenvector problem is

$$\phi(x) = Ae^{ipx/\hbar}.$$

for some arbitrary constant A. It should be noted that $i, \hbar$ are both constants, and p is the eigenvalue corresponding to the eigenvector, which is not a single constant unless we are only concerned with a single eigenvector. Therefore, x is the only variable which changes the value of the position space momentum eigenfunction as it is itself varied, which is why we can call it a function of the position variable x. Now, it might seem like our job is done in this section, but the form of the momentum eigenvectors bring significant imperfections to the Hilbert space we have so far seen as quite perfect for our job of describing physical reality. We must deal with these in any valid discussion of quantum mechanics.

As mentioned in the section on infinities, an exponential function rises exponentially to infinity as x goes to infinity (when its exponent is positive, as it is here). The reader should therefore realise that, no matter what constant A we try to multiply it with, we will never be able to gain a finite integral over all space and therefore position x. As there is no normalisation which produces a finite norm for a momentum eigenstate, we can only do the next best thing, which is try to normalise it in a way which encapsulates this infinite behaviour while being still mainly serviceable for the needs of algebraic manipulations. Of course, this can be best done through the Dirac delta function. We have seen how useful it is, in that its derivatives are defined, it has secondary properties like sampling which greatly simplify algebra with it, and it is very commonly seen across what we have already done. Thus, for such non-finite norm vectors we want to use the process of *normalising to the Dirac delta function*, the continuous analogue of normalising to unity!

In the case of momentum eigenvectors, a constant A is already set up in the general solution. So we can try to find a constant A which normalises the momentum eigenvectors to the Dirac delta function. In arbitrary terms, we have, still in the position basis,

$$\langle p|p'\rangle = \int \langle p|x\rangle\langle x|p'\rangle \, dx$$

and we would like $\langle p|p'\rangle$ to be equal to $\delta(p - p')$ after we have modified $|p\rangle$ (we can use the variable label p since it should hold for all eigenkets) by

multiplying it by some constant. Currently the above is

$$\begin{aligned}\langle p|p'\rangle &= \int \langle p|x\rangle\langle x|p'\rangle\, dx = \int \langle x|p\rangle^*\langle x|p'\rangle\, dx \\ &= \int \psi_p^*(x)\psi_{p'}(x)\, dx = \int A^* e^{-ipx/\hbar} A e^{ip'x/\hbar}\, dx \\ &= |A|^2 \int e^{-ix(p-p')/\hbar}\, dx.\end{aligned}$$

Unfortunately, at this point we cannot use any mathematics we know or will be able to quickly learn to move forwards, as the answer lies in a relationship given by Fourier transforms, which are too advanced for this book. Therefore a relationship will just have to be stated and taken as given. This is the relationship

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ik(x-x_0)}\, dk = \delta(x - x_0).$$

It looks haphazard, but at the heart of it lies the connections between complex numbers, exponentials and pi, of which an example is the Euler formula. Any further explanation would be meaningless without the reader understanding Fourier transforms (where from the relationship is derived) in the first place, and that is hardly assumed knowledge here.

This is reasonably closely related to our momentum eigenstate normalisation question, with a bridge step. We had:

$$|A|^2 \int e^{-ix(p-p')/\hbar}\, dx$$

for some normalisation constant A we are trying to find. We can rewrite this as

$$|A|^2 \int e^{-i(\frac{p}{\hbar}-\frac{p'}{\hbar})x}\, dx,$$

which is extremely close to the expression above. The only lemma we need to is that $\delta(\frac{x}{a}) = a\delta(x)$ for some constant a. Now we can use the relations to get

$$|A|^2 \int e^{-ix(p-p')/\hbar}\, dx = \frac{|A|^2}{2\pi}\delta\left(\frac{1}{\hbar}(p-p')\right)$$

and by the lemma this is

$$\frac{|A|^2\hbar}{2\pi}\delta(p-p').$$

Finally, we can see what constant we want A to be! Simply $A = \frac{1}{\sqrt{2\pi\hbar}}$ will work, and, retracing our steps and replacing A with $(2\pi\hbar)^{-1/2}$ we will get

$$|p\rangle \leftrightarrow \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar} \implies \langle p|p'\rangle = \delta(p - p').$$

This has taken a lot of labour and a couple of steps we can only take for granted without more advanced mathematical tools: but one should now be able to fully understand that orthonormalising to the Dirac delta function is the continuous analogue of orthonormalising to the Kronecker delta in the discrete case, and that both are necessary, possible, and useful.

8.5 Probability Distribution Functions

The state problem was said at its introduction in this book to be a problem of encapsulation and extraction. The question of extraction is answered by observable operators and their eigenbases, and elaborated on by the Compatibility Theorem and Heisenberg Uncertainty Principle, which result from those observable and measurement postulates. The question of encapsulation has been already largely answered by state vectors, but both elegance and historical justice can be demonstrated with the following formalisation, which will show that the continuous wavefunction can be interpreted as a probability distribution function which therefore encapsulates not only all the possible measurements, but also their probabilities. This is naturally a direct mirror image of the probability mass function we obtained for the discrete case wavefunction. After that, the only difficulty– though it will prove itself to be much more diverse and challenging than any other– would be to solve for the state vector in some basis given the conditions of a physical problem we are given.

We know how to extract probabilities from our representation of a state:

$$P(\alpha_i) = |(\alpha_i, \Psi)|^2$$

if α_i is some orthonormal eigenvector. It is important to remember that there are bases of the state space which do not consist of eigenvectors of some operator, but since we have already seen that we are exclusively interested in observable eigenbases when we work in quantum mechanics, we might often

write eigenvector instead of vector with the implicit assumption that the basis vector we are using would be an eigenbasis vector. We also know that

$$(\alpha_i, \Psi) = c_i,$$

the component of the state vector in the orthonormal basis $\{\alpha_i\}$. Therefore the components of the state vector in some observable basis are the links to probabilities in the state problem. We call them **probability amplitudes**, and the modulus squared of these probability amplitudes are called **probability densities**. If we work in discrete cases, probability densities, called probability masses in the discrete case, are simply synonymous with probabilities, as the postulate shows. Very minor differences exist in the continuous case, which we now have the correct definitions to tackle as well.

The generalisation to continuous dimensions of probability mass functions are well covered in mathematics already. The problem we have is that continuity implies infinite eigenstates, and therefore, from the perspective of probability, infinite events. If there are infinite events, then all of them must have technically have 0 probability because we cannot pin a probability down to a single value when there is a value which corresponds to shifting that value by an infinitely small increment.

Again, this is pure mathematics rather than quantum mechanics speaking. It is convention to call the probability distribution functions **probability density functions** in the continuous case. Since it is a continuous function, we cannot take a certain value for a probability as discussed, but we can integrate it between certain bounds!

The integral

$$\int_{x=a}^{x=b} |\psi(x)|^2 \, dx$$

with a continuous wavefunction, here the position wavefunction, is the probability that the state is measured to be somewhere between the eigenstates $|a\rangle$ and $|b\rangle$ corresponding to position values a and b. Thus we have a way to find the probabilities we need now for the continuous case as well, and see why wavefunctions are probability density functions with respect to certain measurements of observables.

The state problem, which we left as finished for the discrete case, is now complete for the continuous case as well. The physical state corresponds to the state vector, which can be transformed into wavefunctions by eigenvector inner products. The resulting functions are either probability mass functions, in the discrete case, or probability distribution functions, in the continuous case. Finally, I am justified in calling wavefunctions 'tangible' when I sounded strange beforehand. More importantly, we are done with all the conceptual learning we will be covering in this book!

8.6 Exercises from Chapter 7

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

Chapter 9

Chapter 9: Problems in Quantum Mechanics

This chapter, and the next, will practice all the concepts we have learnt with physical problems. The algebraic labour involved will be extensive, but how could it be otherwise? I have kept all the solutions in Dirac notation, because this is conventional and because manipulation with Dirac notation is more powerful and fruitful.

9.1 The Propagator

We have mentioned how one can use energy eigenstates to conveniently find solutions for the Schrödinger equation. However, beyond this algebraic method there is another method, the method of the propagator, which is used ubiquitously by those more advanced in quantum mechanics. In this book, it might seem even at the end counterintuitive to have learnt both the energy eigenstate method and the propagator method, but the crucial point is that for more difficult problems only the latter is viable; while we might not have the full mathematical tools to realise how useful it is, no mature treatment of quantum mechanics is complete without it, and we can still use it for the simpler problems we cover even if its formulation might seem to contribute nothing for these basic examples. It is also, mathematics aside, a conceptually valuable thing to study even at an expository level.

At time t_0 the state must be able to be represented by a ket in the ket

space. At any other time t the state must also be able to be represented by a ket in the ket space. One therefore might be led to consider whether there is an operator which governs the mapping of one state ket to another state ket given some input time $t - t_0$ which has passed.

There is, and it is called the propagator (or time evolution operator). More importantly, the operator turns out to be the same over all time **given the same Hamiltonian** (i.e, given the system is not perturbed). We can try to formulate it, as we know some conditions we expect to be fulfilled.

Denote the operator $\mathcal{U}(t, t_0)$ to be the propagator which carries a state ket from time t_0 to a ket at time t. The first condition we expect the propagator to satisfy is the property of composition:

$$\forall\, t_0 \leq t_1 \leq t_2, \quad \mathcal{U}(t_2, t_0) = \mathcal{U}(t_2, t_1)\mathcal{U}(t_1, t_0).$$

This is because the propagator of a system carrying a ket from t_0 to t_1 to t_2 should be an equivalent transformation to carrying a ket from t_0 to t_2 directly since the start and end kets are the same.

The second condition is subtle but important. Take a ket $|\Psi, t_0\rangle$, We expect that for a proper physical ket it is normalised, so that the total sum of probabilities is always 1 for any measurements of a given observable. Now, if we consider

$$\mathcal{U}(t, t_0)|\Psi, t_0\rangle := |\Psi, t\rangle,$$

then the norm at time t of the state ket becomes

$$\langle \Psi, t|\Psi, t\rangle = \langle \Psi, t_0|\, \mathcal{U}^\dagger(t, t_0)\mathcal{U}(t, t_0)\, |\Psi, t_0\rangle.$$

We expect that the ket at times IS still normalised, or the sums of probabilities for measurements will not equal 1. Therefore we have

$$\langle \Psi, t_0|\Psi, t_0\rangle = \langle \Psi, t|\Psi, t\rangle = 1 = \langle \Psi, t_0|\, \mathcal{U}^\dagger(t, t_0)\mathcal{U}(t, t_0)\, |\Psi, t_0\rangle.$$

This means we must have:

$$\mathcal{U}^\dagger(t, t_0)\mathcal{U}(t, t_0) = 1.$$

Otherwise stated, the propagator is unitary! In fact this is a requirement for operators which map a physical state onto another physical state, due

to the necessity of **conservation of probability**– really, conservation of normalisation–, which unitary operators always fulfil because they preserve the norm of any ket they act on.

Finally, we expect that

$$\lim_{dt\to 0} \mathcal{U}(t_0 + dt, t_0) = 1$$

(the identity operator), due to the continuity of time. Now, we do know that the Schrödinger Equation must still apply- in other words, we know that

$$i\hbar\frac{\partial}{\partial t}|\Psi, t\rangle = \hat{H}|\Psi, t\rangle.$$

Assume the state $|\Psi, t_0\rangle$ was a precedent state to $|\Psi, t\rangle$. We therefore have

$$i\hbar\frac{\partial}{\partial t}\mathcal{U}(t, t_0)|\Psi, t_0\rangle = \hat{H}\mathcal{U}(t, t_0)|\Psi, t_0\rangle$$

so we can equate the two operators since it holds true for any $|\Psi, t_0\rangle$:

$$i\hbar\frac{\partial}{\partial t}\mathcal{U}(t, t_0) = \hat{H}\mathcal{U}(t, t_0).$$

This is called the Schrödinger Equation for the time evolution operator.

We would like there to be a very simple way to find the propagator, because having an explicit representation of the propagator is sufficient to solving the time evolution problem (we can apply it to any initial state $|\Psi, t_0\rangle$). We can try to induce its form through looking at equations we know must hold, since we understand its usage and what properties it must have.

The reader might be relieved to know our starting point is the energy eigenfunction method of solving the Schrödinger Equation, which is not incorrect at all (we derived it very soundly from the Schödinger Equation, and the reader may go back to review that proof if they need do so before continuing) and need not be discarded. In our function formulation, we had

$$\psi_t^{(n)} := e^{-iE_n t/\hbar}\varepsilon_n$$

as the stationary states of any given system- in other words, where time evolution does not change the states at all. In our Dirac notation this is

$$|\Psi_n, t\rangle = e^{-iE_n t/\hbar}|E_n\rangle$$

where we have labelled kets by their eigenvalues as is customary. The derivation

$$\Psi_t = \sum_n (\varepsilon_n, \Psi_0)\psi_t^{(n)}$$

was made where Ψ_t is any arbitrary state at time t and Ψ_0 was the state at time 0. We translate this to Dirac notation as well:

$$|\Psi, t\rangle = \sum_n |\Psi_n\rangle\langle E_n|\Psi, 0\rangle = \sum_n e^{-iE_n t/\hbar}|E_n\rangle\langle E_n|\Psi, 0\rangle.$$

However, this allows us to induce the form of the propagator:

$$|\Psi, t\rangle = \sum_n e^{-iE_n t/\hbar}|E_n\rangle\langle E_n|\Psi, 0\rangle = \left(\sum_n e^{-iE_n t/\hbar}|E_n\rangle\langle E_n|\right)|\Psi, 0\rangle$$

and

$$|\Psi, t\rangle = \mathcal{U}(t, 0)|\Psi, 0\rangle$$

so therefore

$$\mathcal{U}(t, 0) = \sum_n e^{-iE_n t/\hbar}|E_n\rangle\langle E_n| \ .$$

This form of the propagator of course applies when the energy eigenkets are discrete: a phenomenon which does often occur in problems, as we will see shortly. It should be noted that $e^{-iE_n t/\hbar}$ is not a constant, as its value depends on the eigenvalue E_n which varies over the different indexes n, so trying to pull it out of the sum term is invalid. Furthermore, the continuous analog (for continuous energies, which also occur) should be equally clear, as we can simply take an integral:

$$\mathcal{U}(t, 0) = \int_{-\infty}^{\infty} e^{-iE_n t/\hbar}|E_n\rangle\langle E_n| \, dn.$$

In reality, using n as an index for a continuously varying entity, and integrating with respect to that contrived n, is somewhat hideous. However, the idea is that we integrate over the changing values of E_n, which correspond to the labelled kets $|E_n\rangle$. Finally, we will not consider time varying Hamiltonians as they are extremely difficult.

9.1.1 The Free Particle Propagator

We can now formulate the propagator for the free particle, whose solution we studied without a propagator already in chapter 5. The energy eigenkets

of the Hamiltonian, which commutes with the momentum operator, are of the form

$$|E_n;+\rangle = |p = \sqrt{2mE_n}\,\rangle$$

and

$$|E_n;-\rangle = |p = -\sqrt{2mE_n}\,\rangle.$$

The + and − inner labels summarise the fact that these energy eigenvalues are degenerate: they apply to two distinct eigenkets- but with different momenta corresponding to them (the reader should probably revise the section on the Free Particle to recall this). We therefore label the eigenkets by their nondegenerate momenta instead of using the natural E label. This means for the propagator:

$$\mathcal{U}(t,0) = \int_{-\infty}^{\infty} e^{-iE_n t/\hbar} |E_n\rangle\langle E_n|\, dn$$

we can write it for the free particle as

$$\mathcal{U}(t,0) = \int_{-\infty}^{\infty} e^{-ip^2 t/2m\hbar} |p\rangle\langle p|\, dp.$$

Note that, while we before were unscrupulously using an index n to vary over the continuous eigenkets since it made the analog from discrete to continuous clearer, we have replaced that with the continuous index p, which makes far more sense in any case. The term $p^2/2m$ has replaced the term E_n since it is the energy value corresponding to the eigenstate with eigenmomentum p.

We now may choose a basis to work in to evaluate the propagator elements. Evaluating the position space matrix elements

$$\langle x|\mathcal{U}(t,0)|x_0\rangle$$

makes more sense than evaluating the momentum space matrix elements,

$$\langle p|\mathcal{U}(t,0)|p'\rangle$$

since the latter clearly is going to involve delta functions and we might prefer to avoid this for an operator we want to readily apply. In the position space the elements are

$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \int_{-\infty}^{\infty} e^{-ip^2 t/2m\hbar} \langle x|p\rangle\langle p|x_0\rangle\, dp$$

and this can be readily written in terms of the position space momentum eigenfunctions:

$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \int_{-\infty}^{\infty} e^{-ip^2t/2m\hbar}\psi_p(x)\psi_p^*(x_0)\,dp.$$

The eigenfunctions are given by

$$\psi_p(x) = \frac{1}{\sqrt{2\pi\hbar}}e^{ipx}$$

so the above is

$$\begin{aligned}
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \int_{-\infty}^{\infty} e^{-ip^2t/2m\hbar}\psi_p(x)\psi_p^*(x_0)\,dp \\
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \int_{-\infty}^{\infty} e^{-ip^2t/2m\hbar}\left(\frac{1}{\sqrt{2\pi\hbar}}e^{ipx/\hbar}\right)\left(\frac{1}{\sqrt{2\pi\hbar}}e^{-ipx_0/\hbar}\right)dp \\
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{-ip^2t/2m\hbar}e^{ip(x-x_0)/\hbar}\,dp \\
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{-(it/2m\hbar)p^2+(i(x-x_0)/\hbar)p}\,dp
\end{aligned}$$

While to the inexperienced mathematician this integral is not straightforward, in the wider scope of things this integral is in fact trivial, because of the relatively simple fact we can quote that

$$\int_{-\infty}^{\infty} e^{-ax^2+bx}\,dx = e^{b^2/4a}\sqrt{\frac{\pi}{a}}.$$

Plugging in $a := it/2m\hbar$ and $b := i(x-x_0)/\hbar$, and changing the integral to be with respect to p rather than x, we get:

$$\begin{aligned}
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} e^{-(it/2m\hbar)p^2+(i(x-x_0)/\hbar)p}\,dp \\
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \frac{1}{2\pi\hbar}e^{i(x-x_0)/\hbar)^2/4(it/2m\hbar)}\sqrt{\frac{\pi}{it/2m\hbar}} \\
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \left(\frac{1}{2\pi\hbar}\right)e^{-m(x-x_0)/2it\hbar}\sqrt{\frac{2m\pi\hbar}{it}} \\
\langle x|\mathcal{U}(t,0)|x_0\rangle &= \sqrt{\frac{m}{2i\pi\hbar t}}\,e^{im(x-x_0)/2\hbar t}\ .
\end{aligned}$$

It appears laborious, but actually requires neither advanced mathematical skills nor obscure mathematical facts to evaluate. The benefit is now we can plug in any state vector Ψ and we will have a perfect understanding of how it will evolve with time. Thus such a process (with proof of the momentum energy relationship) would be sufficient to solve the problem of the free particle.

To prepare for the next section, we will look at a small modification of the free particle question, and see that it vastly changes the result we get so far as to even discretise the energy spectrum. We will from now onwards continue to use the propagator, as it is ubiquitous in post-expository quantum mechanics texts.

9.2 Particle on an Ellipse∗

Let's start with an interesting modification to the free particle problem. Consider the ellipse below (which could be a circle– the difference does not here matter):

(1,1) ellipse (4 and 2); [black] (1,-1) circle (4pt) node[anchor=south east]; at (1,-1.75) x;

Say that a free particle $\mathfrak{p}$ is confined to be on that ellipse at all times. Consider position x as the particle's starting point, and imagine the particle moving anticlockwise around the ellipse infinitely many times. Now we can draw a helpful visual representation of the linear distance $\mathfrak{p}$ has travelled:

(0,1) – (0,-0.25); (0,0) – (11,0); (3,0) – (3,-0.25); (6,0) – (6, -0.25); (9,0) – (9, -0.25); at (3, -0.5) L; at (6, -0.5) $2L$; at (9, -0.5) $3L$; at (0, -0.5) 0;

...and so on infinitely many times. Take L to be the perimeter of the ellipse. Then it is clear that after moving in a perfect cycle for the length of one perimeter of the ellipse the particle will end back at position x once more. In other words, we write:

$$x \sim x + L$$

This means that the position x is equivalent to position $x + L$, and the relation carries on until infinity as $x + L \sim x + 2L$ and so on.

Now, considering the the system using Schrodinger's equation, we expect

$$\psi(x) = \psi(x + L)$$

for all positions x. Now, the free particle time independent Schrödinger Equation is familiar to us:

$$\hat{H}\Psi = \frac{\hat{P}^2}{2m}\Psi = E\Psi$$

and we also know $\psi(x) = \psi(x + L)$. Note that the (x) indicates we are working in position space, which makes perfect sense considering the interesting cyclic behaviour of varying position along the ellipse meaning we want x to be the central changing variable.

We can come to a solution for the particle on an ellipse quite organically. We first note that due to the form of the Hamiltonian energy and momentum must still be compatible observables, since neither operators are changed by the periodic behaviour bestowed by the elliptical setup. Therefore, the energy eigenkets we derived before are still correct:

$$|E; P^+\rangle = |p = \sqrt{2mE}\,\rangle, \quad |E; P^-\rangle = |p = -\sqrt{2mE}\,\rangle$$

for each E. We again choose the nondegenerate momentum eigenkets to represent the energy spectrum, though we do not know whether it is discrete or continuous yet. The solution comes from noting facts we already know. The momentum eigenkets, which are also energy eigenkets, correspond to:

$$|p\rangle \leftrightarrow \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}$$

and this seems innocuous until we remember the boundary condition:

$$\Psi(x) = \Psi(x + L)$$

That is, due to there being no external potential, we do not expect the wavefunction to change when it travels a length of the perimeter of the ellipse and returns to the same position. We therefore expect the same bounds

to apply to the energy eigenkets- and similarly, the momentum eigenkets. However, this means we expect:

$$|p; x = x\rangle \leftrightarrow \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar} = |p; x = x + L\rangle \leftrightarrow \frac{1}{\sqrt{2\pi\hbar}} e^{ip(x+L)/\hbar}.$$

We then expect

$$\frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar} = \frac{1}{\sqrt{2\pi\hbar}} e^{ip(x+L)/\hbar} \implies e^{i(p/\hbar)x} = e^{i(p/\hbar)x} e^{i(p/\hbar)L}$$

and so we must have

$$e^{i(p/\hbar)L} = 1$$

This is rather dramatic. By Euler's equation we get:

$$e^{ix} = \cos(x) + i\sin(x)$$

Therefore

$$e^{i(p/\hbar)L} = 1 \Rightarrow \ \cos((p/\hbar)L) + i\sin((p/\hbar)L) = 1$$

We need to get rid of the sine function part which is multiplied by an imaginary unit, since the cosine function is real-valued, the sine function is real valued, and 1 is a real number, so we must have $\sin((p/\hbar)L) = 0$ or we would get a complex but not real left hand side and a real right hand side: a contradiction. Then to finish we need $\cos(p/\hbar)L = 1$, which means

$$pL/\hbar = 2n\pi, \ \ n \in \mathbb{Z}.$$

We see now that the above equation is quantized- indexed by the integers n. Therefore momentum p is also quantized and therefore energy must also be quantised as the energy eigenvalues are equal to $p^2/2m$. Such is a very common theme in quantum mechanics: $\sin(n\pi) = 0$ and $\cos 2n\pi = 1$ for $n \in Z$, which indexes by integers n when there is boundary behaviour causing periodic recurrences. Now we can index the eigenvalues with n, and summarise that:

$$p_n = \frac{2\pi n\hbar}{L}, \ \ n \in Z.$$

Then, the eigenenergies must be:

$$E_n = (p_n)^2/2m = \frac{2\pi^2 n^2 \hbar^2}{mL^2}.$$

Now we want to evaluate the new propagator for this free particle on an ellipse. To do this we could solve the time independent Schrödinger to find the energy eigenstates. However, we can also simply solve the momentum eigenvalue equations, which are a bit simpler, since we already know that

$$\hat{P}|\psi_p\rangle = p|\psi_p\rangle \implies \psi_p(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}.$$

Plugging in the momentum eigenvalues p_n and indexing by n, we have

$$\psi_p^{(n)}(x) = \frac{1}{\sqrt{2\pi\hbar}} e^{2\pi nix/L}$$

which are also the position space energy eigenstates corresponding to the energy eigenvalues $E_{|n|}$. The propagator can then be expressed:

$$\mathcal{U}(t,0) = \sum_n e^{-iE_n t/\hbar} |E_n\rangle\langle E_n|$$

and we are working with discrete energy eigenkets so we can use this form of the propagator as well. The position space elements should be clear:

$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \sum_n e^{-iE_n t/\hbar} \langle x|E_n\rangle\langle E_n|x_0\rangle$$
$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \sum_n e^{-iE_n t/\hbar} E_n(x) E_n^*(x_0).$$

Substituting in the energy eigenstates

$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \sum_n e^{-iE_n t/\hbar} \left(\frac{1}{\sqrt{2\pi\hbar}} e^{2\pi nix/L}\right)\left(\frac{1}{\sqrt{2\pi\hbar}} e^{-2\pi nix_0/L}\right)$$

and substituting the energy eigenvalue E_n,

$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \frac{1}{2\pi\hbar} \sum_n e^{-2i\pi^2 n^2 \hbar t/mL^2} e^{2\pi nix/L} e^{-2\pi nix_0/L}$$

simplifying,

$$\langle x|\mathcal{U}(t,0)|x_0\rangle = \frac{1}{2\pi\hbar} \sum_n e^{(2\pi in)/L[\pi n\hbar t/mL]}.$$

we see that the full set of solutions to Schrodinger's Equation for a free particle on a circle can be categorised by:

$$\Psi_n(x) = Ne^{ik_n x}$$

where N was the normalisation constant assumed earlier when we eliminated $\int_0^L \psi^*(x)\psi(x)$. Its value can calculated easily:

$$\begin{aligned}
&\Psi_n(x) = Ne^{ik_n x}, \ \int_0^L \psi_n^*(x)\psi_n(x) = 1 \\
\Rightarrow\ &\int_0^L Ne^{ikx}Ne^{-ikx} = 1 \\
\Rightarrow\ &N^2\int_0^L 1 = 1 \\
\Rightarrow\ &N^2\int \frac{d}{dx}(x+c) = 1 \\
\Rightarrow\ &N^2\Big[x+c\Big]_0^L = 1 \Rightarrow\ \ N^2L = 1 \Rightarrow\ \ N = \frac{1}{\sqrt{L}}
\end{aligned}$$

Overall, we have:

$$\Psi_n(x) = \frac{1}{\sqrt{L}}e^{ik_n x}$$

and since $k_n L = 2\pi n \Rightarrow\ \ k_n = \frac{2\pi n}{L}$, we can also write this as:

$$\Psi_n(x) = \frac{1}{\sqrt{L}}e^{\frac{2\pi n i x}{L}}.$$

This is a normalised set of wavefunctions due to the normalisation coefficient, but is it an orthogonal set too? We can verify this.
Furthermore, we also know that as $V(x) = 0$ then associated energies indexed by integers n are:

$$E_n = \frac{\hat{p}^2}{2m} = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2}{2m}\left(\frac{2\pi n}{L}\right)^2$$

so they are

$$E_n = \frac{2\pi^2\hbar^2 n^2}{mL^2}$$

Note that despite the fact n can be negative, corresponding to negative momentum, the n^2 term in the equation for E ensures that $E > 0$ as we have earlier shown. Clearly there are infinite energy eigenstates as there are infinite integers n to index E_n.

However, since E_n is a function of n^2 it is also clear that all the energy eigenstates for a free particle on an ellipse can correspond to both ψ_n and ψ_{-n} which both have energy E_n, save for E_0. We say that all these non-zero energy eigenstates are degenerate: they can correspond to multiple wavefunctions, here 2. The temptation would be to say that we have an issue herein since that would make ψ_n and ψ_{-n} indistinguishable from the point of view of energy: but of course we very easily know that what distinguishes ψ_n and ψ_{-n} is that they have different momenta.

9.3 Bound States

In the last problem, we made one modification– or really, added one constraint– to the free particle problem, and found this discretised the energy spectrum immediately. Physically, the constraint seems rather significant, given that we expected the particle's motion to loop around in infinite cycles. Mathematically, however, it was only equivalent to adding position-wise periodicity. There exists this question now of how that discretisation occurred in the first place, and specifically, we can point to a specific class of problems which will be valuable to our understanding of this: those problems of bound states.

A General Discussion

A bound state in quantum mechanics is a state where we have the relationship

$$|x| \to \infty \implies \Psi(x) \longrightarrow 0.$$

That is, the wavefunction is focused on one point, which we can call position 0, and the further away we get from that point the smaller the value $\Psi(x)$ (the probability density corresponding to that point) is. The most basic example of this would occur if we had some barrier which prevented or made it practically very difficult for a particle trying to escape outside it. Of course, on a microscopic scale we cannot speak of stone walls or physical barriers, but we will see that barriers nevertheless can exist based on the potential of the system. We start by introducing a fact which exists in classical mechanics and stays valid in quantum mechanics- that a particle cannot have energy less than the minimum value of the potential of the system at any point. The proof is as follows: we know that the Hamiltonian is given by the formula

$$\hat{H} = \frac{\hat{P}}{2m} + V(x).$$

Now pick a normalised energy eigenket $\psi_E(x)$ which has energy E. We have:

$$\langle \psi_E | \hat{H} | \psi_E \rangle = E \langle \psi_E | \psi_E \rangle = E$$

but we also have

$$\langle \psi_E | \hat{H} | \psi_E \rangle = \langle \psi_E | \hat{T} + V(x) | \Psi_E \rangle = \langle \psi_E | \hat{T} | \psi_E \rangle + \langle \psi_E | V(x) | \Psi_E \rangle$$

and for some minimum value of $V(x)$ we have

$$\langle \psi_E | \hat{H} | \psi_E \rangle = \langle \psi_E | \hat{T} | \psi_E \rangle + \langle \psi_E | V(x) | \Psi_E \rangle \geq \langle \psi_E | \hat{T} | \psi_E \rangle + V_{\min} \langle \psi_E | \psi_E \rangle$$

that is,

$$E \geq \langle \psi_E | \hat{T} | \psi_E \rangle + V_{\min}$$

using the algebraic form of the momentum operator, the Schrödinger Equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + (V(x) - E)\Psi = 0.$$

If $E \geq V(x)$ for some point x then $(V(x) - E)$ will clearly be nonpositive. However, at the moment $E < V(x)$, we must get the condition that $V(x) - E$ is positive. If we put denote this difference as $\Delta > 0$, then we get the equation

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + \Delta\Psi = 0 \implies \frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} = \Delta\Psi.$$

This is comparable to the differential equation , whose solution leaks out to infinity and does not fit the definition of a bound state. Therefore in any bound state we can never have energy less than the potential of the system; furthermore, this means that no energy eigenstates exist whose corresponding eigenvalues are less than the minimum value of the potential at some point.

We can now see how we can create a bound state: by setting regions where the potential is higher than the energy of the state and where therefore the wavefunction must disappear. In particular, a region with infinite potential must immediately have the wavefunction vanish as it cannot have an energy greater than infinity. Let us use this to investigate a classic introductory problem in quantum mechanics which involves a bound state, the infinite square well.

9.4 Particle in an Infinite Square Well

Examine the following diagram:

[-to, line width = 0.5mm] (-2.5,0) – (-2.5,4); [-to, line width = 0.5mm] (2.5,0) – (2.5,4); [line width = 0.5mm] (-3,0) – (3,0); [thin,pattern=north east lines, pattern color= black] (3,0) rectangle (2.5,3.85); [thin,pattern=north east lines, pattern color= black] (-2.5,0) rectangle (-3,3.85); at (-2.5, -0.5) $x = 0$; at (2.5, -0.5) $x = L$;

At the shaded regions the potential is defined to be infinite. But between positions $x = 0$ and $x = L$ it is defined as $V(x) = 0$. Due to the fact that the energy of a particle must be greater than the potential it is experiencing, we have two conditions that within the well it must have positive energy, and outside its wavefunction must vanish, so the lines $x = 0$ and $x = L$ act as hard walls which bind the state to being between positions 0 and L.

We can list the properties of the system given by the boundary conditions:

1. $\Psi(x)$ does not exist if $x > L$ or $x < 0$.

2. The new normalisation requirement for the system Ψ is $\langle\Psi|\Psi\rangle = 1 = \int_0^L \Psi^*(x)\Psi(x)\,dx$. This is because the boundary conditions require that the probability of the wavefunction being between 0 and L is 1, not simply just the probability that it is between negative infinity and infinity.

We know that the wavefunction is continuous. This means actually that it must vanish at the walls $x = 0$ and $x = L$, or otherwise it couldn't drop to 0 straight afterwards. So $x = 0$ and $x = L$ are "hard" walls. In other words,

$$\Psi(0) = 0 = \Psi(L)$$

Finally, within the region $x \in (0, L)$, we have $V(x) = 0$. Thus the problem reduces to the free particle inside the square well.

We split the problem into three components: the first, investigating the wavefunction inside the well, the second, investigating the wavefunction outside the well, and finally, investigating the wavefunction on the boundary lines $x = 0$ and $x = L$; we will cover these in that order.

The first part of the problem, the problem of the wavefunction inside the square well, is simply the free particle problem since the potential is 0, within finite position bounds. So, we once again have

$$\frac{d^2\Psi}{dx^2} = -\frac{2mE}{\hbar^2}\Psi$$

which has the solution

$$\Psi(x) = Ae^{ikx} + Be^{-ikx}$$

for some constants A, B, where

$$k = p/\hbar = \frac{\sqrt{2mE}}{\hbar}.$$

We cannot yet determine these constants A, B until we look at the other boundary conditions.

The second component of the problem is the problem of the wavefunction outside the well. We recall that the wavefunction

$$\Psi(x)$$

is a probability distribution function which returns a value given a position x which is its component in the basis and therefore a probability amplitude for achieving that measurement of x. Given that outside the well we have infinite potential, we must have probability zero of finding the particle there. This means that the solution of the wavefunction must also be

$$\Psi(x) = 0 \quad \forall x \in (-\infty, 0) \cup (L, \infty)$$

The most important part of the problem comes, however, with the third component of the problem- that of the wavefunction on the hard walls $x = 0$ and $x = L$. We have

$$\Psi(x) = Ae^{ikx} + Be^{-ikx}$$

for some constants A and B. We can rewrite this with Euler's formula:

$$\begin{aligned} e^{ix} &= \cos(x) + i\sin(x) \\ \implies Ae^{ikx} + Be^{-ikx} &= A\cos(kx) + iA\sin(x) + B\cos(-kx) + iB\sin(-kx). \end{aligned}$$

The cosine function is even and the sine function is odd. Therefore we have

$$Ae^{ikx} + Be^{-ikx} = (A + B)\cos(kx) + i(A - B)\sin(kx).$$

Now, examining the wall conditions, we firstly have:

$$\Psi(x = 0) = 0$$

But since $\sin(0) = 0$ this means

$$(A + B)\cos(kx) = (A + B)\cos(0) = 0.$$

However, $\cos(0) \neq 0$. Therefore we must conclude that $A + B = 0$. However, if we conclude this fact then it must be true for all regions, not simply just the wall $x = 0$, as the wavefunction cannot simply morph as soon as it reaches $x = 0$. Thus for the wavefunction of the whole problem of the infinite square

well, which we will now be covering in trigonometric form, we can omit the first coefficient entirely and therefore vanish the cosine term. We are left with

$$\Psi(x) = i(A - B)\sin kx.$$

Note that we have

$$A + B = 0$$

so neither of the individual constants A and B can be 0, or that would imply the other is 0, and we would have (in exponential form)

$$\Psi(x) = 0e^{ikx} + 0e^{-ikx} = 0$$

which is absurd. Now we can use the other boundary condition.

$$\psi(x = L) = 0 \Rightarrow \ \ i(A - B)\sin kL = 0.$$

If we had

$$A - B = 0$$

then we would have

$$A - B = A + B = 0 \implies B = 0 \implies A = 0$$

which is impossible as shown above. Therefore we must have

$$\sin(kL) = 0.$$

This means that we must have $kL = n\pi, \ \ n \in \mathbb{Z}$. This means the wave number k is again quantized! Thus we can define.

$$k_n = \frac{n\pi}{L}$$

which therefore means we have discrete momenta:

$$p_n = \hbar k_n = \frac{n\pi\hbar}{L}$$

and discrete energy:

$$E_n = \frac{p_n^2}{2m} = \frac{\hbar^2\pi^2 n^2}{2mL^2} \ .$$

Plugging in to our original general solution for the wavefunction, we have the complete set of solutions

$$\Psi_n(x) = N \sin\left(\frac{n\pi x}{L}\right).$$

The constant N is a normalisation constant, which incorporates $i(A - B)$, which we had before. It must be remembered that A and B are arbitrary constants which really are not very important, so rolling them into a new constant N is not a big problem. We can restrict ourselves to considering $n \in \mathbb{Z}^+$ since the parity of the wavefunction is irrelevant in producing the same results. This time, we will solve for the normalisation constant N.

$$\int_0^L \Psi^*(x)\Psi(x)\,dx = N^2 \int_0^L \sin^2\left(\frac{n\pi x}{L}\right)\,dx$$

since the sine function is real valued so its complex conjugate is itself. Then,

$$N^2 \int_0^L \sin^2\left(\frac{n\pi x}{L}\right)\,dx = 1$$

The integral requires some algebra to evaluate. Using the trigonometric identity $\sin^2 kx = \frac{1-\cos 2kx}{2}$, we get:

$$\begin{aligned}
&\sin^2 kx = \frac{1-\cos 2kx}{2} \\
\Rightarrow\ &\int_0^L \sin^2 kx = \int_0^L \frac{1-\cos 2kx}{2} \\
\Rightarrow\ &\int_0^L \sin^2 kx = \int_0^L \frac{1}{2} - \int_0^L \frac{1}{2}\cos 2kx \\
\Rightarrow\ &\int_0^L \sin^2 kx = \int_0^L \frac{d}{dx}\left(\frac{1}{2}x + c\right) - \frac{1}{2}\int_0^L \frac{d}{dx}\left(\frac{1}{2k}\sin 2kx\right) \\
\Rightarrow\ &\int_0^L \sin^2 kx = \left[\frac{1}{2}x + c\right]_0^L - \left[\frac{1}{2k}\sin 2kx\right]_0^L \\
\Rightarrow\ &\int_0^L \sin^2 kx = \frac{L}{2} - \left[\frac{1}{2k}\sin 2kx\right]_0^L
\end{aligned}$$

Now, plugging in $k = \frac{n\pi}{L}$, we get:

$$\begin{aligned}\int_0^L \sin^2\left(\frac{n\pi x}{L}\right)\,dx &= \frac{L}{2} - \left[\frac{L}{n\pi}\sin\left(\frac{2n\pi a}{L}\right) - \frac{1}{2k}\sin 0\right] \\ &= \frac{L}{2} - \left[\frac{L}{n\pi}\sin 2n\pi - 0\right] \\ &= \frac{L}{2}\end{aligned}$$

since $\sin 2n\pi = 0$. So for the normalisation constant we have evaluated the integral from 0 to L and get:

$$N^2\int_0^L \sin^2\left(\frac{n\pi x}{L}\right)\,dx = \frac{N^2 L}{2} = 1 \Rightarrow \;\; N = \sqrt{\frac{2}{L}}$$

as our normalisation constant. That gives us the following set of normalised solutions indexed by positive integers n:

$$\Psi_n = \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x}{L}\right).$$

We can confirm these are orthogonal to each other for different n and thus create an orthonormal set. The normalised property is covered by the normalisation coefficient $\sqrt{2/L}$, and the orthogonality can be proved by considering the inner product- from $x = 0$ to $x = L$ since this is the domain of the wavefunction without it vanishing.

$$\int_0^L \psi_m^*(x)\psi_n(x)\,dx = \left(\sqrt{\frac{2}{L}}\right)^2\int_0^L \sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{n\pi x}{L}\right)\,dx$$

since the complex conjugate of the real valued sine function is itself. Then:

$$\begin{aligned}\sin x\sin y &= \frac{1}{2}[\cos(x-y) + \cos(x+y)] \\ &\Rightarrow \;\; \frac{2}{L}\int_0^L \sin\left(\frac{m\pi x}{L}\right)\sin\left(\frac{n\pi x}{L}\right)\,dx \\ &= \frac{2}{L}\int_0^L \frac{1}{2}\left[\cos\left(\frac{(m-n)\pi x}{L}\right) + \cos\left(\frac{(m+n)\pi x}{L}\right)\right] \\ &= \frac{1}{L}\int_0^L \left[\cos\left(\frac{(m-n)\pi x}{L}\right) + \cos\left(\frac{(m+n)\pi x}{L}\right)\right]\end{aligned}$$

Using some chain rule:

$$\frac{d}{dx}\sin\left(\frac{(m-n)\pi x}{L}\right) = \frac{d}{dg}\sin(g(x)) * \frac{d}{dx}\frac{(m-n)\pi}{L}x$$

where $g(x) = \frac{(m-n)\pi}{L}x$. Then this means:

$$\frac{d}{dx}\sin\left(\frac{(m-n)\pi x}{L}\right) = \frac{(m-n)\pi}{L}\cos\left(\frac{(m-n)\pi}{L}x\right)$$

And therefore,

$$\cos\left(\frac{(m-n)\pi}{L}x\right) = \frac{L}{(m-n)\pi}\frac{d}{dx}\sin\left(\frac{(m-n)\pi x}{L}\right)$$

so we doing something similar with the $\cos(\frac{(m+n)\pi x}{L})$ term we can write the integral as:

$$\frac{1}{L}\int_0^L\left[\left(\frac{L}{(m-n)\pi}\right)\frac{d}{dx}\sin\left(\frac{(m-n)\pi}{L}x\right)+\left(\frac{L}{(m+n)\pi}\right)\frac{d}{dx}\sin\left(\frac{(m+n)\pi}{L}x\right)\right]$$

then, pulling out constants and using the fundamental theorem of calculus to get rid of the integral, we get the expression

$$\left[\frac{1}{(m-n)\pi}\sin\left(\frac{(m-n)\pi}{L}x\right) + \frac{1}{(m+n)\pi}\sin\left(\frac{(m+n)\pi}{L}x\right)\right]_0^L$$

clearly at $x = 0$ we get a bunch of $\sin(0)$ terms so the large bracket is 0. At $x = L$ the value of x cancels with the denominator inside the sine function, and so we get the sine values of $m - n \in \mathbb{Z}^+$ lots of π, which also ends up with 0. So altogether we have proved the whole integral and therefore the inner product of Ψ_m and Ψ_n is 0. Note that the whole process is invalid when $m = n$ since then the fraction $\frac{L}{(m-n)\pi}$ we see in the expression of $\cos(\frac{(m-n)\pi}{L}x)$ as a derivative is clearly invalid due to division by 0. Now we can summarise:

$$\langle\Psi_m|\Psi_n\rangle = \delta_{n,m}$$

which is our original definition of an orthogonal set of wavefunctions $|\Psi\rangle$. So indeed, the set of Ψ_n indexed by integers n is an orthonormal set of states which can then be linearly combined into any superposition of states with different probabilities.

Analysing solutions

We can summarise our numerical analysis of eigenstates as follows. We obtained:

$$\psi_n = \sqrt{\frac{2}{L}} \sin(\frac{n\pi x}{a}), \quad E_n = \frac{\hbar^2 \pi^2 n^2}{2ma^2}, \quad n \in \mathbb{Z}^+$$

We can now plot the first ψ_n on an axis from 0 to a. This is in shown in figure 1. Clearly, we observe a few properties:

1. We call a zero of ψ inside the domain of ψ a node. Then here we see that ψ_n has $n-1$ nodes: the zeros at $x = 0$ and $x = a$ do not count as they are not inside the domain. More importantly, for this set of solutions to ψ for every 1 that you increase the integer indexing ψ_n you also increase the number of nodes by 1. This is actually generally true as you go up from the ground state- the lowest energy state and nearly always ψ_1 in a set of ψ_n, $n \in \mathbb{Z}^+$: for any set of solutions ψ we add 1 node every time we move to the next possible energy level.

2. The ψ_n are clearly alternately symmetric and asymmetric as you go up index numbers n. ψ_1 is symmetric, ψ_2 is asymmetric, ψ_3 is symmetric, etc. Crucially, we call symmetric solutions **even** and the asymmetric solutions **odd**. The definition of an even function $f(x)$ is that $f(x) = f(-x)$. The definition of an odd function is where $f(-x) = -f(x)$. Technically here there are neither even solutions nor odd solutions since this is not true for any ψ_n in the question defined. But if we had taken the midpoint of the well to be $x = 0$ then looking at the first few wavefunctions it is clear this would have been true, so the parity of functions we think in is analogous. We will see in our next problem how important discussion of function parity will be in breaking down questions as they get more and more complex and we seek to categorise more and more realistic situations and problems. But there exists a powerful fact: if a particle is in an even potential then all wavefunctions are either odd or even. In this problem again if we had set the middle of the potential well to be 0 then we would have seen an example of this fact: the potential would be even since it is symmetric about 0 and the solutions are alternately even and odd. Clearly this is an exceptionally

meaningful fact. We can prove it fairly easily for our one-dimensional purposes:

$$\text{By Schrodinger, } -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x) + V(x)\psi(x) = E\psi(x)$$

when we substitute in $x = -x$, we get:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(-x) + V(-x)\psi(-x) = E\psi(-x)$$

we are considering a question where the potential is even and thus $V(x) = V(-x)$ So then:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(-x) + V(x)\psi(-x) = E\psi(-x)$$

we step back out of the algebraic manipulation and see that clearly this means that if $\psi(x)$ is a solution in an even potential then $\psi(-x)$ is also a solution. This in itself is very important, naturally. But since we are working in the same one-dimensional vector space of solutions to the same Schrodinger then it is impossible for these two functions to be linearly independent, or orthogonal in a one-dimensional space. Therefore we can say

$$\psi(x) = a\psi(-x)$$

for some constant a. Then $a\psi(-x) = \psi(x)$, so it satisfies the conditions, and therefore is normalised. So we have:

$$|a|^2 \int_{-\infty}^{\infty} \psi^*(-x)\psi(-x)\, dx = 1$$

But $\psi(-x)$ itself is a normalised solution so the integral is equal to 1. Therefore

$$|a|^2 = 1 \Rightarrow \quad a = \pm 1$$

By definition when $a = 1$ then ψ is even since we get $\psi(x) = \psi(-x)$, and when $a = -1$ then ψ is odd, since we get $\psi(x) = -\psi(x)$. Therefore for an even potential all solutions ψ, when normalised, must be even or odd.

9.5 Harmonic Oscillator ∗

We now introduce a third less fundamental but perhaps more impactful problem which must be included in any discussion of quantum mechanics. We will take one system we know very well from classical mechanics, the spring, which has a potential of

$$V(x) = \frac{1}{2}kx^2$$

for the spring constant k. This can also be written via the angular frequency, $\omega = k/m$, as

$$V(x) = \frac{1}{2}mX\omega^2.$$

There is the standard method of solving this problem in the quantum mechanical version, where the Hamiltonian is

$$\hat{H} = \frac{\hat{P}^2}{2m} + \frac{1}{2}m\hat{X}^2\omega^2.$$

We note that for the system the spring constant and mass are both constants, and so $\omega = k/m$ is a constant and therefore should not and cannot be replaced by an expression in the position and momentum operators for the quantum Hamiltonian. Now the conventional method would be to solve the problem in position space and implement strategies much like we have shown above. But there is a method, courtesy of Dirac, which shows that we can use the energy eigenbasis as well.

In principle this should be difficult, as to find the energy eigenkets which form the energy eigenbasis is tantamount to solving for the propagator for time evolution, in which case it is unclear why we would ever need to be in the energy eigenbasis after that. However, elegance will show in a clever way this limitation is not quite concrete for all problems, and the energy eigenbasis can be very useful even before we know the eigenvectors.

To start, we define the operator

$$a = \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{X} + \frac{i\hat{P}}{m\omega}\right).$$

Then we consider its hermitian adjoint,

$$a^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{X} - \frac{i\hat{P}}{m\omega}\right).$$

It is clear these operators are not hermitian. What relationship is held by the two operators? They turn out not to be unitary either. Instead, we get

$$[a, a^\dagger] = 1$$

which can be easily verified:

$$\begin{aligned}
[a, a^\dagger] &= \frac{m\omega}{2\hbar}\left[\left(\hat{X} + \frac{i\hat{P}}{m\omega}\right)\left(\hat{X} - \frac{i\hat{P}}{m\omega}\right)\right] - \frac{m\omega}{2\hbar}\left[\left(\hat{X} - \frac{i\hat{P}}{m\omega}\right)\left(\hat{X} + \frac{i\hat{P}}{m\omega}\right)\right] \\
&= \frac{m\omega}{2\hbar}\left[\hat{X}^2 - \frac{i}{m\omega}[\hat{X}, \hat{P}] - \left(\frac{i\hat{P}}{m\omega}\right)^2 - \hat{X}^2 - \frac{i}{m\omega}[\hat{X}, \hat{P}] + \left(\frac{i\hat{P}}{m\omega}\right)^2\right] \\
&= \frac{m\omega}{2\hbar}\left[-\frac{i}{m\omega}(2i\hbar)\right] = \frac{m\omega}{2\hbar}\frac{2\hbar}{m\omega} = 1
\end{aligned}$$

Finally, we can define an operator, N, to be $a^\dagger a := N$. This operator is hermitian as $a^\dagger a$. This operator has the form, as seen above:

$$\begin{aligned}
N &= \frac{m\omega}{2\hbar}\left[\hat{X}^2 + \frac{i}{m\omega}[\hat{X}, \hat{P}] - \left(\frac{i\hat{P}}{m\omega}\right)^2\right] = \left(\frac{m\omega}{2\hbar}\right)\left[\hat{X}^2 + \frac{\hat{P}^2}{m^2\omega^2} + \frac{i}{m\omega}[\hat{X}, \hat{P}]\right] \\
&= \frac{1}{\hbar\omega}\frac{\hat{P}^2}{2m}\frac{1}{2\hbar} + \hat{X}^2 m\omega + \frac{i}{2\hbar}i\hbar
\end{aligned}$$

Regarding the Hamiltonian of the system again, this is

$$N = \frac{\hat{H}}{\hbar\omega} - \frac{1}{2} \implies \hat{H} = \hbar\omega\left(N + \frac{1}{2}\right).$$

A bit of clever compatibility thinking is again the step here. Technically N is not an observable operator, but note that what we proved about compatible observables did not hinge on the fact that the operators were Hermitian or had real eigenvalues. Only the third component, that about successive measurements of different observables and whether or not they affect the eigenstate, was contingent on the discussion being about observables in the

other place. In other words, if two operators commute they possess a common eigenbasis; this is not simply limited to observable operators, as a review of the proof we gave will show.

We have seen above that the operator N is a linear combination of the Hamiltonian. This means that they will commute. This in turn means that they possess a common eigenbasis! Thus there exist energy eigenkets $|E\rangle$ which are also eigenkets of the operator N. Thus we have the relationship

$$\hat{H}|E\rangle = E|E\rangle$$

as usual, but also the relationship

$$N|E\rangle = n|E\rangle$$

for some eigenvalue n. We can now equate the eigenvalues!

$$\hat{H}|E\rangle = \hbar\omega\left(N + \frac{1}{2}\right)|E\rangle = \hbar\omega N|E\rangle - \frac{\hbar\omega}{2}|E\rangle = \hbar\omega\left(n + \frac{1}{2}\right)|E\rangle$$

So the energy eigenvalues are given by

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega.$$

The author has tried to instil a justified suspicion in the reader for indexing by n without knowing there is a discrete case which can be indexed by integers; indeed, we still do not know anything about the eigenvalues n so energy could still well be continuous at this stage, and indexing by n would be meaningless. However, we will soon prove that in this case this step is fine as the eigenvalues of N are in fact nonnegative integers n! We will from now on refer to N as the counting operator, as is widespread convention. Let us consider some more commutation relations.

By the commutation relation

$$[AB, C] = A[B, C] + [A, C]B$$

we have

$$[N, a^\dagger] = [a^\dagger a, a^\dagger] = a^\dagger[a, a^\dagger] + [a^\dagger, a^\dagger]a = a^\dagger(1) + 0 = a^\dagger.$$

Now we also have

$$[N, a^\dagger] = Na^\dagger - a^\dagger N \implies Na^\dagger = a^\dagger N + [N, a^\dagger] = a^\dagger N + a^\dagger.$$

Thus applying the operator to an energy eigenket, now indexed by n,

$$Na^\dagger|E_n\rangle = (a^\dagger N + a^\dagger)|E_n\rangle = a^\dagger N|E_n\rangle + a^\dagger|E_n\rangle = (n+1)a^\dagger|E_n\rangle.$$

The reader will recognise the above as another eigenvalue equation, which states that for the counting operator, we have

$$N|E_n\rangle = n|E_n\rangle$$

as one eigenvalue equation, but also the ket $a^\dagger|E_n\rangle$ as an eigenket which has an eigenvalue $n+1$ such that as above, the equation

$$N(a^\dagger|E_n\rangle) = (n+1)(a^\dagger|E_n\rangle)$$

applies. We have mentioned that n is a nonnegative integer, so what this is really saying is that if we apply the counting operator $|E_n\rangle$, we get the "measurement" of an eigenvalue n, but if we then change the eigenket by applying the operator $a^\dagger$ to it first, and then apply the counting operator, we will now count the eigenvalue $n+1$. What is the physical meaning of n? Well, we have

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega.$$

The meaning of $\hbar\omega$ is interesting: one unit we can choose for the Planck's constant $\hbar$ is joules per hertz, and the angular frequency, as suggested by the name, can be measured in hertz, which means that units of $\hbar\omega$ are measured in joules- thereby a measure of energy. Of course, the value of $\hbar\omega$ is not 1 Joule, since the angular frequency would have to be unthinkable for this to be true, but we can think of $\hbar\omega$ as small "packets" of energy (in a loose intuitive manner) which we can use to measure the energy of the system it turns out very appropriately on the microscopic scale. The basic analogy one has already seen in their own studies? Coulombs as a measure of charge, which are also technically "packets" of electrons (again in a very loose manner) which help make the counting process easier!

Thus every time we increase n we increase the energy reading. When n is measured that corresponds to when the counting operator acts on the energy

eigenket $|E_n\rangle$, and the corresponding energy is $n+1/2$ in units $\hbar\omega$. However, if we apply $a^\dagger$ to $|E_n\rangle$ first, we replace n with $n+1$, thereby replacing the corresponding energy $n+1/2$ with the energy $n+1+1/2$. Thus the operator $a^\dagger$ **creates** one unit of energy, measured in $\hbar\omega$- giving it is name. Similarly:

$$[N,a] = [a^\dagger a, a] = a^\dagger[a,a] + [a^\dagger,a]a = a^\dagger(-1) + 0 = -a.$$

Now we also have

$$[N,a] = Na - aN \implies Na = aN + [N,a] = aN - a.$$

Thus applying the operator to an energy eigenket, now indexed by n,

$$Na|E_n\rangle = (aN-a)|E_n\rangle = aN|E_n\rangle - a|E_n\rangle = (n-1)a|E_n\rangle.$$

Which, by the same argument as above, shows that the operator a acting on a energy eigenstate E_n **annihilates** (dramatic, but conventional) one unit of energy $\hbar\omega$! Thus the names the creation operator for $a^\dagger$ and the annihilation operator for a.

Next, we wonder if there is an easier way to label the kets $a^\dagger|E_n\rangle$ and $a|E_n\rangle$, which we have also noted are eigenkets of the counting operator. Well, of course there is-

$$N|E_{n+1}\rangle = (n+1)|E_{n+1}\rangle, \quad Na^\dagger|E_n\rangle = (n+1)a^\dagger|E_n\rangle \implies a^\dagger|E_n\rangle \equiv |E_{n+1}\rangle.$$

This is completely reasonable, because $a^\dagger$ on the state $|E_n\rangle$ raises the energy by 1. However, by the formula

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega,$$

and the fact we have taken it as a given that $n \in \mathbb{Z}^+$, we must have $E_{n+1} = E_n + 1 \times \hbar\omega$, in which case the equivalence of the states $a^\dagger|E_n\rangle$ and $|E_{n+1}\rangle$ makes perfect sense. Similarly, of course, $a|E_n\rangle$ and $|E_{n-1}\rangle$ are equivalent states. Yet we know that equivalence is not algebraic equality: rather,

$$a|E_n\rangle = c|E_{n-1}\rangle$$

for some multiplicative constant c is sufficient for them to be equivalent states. The norm of $a|E_n\rangle$ is

$$\langle E_n|a^\dagger\, a|E_n\rangle = \langle E_n|N|E_n\rangle$$

by our correspondence

$$\Omega|X\rangle \leftrightarrow \langle X|\Omega^\dagger.$$

This is then equal to

$$|c|^2\langle E_{n-1}|E_{n-1}\rangle = |c|^2$$

given the definition of c above and the assumption that $|E_{n+1}\rangle$ has already been normalised. So we have

$$\langle E_n|N|E_n\rangle = n\langle E_n|E_n\rangle = n = |c|^2 \implies c = \sqrt{n}.$$

Thus we have

$$a|E_n\rangle = \sqrt{n}|E_{n-1}\rangle.$$

We can do the exact same thing for $a^\dagger$ to get

$$a^\dagger|E_n\rangle = \sqrt{n+1}|E_{n+1}\rangle.$$

By the semidefinite metric, the norm of any ket, including $a|E_n\rangle$ is positive. However, the norm of $a|E_n\rangle$ is n as already shown. Thus, n must be nonnegative. We next realise that all we have covered so far means that

$$(a^2)|E_n\rangle = a\sqrt{n}|E_{n-1}\rangle = \sqrt{n}a|E_{n-1}\rangle = \sqrt{n(n-1)}\,|E_{n-2}\rangle$$

and so on- in other words, applying the annihilation operator means we should be able keep annihilating energy in units of $\hbar\omega$ until we reach a negative value of n, no matter how great the n and therefore energy. The reason this is problematic, however, is that we have already proven that the counting operator eigenvalue n corresponding to the energy and counting simultaneous eigenstate $|E_n\rangle$ must be nonnegative due to the semidefinite postulate. So we cannot expect a negative value of n to appear. As we have

$$a|E_n\rangle = \sqrt{n}|E_{n-1}\rangle$$

We conclude that if we reach a E_n with $n < 0$ the definition means we must have reached an eigenvalue $n < 0$. The fact above with the sequence of square roots shown by repeatedly applying the annihilation operator shows that clearly we should be able to reach a state E_n with $n < 0$. This is unless we terminate at a value before we reach below $n = 0$. If after annihilating a sufficient number of $\hbar\omega$ units of energy we had $n \in (0, 1)$ then we would be able to apply the annihilation operator again to get a new $n' = n-1 \in (-1, 0)$

which is negative. If we had a value of $n > 1$ which was not an integer then we conclude we must be able to keep applying the annihilation operator until we get a value in the region $(0, 1)$, after which the same argument applies that we reach a negative n. Thus we must have the value $n = 1$ occurring, after which reapplying the annihilation operator should give $n = 0$. What if we now try to reapply the annihilation operator? Well then, we have

$$a|E_n\rangle = \sqrt{n}|E_{n-1}\rangle \implies a|E_0\rangle = 0|E_{-1}\rangle = 0.$$

In other words, all negative index E_n, and therefore negative n disappear! This satisfies our condition of being able to apply the annihilation operator repeatedly but not ever having states where the counting operator eigenvalue n is negative. Thus we have shown with this argument, which should be reread if one is not convinced, that all we have done thus far on the assumption that the eigenvalues n are integers was done on the correct assumption: n must be a nonnegative integer, and there is a terminating state $|E_0\rangle$ which has energy $E_0 = (0 + 1/2)\hbar\omega = (1/2)\hbar\omega$.

Chapter 10

Appendix: A Mathematical Toolbox

This book is designed so that even high schoolers can access quantum mechanics. As such, this section is a reference section, which acknowledges the fact that high school curricula might teach different topics in slightly different orders and detail, and also that some of the prerequisite mathematics assumed here might be just a small reach beyond high school syllabus. The reference section contains some instruction, but it should not be used for anyone learning these topics for the first time, as I have been quite brief so as not to distract from the main topic of this book. For those who are unfamiliar with using any of the mathematical rules referenced as assumed knowledge, they should turn to any textbook which covers these topics and complete a thorough run through some exercises before reading any other part of this book. For many other students, they can simply look at the exercises in these sections, and if they do not see anything unfamiliar, they can simply skip this reference chapter.

10.1 Mathematical Syntax

There are multiple symbols used as figures of proof in this book, which represent common phrases one might turn to were they reciting a proof out loud. As these are ubiquitous symbols, being fluent reading them will be nec-

essary, as otherwise they can be especially easy to confuse with each other.

Symbol	Meaning
$:=$	'define equals'
$\forall$	'For all'
$\exists$	'There exists'
$\in$	'In' (for a set)
$\implies$	'Implies'
$\iff$	'Implies each other'
'Iff'	'If and only if'
$\sum_{i \neq j}$	'Sum over all i not equal to j'
$\sum_{\{i\}}$	'Sum over all x in a set $\{x\}$'
$\square$	Q.E.D
$f'(x)$	First Derivative of $f(x)$
$f''(x)$	Second Derivative of $f(x)$
$\{x_i\}$	'the set of values x_i'
$x \to y$	'x approaches y'
$\mathbb{Z}$	the set of all integers
$\mathbb{R}$	the set of all real numbers
$\mathbb{Z}^+$	the set of all positive integers
$\equiv$	'is equivalent to'

Fluency with basic summation and product notation is assumed. There is also interval notation commonly used for inequalities:

- $x \in [a, b) \iff a \leq x < b$
- $x \in (a, b] \iff a < x \leq b$
- $x \in (a, b) \iff a < x < b$
- $x \in [a, b] \iff a \leq x \leq b$

and round brackets are always used for any side of the inequality bounded by $\pm\infty$.

10.2 Probability*

In this book, as correct and conventional, probabilities are numbers between 0 and 1, sometimes represented by fractions.

For a random X variable which can take multiple values x_i

10.3 Complex Numbers∗

Define the imaginary unit, i, to be equal to the square root of -1. We then have:

- $i^2 = (\sqrt{-1})^2 = -1$
- $\forall\ k \in \mathbb{R},\ \ \frac{k}{i} = \frac{ik}{i^2} = \frac{ik}{-1} = -ik$
- $\forall\ k \in \mathbb{R}^+,\ \ \sqrt{-k} \equiv \sqrt{-1} \times \sqrt{k} = i\sqrt{k}$
- $\forall\ a, b \in \mathbb{R},\ \ ai \pm bi = a\sqrt{-1} \pm b\sqrt{-1} = (a \pm b)\sqrt{-1} = (a \pm b)i$

Complex numbers are numbers usually represented in the form

$$z := a + bi$$

for some $a, b \in \mathbb{R}$. Then there are two functions commonly referred too:

- $\mathrm{Re}(z)$ is called the 'real part' of the imaginary number z, which is, the real number a if $z := a + bi$.
- $\mathrm{Im}(z)$ is called the 'imaginary part' of the imaginary number z, which is, the real number b if $z := a + bi$. Note that $\mathrm{Im}(z) = b$ and not $\mathrm{Im}(z) = bi$.

It is then clear that a real number is also a complex number, but with imaginary part 0. Conversely, if a complex number has imaginary part 0, then all that remains is its real part and so it must be a real number. Therefore,

- $z \in \mathbb{R} \iff \mathrm{Im}(z) = 0$.

We may represent any complex number on a real-valued two dimensional plane called an Argand diagram (figure). From this figure we can create a few more useful definitions:

- The modulus of a complex number z is denoted $|z|$ and defined to be $|z| := \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}$. This is the application of Pythagoras' Theorem to find the distance from any point z represented on the Argand diagram to the origin.

- For all complex numbers $z := a + bi$, we denote with z^* the complex conjugate of the complex number, defined by $z^* = a - bi$ (in other words, reversing the sign of the imaginary part).
- For all complex numbers z, we define the modulus squared (or square modulus) of that complex number to be $|z|^2 := zz^*$. This modulus squared is always real.

Exercises on Complex Numbers*

1. Express as a single complex number with distinct real and imaginary parts:
   (a) $(4i)^3$

   (b) $(5 - 7i)(6 - 8i)$

   (c) $\frac{3+5i}{2-4i}$

   (d) $(-4 - 7i)^*$

   (e) $|2 + 3i|^2$

2. Show, with proof, whether each of the following statements are true
   (a) $|z| > |\text{Re}(z)|$ and $|z| \geq |\text{Im}(z)|$

   (b) $\forall\ z \in \mathbb{C},\ \ |z|^2 \in \mathbb{R}$

3. Show that $|r\cos\theta + ir\sin\theta| = r$

4.

10.4 Matrices*

A matrix is an array of values. They are placed into rectangular arrangements in rows and columns, and are an extremely important way of

mathematically listing values. We shall see that the structure of a matrix allows for powerful computations to be done. For now, there are many definitions to cover.

A matrix with m rows and n columns is said to be an $m \times n$ matrix. A $n \times n$ matrix is called a square matrix, a $n \times 1$ matrix is usually called a column vector, and a $1 \times n$ matrix is usually called a row vector.

The values of a matrix are usually called the element of that matrix. For a matrix denoted by any symbol, for example M, the notation M_{ij} usually represents the element in the i'th row and j'th column of the matrix. Therefore, we can show the four usual types of matrices we might often see:

1. An arbitrary $m \times n$ matrix:

$$\begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} & \dots & \dots & \mathrm{M}_{1n} \\ \mathrm{M}_{21} & \ddots & \dots & \dots & \vdots \\ \vdots & \dots & \ddots & \dots & \vdots \\ \mathrm{M}_{m1} & \dots & \dots & \dots & \mathrm{M}_{mn} \end{bmatrix}$$

2. A $n \times n$ square matrix:

$$\begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} & \dots & \mathrm{M}_{1n} \\ \mathrm{M}_{21} & \ddots & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ \mathrm{M}_{n1} & \dots & \dots & \mathrm{M}_{nn} \end{bmatrix}$$

3. A $n \times 1$ column vector:

$$\begin{bmatrix} \mathrm{M}_{11} \\ \mathrm{M}_{21} \\ \vdots \\ \vdots \\ \mathrm{M}_{n1} \end{bmatrix}$$

4. A $1 \times n$ row vector:

$$\begin{bmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} & \dots & \dots & \mathrm{M}_{1n} \end{bmatrix}$$

As shown in the examples, a variety of dots are used often to show the idea that we can have a large number of unlisted elements between elements explicitly written. This should not be a great worry, since the arrangements of elements in a matrix are always in ordered rows and columns anyway. We now move onto defining the operations:

1. Scalar multiplication of a matrix M by a scalar a results in a new matrix aM with each previous element M_{ij} replaced with a new element aM_{ij}. So for example,

$$-1 \times \begin{bmatrix} 5 & 4 & -9 \\ 0 & -1 & 6 \end{bmatrix} = \begin{bmatrix} -5 & -4 & 9 \\ 0 & 1 & -6 \end{bmatrix}$$

2. Two matrices can be summed iff they are of the same dimensions. If they are of the same dimensions, the sum of two matrices M and N are obtained by creating a new matrix $S := M + N$, of the same dimensions, where each new element of the resultant matrix S_{ij} is the result of summing corresponding position elements in M: in other words, $\forall\ i, j, S := M + N,\ \ S_{ij} = M_{ij} + N_{ij}$. So for example,

$$\begin{bmatrix} 3 & 2 \\ -1 & 8 \\ 0 & 5 \end{bmatrix} + \begin{bmatrix} 7 & -3 \\ -1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 3+7 & 2-3 \\ -1-1 & 8+0 \\ 0+0 & 5+2 \end{bmatrix} = \begin{bmatrix} 10 & -1 \\ -2 & 8 \\ 0 & 7 \end{bmatrix}$$

3. A matrix multiplying another matrix is more complicated. First of all, for two matrices M and N the matrix products $M \times N$ and $N \times M$ are rarely the same: whichever matrix is on the left of the product changes the result. We say that matrix multiplication is non-commutative. The terms left-multiply and right-multiply emerge naturally from this non-commutativity, where left-multiplying a matrix M by a matrix N means performing a matrix multiplication with that matrix N on the left of the multiplication, and vice versa.

   Next, matrix multiplication can only be performed if the number of columns of the matrix on the left are the same as the number of rows of the matrix on the right. If we have an $a \times b$ matrix M and a $c \times d$ matrix N, then the product $M \times N$ is only possible if $b = c$. The result can be best memorised by the crude representation:

$$b = c \implies (a \times b) \times (c \times d) = (a \times b) \times (b \times d) = a \times d.$$

Here, we get the idea that the 'middle two' values when we list the dimensions of the two matrices being multiplied together in the correct order must match up and then are discarded, leaving the number of rows of the left matrix and the number of columns of the right matrix as the dimensions of the resulting matrix after multiplication.

10.5 Calculus*

10.6 Misc.*