

Advanced Quantum Theory

AMATH473/673, PHYS454

Achim Kempf

Department of Applied Mathematics
University of Waterloo
Canada

© Achim Kempf, October 2016

(Please do not copy: textbook in progress)

Contents

1	A brief history of quantum theory	7
1.1	The classical period	7
1.2	Planck and the “Ultraviolet Catastrophe”	7
1.3	Discovery of h	8
1.4	Mounting evidence for the fundamental importance of h	9
1.5	The discovery of quantum theory	9
1.6	Relativistic quantum mechanics	10
1.7	Quantum field theory	12
1.8	Beyond quantum field theory?	14
1.9	Experiment and theory	16
2	Classical mechanics in Hamiltonian form	17
2.1	Newton’s laws for classical mechanics cannot be upgraded	17
2.2	Levels of abstraction	18
2.3	Classical mechanics in Hamiltonian formulation	19
2.3.1	The energy function H contains all information	19
2.3.2	The Poisson bracket	21
2.3.3	The Hamilton equations	23
2.3.4	Symmetries and Conservation laws	25
2.3.5	A representation of the Poisson bracket	27
2.4	Summary: The laws of classical mechanics	28
2.5	Classical field theory	28
3	Quantum mechanics in Hamiltonian form	31
3.1	Reconsidering the nature of observables	32
3.2	The canonical commutation relations	33
3.3	From the Hamiltonian to the Equations of Motion	36
3.4	From the Hamiltonian to predictions of numbers	39
3.4.1	A matrix representation	40
3.4.2	Solving the matrix differential equations	42
3.4.3	From matrix-valued functions to number predictions	43
3.5	Initial conditions	46

3.6	Emergence of probabilities	47
3.7	The Hilbert space and Dirac's notation	50
3.7.1	Hilbert spaces	51
3.7.2	Hilbert bases	53
3.7.3	Discrete wave functions and matrix representations	54
3.7.4	The domain of operators	55
3.7.5	Changes of basis	56
4	Uncertainty principles	59
4.1	The Heisenberg uncertainty relations	59
4.2	The impact of quantum uncertainty on the dynamics	63
4.3	The time and energy uncertainty relation	64
5	Pictures of the time evolution	67
5.1	The time-evolution operator	67
5.1.1	Calculating $\hat{U}(t)$	68
5.1.2	Significance of $\hat{U}(t)$	71
5.2	The pictures of time evolution	73
5.2.1	The Heisenberg picture	73
5.2.2	The Schrödinger picture	73
5.2.3	The Dirac picture	76
6	Measurements and state collapse	81
6.1	Ideal measurements	81
6.2	The state collapse	83
6.3	Simultaneous measurements	84
6.4	States versus state vectors	85
7	Quantum mechanical representation theory	87
7.1	Self-adjointness	87
7.2	The spectrum of an operator	88
7.3	Stieljes Integration	90

Introduction

Quantum theory, together with general relativity, represents humanity's so-far deepest understanding of the laws of nature. And quantum phenomena are not rare or difficult to observe. In fact, we experience quantum phenomena constantly! For example, the very stability of the desk at which you are sitting now has its origin in a quantum phenomenon. This is because atoms are mostly empty space and the only reason why atoms don't collapse is due to the uncertainty relations. Namely, the uncertainty relations imply that it costs plenty of momentum (and therefore energy) to compress atoms. Also, for example, the spectrum of sunlight is shaped by quantum effects - if Planck's constant were smaller, the sun would be bluer.

Over the past century, the understanding of quantum phenomena has led to a number of applications which have profoundly impacted society, applications ranging from nuclear power, lasers, transistors and photovoltaic cells, to the use of MRI in medicine. Ever new sophisticated applications of quantum phenomena are being developed, among them, for example, quantum computers which have the potential to revolutionize information processing.

Also on the level of pure discovery, significant progress is currently being made, for example, in the field of cosmology, where both quantum effects and general relativistic effects are important: high-precision astronomical data obtained by satellite telescopes over the past 15 years show that the statistical distribution of matter in the universe agrees with great precision with the distribution which quantum theory predicts to have arisen from quantum fluctuations shortly after the big bang. We appear to originate in quantum fluctuations. New satellite-based telescopes are being planned.

The aim of this course is to explain the mathematical structure of all quantum theories and to apply it to nonrelativistic quantum mechanics. Nonrelativistic quantum mechanics is the quantum theory that replaces Newton's mechanics and it is the simplest quantum theory. The more advanced quantum theory of fields, which is necessary for example to describe the ubiquitous particle creation and annihilation processes, is beyond the scope of this course, though of course I can't help but describe some of it. For example, the first chapter of these notes, up to section 1.5, describes the history of quantum theory as far as we will cover it in this course. The introduction goes on, however, with a historical overview that outlines the further developments, from relativistic quantum mechanics to quantum field theory and on to the modern day quest

for a theory of quantum gravity with applications in quantum cosmology. Quantum theory is still very much a work in progress and original ideas are needed as much as ever!

Note: This course is also a refresher course for beginning graduate students, as AMATH673 at the University of Waterloo. Graduate do the same homework and write the same midterm and final but write also an essay. If you are a grad student taking this course, talk with me about the topic.

Here at Waterloo, there are a number of graduate courses that build on this course. For example, I normally teach every other year Quantum Field Theory for Cosmology (AMATH872/PHYS785).

Chapter 1

A brief history of quantum theory

1.1 The classical period

At the end of the 19th century, it seemed that the basic laws of nature had been found. The world appeared to be a mechanical clockwork running according to Newton's laws of mechanics. Light appeared to be fully explained by the Faraday-Maxwell theory of electromagnetism which held that light was a wave phenomenon. In addition, heat had been understood as a form of energy. Together, these theories constituted "Classical Physics". Classical physics was so successful that it appeared that theoretical physics was almost complete, the only task left being to add more digits of precision. And so, Max Planck's teacher, Scholli, advised his student against a career in physics. Soon after classical physics was overthrown.

1.2 Planck and the "Ultraviolet Catastrophe"

The limits to the validity of classical physics first became apparent in measurements of the spectrum of heat radiation. It had been known that very hot objects, such as a smith's hot iron, are emitting light. They do because matter consists of charged particles which can act like little antennas that emit and absorb electromagnetic waves. This means that also cold objects emit and absorb electromagnetic radiation. Their heat radiation is not visible because it too weak and too red for our eyes to see. Black objects are those that absorb electromagnetic radiation (of whichever frequency range under consideration) most easily and by time reversal symmetry they are therefore also the objects that emit electromagnetic radiation of that frequency range most readily. Tea in a black tea pot cools down faster than tea in a white or reflecting tea pot.

Now at the time that Planck was a student, researchers were ready to apply the laws of classical physics to a precise calculation of the radiation spectrum emitted by black bodies. To everybody's surprise the calculations, first performed by Rayleigh and Jeans, predicted far more emission of waves of short wavelengths (such as ultraviolet)

than what experimental measurements seemed to indicate. This was not a subtle discrepancy: the laws of classical physics were found to predict that any object would actually emit an infinite amount of heat radiation in an arbitrarily short time, especially at very short wave lengths.

At first, this was not seen as a reason to doubt the laws of classical physics. It seemed obvious that this nonsensical prediction could only be due to an error in the calculation. Eventually, however, as time passed and nobody appeared to be able to find a flaw in the calculation, the problem became considered serious enough to be called the “ultraviolet catastrophe”. Scholli suggested to Planck to look into this problem.

1.3 Discovery of h

From about 1890 to 1900, Planck dedicated himself to thoroughly analyzing all assumptions and steps in the calculations of Rayleigh and Jeans. To his great disappointment and confusion he too did not find an error. In the year 1900, Planck then learned of a new precision measurement of the heat radiation spectrum. Those measurements were precise enough to allow curve fitting. Planck had so much experience with the calculations of the heat radiation that on the same day that he first saw the curve of the heat radiation spectrum he correctly *guessed* the formula for the frequency spectrum of heat radiation, i.e., the formula that is today called Planck’s formula. After two further months of trying he was able to derive his formula from a simple but rather radical hypothesis. Planck’s hypothesis was that matter cannot radiate energy continually, but only in discrete portions of energy which he called “quanta”.

Concretely, Planck postulated that light of frequency f could only be emitted in packets of energy $E_q = hf$, as if light was consisting of particles. He found that the value of this constant, h , must be about $6.6 \cdot 10^{-34} \text{Kg m}^2/\text{s}$ for the prediction of the heat radiation spectrum to come out right. Planck’s quantum hypothesis was in clear contradiction to classical physics: light was supposed to consist of continuous waves - after all, light was known to be able to produce interference patterns¹. Nevertheless, most researchers, including Planck himself, still expected to find an explanation of his quantum hypothesis within classical physics.

¹It is easy to see these interference patterns: in a dark room, have a candle burning on a desk, then sit a few meters away from it. Close one of your eyes and hold a hair in front of your other eye, about 1cm in front of the eye, vertically. Align the hair with the flame of the candle. Do you see an interference pattern, i.e., the flame plus copies of it to its left and right? From the apparent distance between the copies of the flame and the distance of the hair to the flame you can work out the ratio of the thickness of the hair to the wavelength of the light.

1.4 Mounting evidence for the fundamental importance of h

The significance of Planck's constant was at first rather controversial. Einstein, however, was prepared to take Planck's finding at face value. In 1906, Einstein succeeded in quantitatively explaining the photoelectric effect². Then, he reasoned, the light's energy packets must be of high enough energy and therefore of high enough frequency to be able to free electrons from the metal. For irrational reasons, Einstein's explanation of the photoelectric effect is the only result for which he was awarded a Nobel prize.

At about the same time, work by Rutherford and others had shown that atoms consist of charged particles which had to be assumed to be orbiting another. This had led to another deep crisis for classical physics: If matter consisted of charged particles that orbit another, how could matter ever be stable? When a duck swims in circles in a pond, it continually makes waves and the production of those waves costs the duck some energy. Similarly, an electron that orbits a nucleus should continually create electromagnetic waves. Just like the duck, also the electron should lose energy as it radiates off electromagnetic waves. A quick calculation showed that any orbiting electron should rather quickly lose its energy and therefore fall into the nucleus.

Finally, in 1913, Bohr was able to start explaining the stability of atoms. However, to this end he too had to make a radical hypothesis involving Planck's constant h : Bohr hypothesized that, in addition to Newton's laws, the orbiting particles should obey a strange new equation. The new equation says that a certain quantity calculated from the particle's motion (the so called "action"), can occur only in integer multiples of h . In this way, only certain orbits would be allowed. In particular, there would be a smallest orbit of some finite size, and this would be the explanation of the stability of atoms. Bohr's hypothesis also helped to explain another observation which had been made, namely that atoms absorb and emit light preferably at certain discrete frequencies.

1.5 The discovery of quantum theory

Planck's quantum hypothesis, Einstein's light quanta hypothesis and Bohr's new equation for the hydrogen atom all contained Planck's h in an essential way, and none of this could be explained within the laws of classical physics. Physicists, therefore, came to suspect that the laws of classical physics might have to be changed according to some overarching new principle, in which h would play a crucial role. The new physics

²Under certain circumstances light can kick electrons out of a metal's surface. Classical physics predicted that this ability depends on the brightness of the light. Einstein's quantum physics correctly explained that it instead depends on the color of the light: Einstein's radical idea was that light of frequency ω comes in quanta, i.e., in packets of energy $h\omega$

would be called quantum physics. The theoretical task at hand was enormous: One would need to find a successor to Newton's mechanics, which would be called quantum mechanics. And, one would need to find a successor to Faraday and Maxwell's electromagnetism, which would be called quantum electrodynamics. The new quantum theory would have to reproduce all the successes of classical physics while at the same time explaining in a unified way all the quantum phenomena, from Planck's heat radiation formula, to the stability and the absorption and emission spectra of atoms.

The task took more than twenty years of intense experimental and theoretical research by numerous researchers. Finally, in 1925, it was Heisenberg who first found "quantum mechanics", the successor to Newton's mechanics. (At the time, Heisenberg was a 23 year old postdoctoral fellow with a Rockefeller grant at Bohr's institute in Copenhagen). Soon after, Schrödinger found a seemingly simpler formulation of quantum mechanics which turned out to be equivalent. Shortly after, Dirac was able to fully clarify the mathematical structure of quantum mechanics, thereby revealing the deep principles that underlie quantum theory. Dirac's textbook "Principles of Quantum Mechanics" is a key classic.

The new theory of "Quantum Mechanics", being the successor to Newton's mechanics, correctly described how objects move under the influence of electromagnetic forces. For example, it described how electrons and protons move under the influence of their mutual attraction. Thereby, quantum mechanics explained the stability of atoms and the details of their energy spectra. In fact, quantum mechanics was soon applied to explain the periodic table and the chemical bonds.

What was still needed, however, was the quantum theory of those electromagnetic forces, i.e., the quantum theoretic successor to Faraday and Maxwell's electromagnetism. Planck's heat radiation formula was still not explained from first principles! Fortunately, the discovery of quantum mechanics had already revealed most of the deep principles that underlie quantum theory. Following those principles, Maxwell's theory of electromagnetism was "quantized" to arrive at quantum electrodynamics so that Planck's formula for the heat radiation spectrum could be derived.

It then became clear that quantum mechanics, i.e., the quantization of classical mechanics, was merely the starting point. Somehow, quantum mechanics would have to be upgraded to become consistent with the brand new theory of relativity which Einstein had discovered! And then it would have to be covariantly combined with the quantization of electrodynamics in order to be able to describe both matter and radiation and their interactions.

1.6 Relativistic quantum mechanics

Already by around 1900, Lorentz, Einstein and others had realized that Newton's mechanics was in fact incompatible with Faraday and Maxwell's theory of electromagnetism, for reasons unrelated to quantum theory, thereby contributing to the crisis of

classical physics. In a daring move, Einstein accepted Faraday and Maxwell's theory of electromagnetism as correct and questioned the validity of Newton's notion of absolute space and time:

Maxwell was able to calculate the speed of electromagnetic waves from first principles, and found it to match with the measured speed of light. His calculations also showed, however, that a traveller would with some large constant velocity would find the same speed of light. (Today we would say that this is because the Maxwell equations are covariant).

At the time, this was rather surprising as it clearly contradicted Newton's classical mechanics which says that velocities are simply additive. For example, according to Newton, a passenger who walks forward at $v_1 = 5\text{km/h}$ in a train travelling at $v_2 = 100\text{km/h}$ has a speed of $v_3 = v_1 + v_2 = 105\text{km/h}$ relative to the ground. In fact, he does not. His speed to the ground is $v_3 = (v_1 + v_2)/(1 + v_1 v_2/c^2) = 104.9999994\dots\text{km/h}$. Today, the nonadditivity of velocities is an easy-to-measure everyday phenomenon. At the time, the nonadditivity of velocities was first confirmed experimentally by Michelson and Moreley, who compared the speed of two light rays travelling parallel and orthogonal to the motion of the earth around the sun. The new theory that explained it all was of course Einstein's special relativity. By 1916, he developed it into general relativity, which supersedes Newton's laws of gravity. General relativity very elegantly explains gravity as curvature of space-time.

Historically, the discovery of relativity therefore happened more or less simultaneously with the discovery of quantum theory. Yet, the two theories were developed virtually independently of another. In actual experiments, special relativity effects seemed of little importance to quantum mechanical effects and vice versa. For example, it was easy to estimate that an electron which orbits the nucleus of a hydrogen atom would travel at most at speeds smaller than one percent of the speed of light. Also, since gravity is extremely weak compared to the electromagnetic forces that rule the atom it was clear that general relativity would be even less important than special relativity for those early quantum mechanical studies. Conversely, the uncertainty principle appeared irrelevant at the astrophysical scales where general relativity was applied.

Nevertheless, soon after quantum mechanics had been found in 1925 it became apparent that at least the tiny special relativistic effect of the speed of an electron orbiting a nucleus was indeed measurable. This meant that there was experimental guidance for the development of an improved version of quantum mechanics that would be compatible with special relativity. Indeed, Klein, Gordon, Dirac and others soon developed "relativistic quantum mechanics"³. Dirac's analysis, in particular, led him to correctly predict surprising magnetic properties of electrons, and it led him to correctly

³This relativistic quantum mechanics is an improvement of quantum mechanics which is consistent merely with *special* relativity. The search for a quantum theory that is consistent also with general relativity is still on today.

predict the existence and properties of antiparticles such as the positron!

However, the fact that particles are able to create and annihilate another in collisions, which had clearly been observed, was beyond the power of even relativistic quantum mechanics. It was clear that a significant enlargement of the framework of quantum theory was needed.

1.7 Quantum field theory

The way forward was called “second quantization”. The starting observation was that, in quantum mechanics, the wave functions behave completely deterministically, namely according to the Schrödinger equation. Given the initial wave function, one can calculate its evolution with absolute certainty. It was felt that to be able to predict the evolution of something, here the wavefunction, with absolute certainty was unusual for a quantum theory. The idea of second quantization was, therefore, to apply quantum theory to quantum theory itself. To this end, the quantum mechanical wave functions were to be treated as classical fields, much like the classical electromagnetic fields. Then, the aim was to find the quantum version of those fields. Since quantum theory was to be applied to the wave functions themselves, the amplitudes of wave functions would no longer be numbers but they would be operators instead. (An operator is a linear map on an infinite dimensional vector space). As a consequence, in quantum field theory, the amplitudes of the wave functions would be subject to uncertainty relations. One should not be able to be sure of the values of the wave function, nor should one be able to be sure of the norm of the wave function. Since in quantum mechanics the normalization of the wave function to norm one means that there is exactly one particle, somewhere, i.e., one would with second quantization not necessarily be sure how many particles there are. Roughly speaking, it is in this way that the quantum fluctuations of the wave functions themselves would then account for the creation and annihilation of particles⁴.

The problem of finding a quantum theory for fields had of course already been encountered when one had first tried to find the quantum theoretic successor to Faraday and Maxwell’s electrodynamics (which was consistent with special relativity from the start). As it turned out, guided by the general principles underlying quantum mechanics the quantum theory of the electromagnetic fields alone was not too hard to find. Following these lines, one was eventually able to write down a unifying quantum theory both of charged particles and their antiparticles, and also of their interaction through electromagnetic quanta, i.e., photons. While this theory succeeded well in describing all the interactions, including annihilation and creation processes, it did yield much more than one had bargained for. The reason was that, since now particle number was no longer conserved, the time-energy uncertainty principle made it possible for

⁴Yes, third and higher quantization has been considered, but with no particular successes so far.

short time intervals that energy (and therefore all kinds of particles) could be virtually “borrowed” from the vacuum.

As a consequence, the new quantum field theory, called quantum electrodynamics, necessarily predicted that, for example, that an electron would sometimes spontaneously borrow energy from the vacuum to emit a photon which it then usually quickly reabsorbs. During its brief existence, this so-called “virtual” photon even has a chance to split into a virtual electron-positron pair which shortly after annihilates to become the virtual photon again. In fact, the virtual electron (or the positron) during its short existence, might actually emit and quickly reabsorb a virtual photon. That photon, might briefly split into an electron positron pair, etc etc ad infinitum. Even more intriguing is that even without a real electron to start with, the vacuum alone is predicted to have virtual particles continually appearing and disappearing!

In spite of all this new complexity, it turned out that the theory’s predictions for the very simplest interactions and creation-annihilation processes were in very good agreement with experimental results. However, the calculation of those predicted endless chain reactions of virtual processes typically yielded divergent integrals! To take those virtual processes into account should have increased the precision of predictions. Instead, one only obtained seemingly meaningless predictions of infinite numbers. It took the combined efforts of numerous scientists, such as Feynman, Tomonaga, Weisskopf, Dyson and others, over about twenty years, to solve this problem.

It turned out that those calculations that had yielded infinities did make sense after all, if one suitably recalibrated the parameters of the theory, such as the fundamental masses and charges. This process of recalibration, called renormalization, also occurs in condensed matter physics, where it is easier to understand intuitively: Consider an electron that is traveling through a crystal. It has the usual mass and charge. But if you want to influence the electron’s motion you will find that the traveling electron behaves as if it had a several times larger mass and a smaller charge. That’s because the electron slightly deforms the crystal by slightly displacing the positive and negative charges that it passes by. It is these deformations of the crystal, which travel with the electron, which make the electron behave as if it were heavier and they also shield its charge. Also, the closer we get to the electron with our measurement device, the less is its charge shielded, i.e., the more we see of the bare charge of the electron.

The key lesson here is that the masses and charges that one observes in a crystal are generally not the “bare” masses and charges that the particles fundamentally possesses. The observed masses and charges even depend on how closely one looks at the electron.

Now when fundamental particles travel through the vacuum, then they deform the distribution of those virtual particles that pop in and out of existence due to the time-energy uncertainty principle. Again, this makes particles behave as if they had a different mass and a different charge. The masses and charges that are observed are not the “bare” masses and charges that the particles fundamentally possess. The observed masses and charges actually depend again on how closely one looks at the particles, i.e., at what energy one observes them, say with an accelerator. In quantum

field theory, it turns out that the bare masses and charges may formally even tend to zero or be divergent. This is OK, as long as the predicted measured values come out right.

Technically, if you would like to know the gist of it already, renormalization consists of the following steps: First, artificially render all predictions finite, say by cutting of the divergent integrals. It turned out that this can be achieved by postulating the existence of a smallest possible distance ϵ between any two particles and by calculating virtual processes accordingly. Next, adjust the parameters of the theory (charges, masses etc) such that a handful of predictions come out in agreement with experiment (namely as many as there are free parameters such as masses and charges in the theory). Now let $\epsilon \rightarrow 0$, while at the same time letting the bare parameters of the theory run so that the same handful of predictions comes out right. (The parameters of the theory will thereby usually tend to 0 or ∞ .) Crucially, all other (infinitely many!) possible predictions of the theory will now also come out finite in the limit $\epsilon \rightarrow 0$ - and they can be compared to experiment. Indeed, predictions so-obtained through renormalization, for example for the energy levels of the hydrogen atom, match the experimental results to more than a dozen digits behind the comma!

Of course, renormalization has always been seen as mathematically and conceptually unsatisfactory. Nevertheless, it did open the door to the successful application of quantum field theory for the description of all the many species of particles that have been discovered since, from neutrinos and muons to quarks.

It is important also to mention two developments related to quantum field theory: First, on the applied side, it turned out that quantum field theoretic methods can also be used for the description of wave phenomena in solids. These are crucial, for example, for the understanding of superconductivity. Second, on the theoretical side, Feynman in his work on quantum electrodynamics, found an equivalent but very insightful and mathematically powerful new formulation for the principles of quantum theory, called the path integral formulation. I will briefly outline the path integral formulation of quantum mechanics later in this course.

1.8 Beyond quantum field theory?

Today, quantum field theory has served as the basis of elementary particle physics (and therefore as the basis for the description of all that we are made of) for about fifty years. Even though numerous new particles and even new forces have been discovered over the years, quantum field theory itself never needed to undergo any fundamental changes. Similarly successful has been Einstein's general relativity, which has now served as the basis of all gravitational physics for over 80 years. Even the most sophisticated experiments seem to merely confirm the validity of quantum field theory and general relativity with more and more precision.

Could it be, therefore, that these two theories constitute the final laws of nature and

that this is all there is? Should one discourage students from a career in the subject? Certainly not! In fact, the situation resembles in many ways the situation at the time Planck was a student. We have two highly successful theories - but they are inconsistent! As long as we consider gravity to be a fixed background for quantum theory some calculations can be performed. Hawking's prediction of black hole radiation is of this kind. However, once we fully take into account the dynamics of general relativity, we face a problem: The predictions of infinities in quantum field theory appear to persist. In the renormalization procedure, the limit $\epsilon \rightarrow 0$ does no longer seem to work (not for lack of trying!).

This problem is very deep. Many believe that this indicates that there actually exists a finite shortest length, ϵ , in nature, much like there is a finite fastest speed. Indeed, if we put together what we know from general relativity and what we know from quantum theory, we can conclude that we cannot even in principle devise an experimental operation that would allow us to resolve distances as small as about $10^{-35}m$, which is the so-called Planck scale:

Consider the task of resolving some very small structure. To this end, we need to shine on it some probing particles of very short wavelength. Due to quantum theory, the shorter the wavelength, the higher is the energy uncertainty of the probing particle. According to general relativity, energy gravitates and curves space. Thus, the probing particles will randomly curve space to the extent of their energy uncertainty. Assume now that a distance of $10^{-35}m$ or smaller is to be resolved. A short calculation shows that to this end the probing particles would have to be of such short wavelength, i.e., of such high energy uncertainty that they would significantly curve and thereby randomly disturb the region that they are meant to probe. It therefore appears that the very notion of distance loses operational meaning at distances of $10^{-35}m$ or so.

In order to describe the structure of space-time and matter at such small scales we will need a unifying theory of quantum gravity. Much effort is being put into this. In this field of research, it is widely expected that within the unified quantum gravity theory there will be a need for renormalization, but not for infinite renormalization. This yet-to-be found theory of quantum gravity may also solve several other major problems of quantum theory. In particular, it could yield an explanation for the particular masses and charges of the elementary particles, and perhaps even an explanation for the statistical nature of quantum theoretical predictions.

A very concrete major problem awaiting resolution in the theory of quantum gravity is the derivation of the cosmological constant, which represents the energy of the vacuum. Quantum field theory predicts the vacuum to possess significant amounts of energy due to vacuum fluctuations: Each field can be mathematically decomposed into a collection of quantum theoretical harmonic oscillators, each of which contributes a finite ground state energy of $\hbar\omega/2$. General relativity predicts that the vacuum energy should gravitate, just like any other form of energy.

Evidence from recent astronomical observations of the expansion rate of the universe indicates that the cosmological constant has a small but nonzero value. How much

vacuum energy does quantum field theory predict? Straightforwardly, quantum field theory predicts the vacuum energy density to be infinite. If we augment quantum field theory by the assumption that the Planck length is the shortest length in nature, then quantum field theory predicts a very large vacuum energy. In fact, it is by a factor of about 10^{120} larger than what is experimentally observed. This is the today's "ultraviolet catastrophe". It appears that whoever tries to reconcile quantum theory with general relativity must be prepared to question the very foundations of all we know of the laws of nature. Original ideas are needed that may be no less radical than those of Planck or Einstein. Current attempts are, for example, string theory and loop quantum gravity.

1.9 Experiment and theory

In the past, progress in the search for the theory of quantum gravity has been severely hampered by the fact that one cannot actually build a microscope with sufficiently strong resolving power to probe Planck scale physics. Even the best microscopes today, namely particle accelerators, can resolve distances only down to at most $10^{-20}m$, which is still very far from the Planck scale of $10^{-35}m$. Of course, guidance from experiments is not strictly necessary, as Einstein demonstrated when he first developed general relativity. Nevertheless, any candidate theory must be tested experimentally before it can be given any credence.

In this context, an important recent realization was that there are possibilities for experimental access to the Planck scale other than through accelerators! One possibility could be the study of the very highly energetic cosmic rays that occasionally hit and locally briefly light up the earth's atmosphere. Another recently much discussed possibility arises from the simple fact that the universe itself was once very small and has dramatically expanded since. The idea is, roughly speaking, that if the early expansion was rapid enough then the universe might have acted as a microscope by stretching out everything to a much larger size. Astronomical evidence obtained over the past few years indicate that this did happen.

The statistical distribution of matter in the universe is currently being measured with great precision, both by direct observation of the distribution of galaxies, and through the measurement of the cosmic microwave background. Experimental evidence is mounting for the theory that the matter distribution in the universe agrees with what one would expect if it originated as tiny primordial quantum fluctuations - which were inflated to cosmic size by a very rapid initial expansion of the universe! It appears that the universe itself has acted as a giant microscope that enlarged initially small quantum phenomena into an image on our night sky! It is just possible that even the as yet unknown quantum phenomena of Planck length size have left an imprint in this image. Some of my own research is in this area. New satellite based telescopes are currently further exploring these phenomena.

Chapter 2

Classical mechanics in Hamiltonian form

2.1 Newton's laws for classical mechanics cannot be upgraded

When physicists first tried to find the laws of quantum mechanics they knew from experiments that Planck's constant h would have to play an important role in those laws. Imagine yourself in the situation of these physicists. How would you go about guessing the laws of quantum mechanics? Clearly, quantum mechanics would have to strongly resemble classical mechanics. After all, quantum mechanics should be an improvement over classical mechanics. Thus, it would have to reproduce all the successful predictions of classical mechanics, from the motion of the planets to the forces in a car's transmission. So how if we try to carefully improve one or several Newton's three axioms of classical mechanics by suitably introducing Planck's constant?

For example, could it be that $F = ma$ should really be $F = ma + h$ instead? After all, h is such a small number that one could imagine that this correction term might have been overlooked for a long time. However, this attempt surely can't be right on the trivial grounds that h does not have the right units: F and ma have the units Kgm^2/s^2 while the units of h are Kgm^2/s . But then, could the correct second law perhaps be $F = ma(1 + h/xp)$? The units would match. Also this attempt can't be right because whenever x or p are small, the term h/xp would be enormous, and we could therefore not have overlooked this term for all the hundreds of years since Newton. Similarly, also $F = ma(1 + xp/h)$ can't be right because for the values of x and p that we encounter in daily life the term xp/h would usually be big enough to have been seen.

In fact, no attempt to improve on Newton's laws in such a manner works. This is why historically this search for the laws of quantum mechanics actually took a quarter century! When the first formulations of quantum mechanics were eventually found by

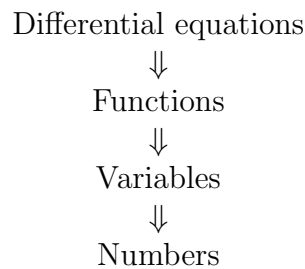
Heisenberg and Schrödinger, they did not at all look similar to classical mechanics.

It was Dirac who first clarified the mathematical structure of quantum mechanics and thereby its relation to classical mechanics. Dirac remembered that a more abstract formulation of classical mechanics than Newton's had long been developed, namely Hamiltonian's formulation of classical mechanics. Hamilton's formulation of classical mechanics made use of a mathematical tool called Poisson brackets. Dirac showed that the laws of classical mechanics, once formulated in their Hamiltonian form, can be upgraded by suitably introducing \hbar into its equations, thereby yielding quantum mechanics correctly. In this way, Dirac was able to show how quantum mechanics naturally supersedes classical mechanics while reproducing the successes of classical mechanics. We will follow Dirac in this course¹.

2.2 Levels of abstraction

In order to follow Dirac's thinking, let us consider the levels of abstraction in mathematical physics: Ideally, one starts from abstract laws of nature and at the end one obtains concrete number predictions for measurement outcomes. In the middle, there is usually a hierarchy of mathematical problems that one has to solve.

In particular, in Newton's formulation of classical mechanics one starts by writing down the equations of motion for the system at hand. The equations of motion will generally contain terms of the type $m\ddot{x}$ and will therefore be of the type of differential equations. We begin our calculation by solving those differential equations, to obtain functions. These functions we then solve for variables. From those variables we eventually obtain some concrete numbers that we can compare with a measurement value. The hierarchy of abstraction is, therefore:



This begs the question if there is a level of abstraction above that of differential equations? Namely, can the differential equations of motion be obtained as the *solution* of

¹Actually, Schrödinger in his paper introducing the Schrödinger equation already tried to motivate his equation by an analogy with some aspect of Hamilton's work (the so called Hamilton Jacobi theory). This argument did not hold up. In fact, he came to correctly guess his equation by following his intuition that the discreteness of quantum phenomena might mathematically arise as discrete eigenvalues - which had been known to arise in solving wave equations.

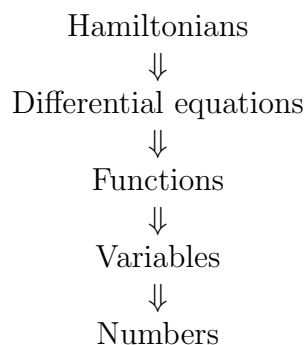
some higher level mathematical problem? The answer is yes, as Dirac remembered: Already in the first half of the 19th century, Lagrange, Hamilton and others had found this higher level formulation of classical mechanics. Their methods had proven useful for solving the dynamics of complicated systems, and some of those methods are still being used, for example, for the calculation of satellite trajectories. Dirac thought that if Newton's formulation of classical mechanics was not upgradable, it might be worth investigating if the higher level formulation of Hamilton might be upgradable to obtain quantum mechanics. Dirac succeeded and was thereby able to clearly display the similarities and differences between classical mechanics and the quantum mechanics of Heisenberg and Schrödinger. To see this is our first goal in this course.

Remark: For completeness, I should mention that there are two equivalent ways to present classical mechanics on this higher level of abstraction: One is due to Hamilton and one is due to Lagrange. Lagrange's formulation of classical mechanics is also upgradable, i.e., that there is a simple way to introduce \hbar to obtain quantum mechanics from it, as Feynman first realized in the 1940s. In this way, Feynman discovered a whole new formulation of quantum mechanics, which is called the path integral formulation. I will explain Feynman's formulation of quantum mechanics later in the course.

2.3 Classical mechanics in Hamiltonian formulation

2.3.1 The energy function H contains all information

What was Hamilton's higher level of abstraction? How can classical mechanics be formulated so that Newton's differential equations of motion are themselves the solution of a higher level mathematical problem? Hamilton's crucial observation was the following: the expression for the total energy of a system already contains the complete information about that system! In particular, if we know a system's energy function, then we can derive from it the differential equations of motion of that system. In Hamilton's formulation of classical mechanics the highest level description of a system is therefore through its energy function. The expression for the total energy of a system is also called the Hamiltonian. The hierarchy of abstraction is now:



As a very simple example, let us consider a system of two point² masses, m_1 and m_2 , which are connected by a spring with spring constant k . We write their respective position vectors as $\vec{x}^{(r)} = (x_1^{(r)}, x_2^{(r)}, x_3^{(r)})$ and their momentum vectors as $\vec{p}^{(r)} = (p_1^{(r)}, p_2^{(r)}, p_3^{(r)})$, where r is 1 or 2 respectively (we will omit the superscript (r) when we talk about one mass only). The positions and momenta are of course functions of time. Let us, therefore, keep in mind that for example $x_3^{(1)}$ is just a short hand notation for the function $x_3^{(1)}(t)$. Since this is a simple system, it is easy to write down its equations of motion:

$$\frac{d}{dt} x_i^{(r)} = \frac{p_i^{(r)}}{m_r} \quad (2.1)$$

$$\frac{d}{dt} p_i^{(1)} = -k(x_i^{(1)} - x_i^{(2)}) \quad (2.2)$$

$$\frac{d}{dt} p_i^{(2)} = -k(x_i^{(2)} - x_i^{(1)}) \quad (2.3)$$

Here, $r \in \{1, 2\}$ labels the objects and $i \in \{1, 2, 3\}$ labels their coordinates. Hamilton's great insight was that these equations of motion (as well as those of arbitrarily complicated systems) can all be *derived* from just one piece of information, namely the expression for the system's total energy H alone! This is to say that Hamilton discovered that the expression for the total energy is what we now call the generator of the time evolution. The Hamiltonian H , i.e., the total energy of the system, is the kinetic energy plus the potential energy. In our example:

$$H = \frac{(\vec{p}^{(1)})^2}{2m_1} + \frac{(\vec{p}^{(2)})^2}{2m_2} + \frac{k}{2} (\vec{x}^{(1)} - \vec{x}^{(2)})^2 \quad (2.4)$$

Here, $(\vec{p}^{(1)})^2 = \sum_{i=1}^3 (p_i^{(1)})^2$ etc. Now imagine that the system in question is instead a complicated contraption with plenty of wheels, gears, discs, levers, weights, strings, masses, bells and whistles. Using Newton's laws it is possible to determine the equations of motion for that system but it will be complicated and will typically involve drawing lots of diagrams with forces. Hamilton's method promises a lot of simplification here. We just write down the sum of all kinetic and potential energies, which is generally not so difficult, and then Hamilton's methods should yield the equations of motion straightforwardly. In practice we won't be interested in complicated contraptions. We'll be interested in systems such as molecules, quantum computers or quantum fields, which all can be quite complicated too.

²In this course, we will always restrict attention to point masses: all known noncomposite particles, namely the three types of electrons and neutrinos, six types of quarks, the W and Z particles (which transmit the weak force responsible for radioactivity), the gluons (which transmit the strong force responsible for the nuclear force) and the photon are all point-like as far as we know.

But what is the technique with which one can derive the equations of motion from a Hamiltonian, for example, Eqs.2.1-2.3 from Eq.2.4? Exactly how does the generator, H , of the time evolution generate the time evolution equations Eqs.2.1-2.3?

2.3.2 The Poisson bracket

The general procedure by which the equations of motion can be derived from a Hamiltonian H requires the use of a powerful mathematical operation, called ‘‘Poisson bracket’’³:

The Poisson bracket is a particular kind of multiplication: Assume that f and g are polynomials in terms of the positions and momenta of the system, say $f = -2p_1$ and $g = 3x_1^2 + 7p_3^4 - 2x_2^3p_1^3 + 6$. Then, the Poisson bracket of f and g is written as $\{f, g\}$ and the evaluation of the bracket will yield another polynomial in terms of the position and momenta of the system. In this case:

$$\{-2p_1, 3x_1^2 + 7p_3^4 - 2x_2^3p_1^3 + 6\} = 12x_1 \quad (2.5)$$

But how does one evaluate such a Poisson bracket to obtain this answer? The rules for evaluating Poisson brackets are tailor-made for mechanics. There are two sets of rules:

A) By definition, for each particle, the Poisson brackets of the positions and momenta are:

$$\{x_i, p_j\} = \delta_{i,j} \quad (2.6)$$

$$\{x_i, x_j\} = 0 \quad (2.7)$$

$$\{p_i, p_j\} = 0 \quad (2.8)$$

for all $i, j \in \{1, 2, 3\}$. Here, $\delta_{i,j}$ is the Kronecker delta, which is 1 if $i = j$ and is 0 if $i \neq j$. But these are only the Poisson brackets between linear terms. How to evaluate then the Poisson bracket between two polynomials? The second set of rules allow us to reduce this general case to the case of the Poisson brackets between linear terms:

B) By definition, the Poisson bracket of two arbitrary expressions in the positions and momenta, $f(x, p)$ and $g(x, p)$, obey the following rules:

$$\{f, g\} = -\{g, f\} \quad \text{antisymmetry} \quad (2.9)$$

$$\{cf, g\} = c\{f, g\}, \quad \text{for any number } c \quad \text{linearity} \quad (2.10)$$

$$\{f, g+h\} = \{f, g\} + \{f, h\} \quad \text{addition rule} \quad (2.11)$$

³Remark: In elementary particle physics there is a yet higher level of abstraction, which allows one to *derive* Hamiltonians. The new level is that of so-called ‘‘symmetry groups’’. The Poisson bracket operation plays an essential role also in the definition of symmetry groups. (Candidate quantum gravity theories such as string theory aim to derive these symmetry groups from a yet higher level of abstraction which is hoped to be the top level.)

$$\{f, gh\} = \{f, g\}h + g\{f, h\} \quad \text{product rule} \quad (2.12)$$

$$0 = \{f, \{g, h\}\} + \{h, \{f, g\}\} + \{g, \{h, f\}\} \quad \text{Jacobi id.} \quad (2.13)$$

Let us postpone the explanation for why these definitions had to be chosen in exactly this way⁴. For now, note that an immediate consequence of these rules is that the Poisson bracket of a number always vanishes:

$$\{c, f\} = 0 \quad \text{if } c \text{ is a number} \quad (2.14)$$

The point of the second set of rules is that we can use them to successively break down the evaluation of a Poisson bracket like that of Eq.2.5 into sums and products of expressions that can be evaluated by using the first set of rules, Eqs.2.6,2.7,2.8. Using the product rule we immediately obtain, for example:

$$\{x_3, p_3^2\} = \{x_3, p_3\}p_3 + p_3\{x_3, p_3\} = 1p_3 + p_3 \cdot 1 = 2p_3 \quad (2.15)$$

Exercise 2.1 Prove Eq.2.14.

Exercise 2.2 Show that $\{f, f\} = 0$ for any f .

Exercise 2.3 Assume that n is a positive integer.

a) Evaluate $\{x_1, p_1^n\}$

b) Evaluate $\{x_2^n, p_2\}$

Exercise 2.4 Verify Eq.2.5.

Exercise 2.5 Evaluate $\{3 + x_1p_2^2, p_1x_2^2p_3\}$.

Exercise 2.6 Show that the Poisson bracket is not associative by giving a counter example.

So far, we defined the Poisson brackets of polynomials in the positions and momenta of one point mass only. Let us now consider the general case of a system of n point masses, $m^{(r)}$ with position vectors $\vec{x}^{(r)} = (x_1^{(r)}, x_2^{(r)}, x_3^{(r)})$ and momentum vectors $\vec{p}^{(r)} = (p_1^{(r)}, p_2^{(r)}, p_3^{(r)})$, where $r \in \{1, 2, \dots, n\}$. How can we evaluate the Poisson brackets of expressions that involve all those positions and momentum variables? To this end, we need to define what the Poisson brackets in between positions and momenta of different particles should be. They are defined to be simply zero. Therefore, to summarize, we define the basic Poisson brackets of n masses as

$$\{x_i^{(r)}, p_j^{(s)}\} = \delta_{i,j} \delta_{r,s} \quad (2.16)$$

$$\{x_i^{(r)}, x_j^{(s)}\} = 0 \quad (2.17)$$

$$\{p_i^{(r)}, p_j^{(s)}\} = 0 \quad (2.18)$$

⁴If the product rule already reminds you of the product rule for derivatives (i.e., the Leibniz rule) this is not an accident. As we will see, the Poisson bracket can in fact be viewed as a sophisticated generalization of the notion of derivative.

where $r, s \in \{1, 2, \dots, n\}$ and $i, j \in \{1, 2, 3\}$. The evaluation rules of Eqs.2.9-2.13 are defined to stay just the same.

Exercise 2.7 *Mathematically, the set of polynomials in positions and momenta is an example of what is called a Poisson algebra. A general Poisson algebra is a vector space with two extra multiplications: One multiplication which makes the vector space into an associative algebra, and one (non-associative) multiplication $\{, \}$, called the Lie bracket, which makes the vector space into what is called a Lie algebra. If the two multiplications are in a certain sense compatible then the set is said to be a Poisson algebra. Look up and state the axioms of a) a Lie algebra, b) an associative algebra and c) a Poisson algebra.*

2.3.3 The Hamilton equations

Let us recall why we introduced the Poisson bracket: A technique that uses the Poisson bracket is supposed to allow us to *derive* all the differential equations of motion of a system from the just one piece of information, namely from the expression of the total energy of the system, i.e., from its Hamiltonian.

To see how this works, let us consider an arbitrary polynomial f in terms of the positions and momentum variables $x_i^{(r)}, p_j^{(s)}$ of the system in question, for example, something like $f = 7x_2^{(3)} \left(x_3^{(1)}\right)^3 - 2 \cos(4t^2)(p_1^{(1)})^7 + 3/2$. This f depends on time for two reasons: There is an explicit dependence on time through the cosine term, and there is an implicit dependence on time because the positions and momenta generally depend on time. According to Hamilton's formalism, the equation of motion for f is then given by:

$$\frac{df}{dt} = \{f, H\} + \frac{\partial f}{\partial t} \quad (2.19)$$

This famous equation is called the Hamilton equation. If you know how to evaluate Poisson brackets then the Hamilton equation Eq.2.19 encodes for you all of classical mechanics! Namely, given H , equation Eq.2.19 yields the differential equation of motion for any entity f by the simple procedure of evaluating the Poisson bracket on its right hand side.

If f is dependent on time only through x and p (say if we choose for f a polynomial in x and p 's with constant coefficients) then $\partial f / \partial t = 0$ and Hamilton's equation simplifies to:

$$\frac{d}{dt} f = \{f, H\} \quad (2.20)$$

Unless otherwise specified, we will in the following choose f 's that depend on time only through the x and p 's. In particular, the most important choices for f are of this kind: $f = x_i^{(r)}$ or $f = p_i^{(r)}$. For these choices of f we immediately obtain the fundamental

equations of motion:

$$\frac{d}{dt} x_i^{(r)} = \{x_i^{(r)}, H\} \quad (2.21)$$

$$\frac{d}{dt} p_i^{(r)} = \{p_i^{(r)}, H\} \quad (2.22)$$

Here is a concrete example: A single free particle of mass m possesses only kinetic energy. Its Hamiltonian is:

$$H = \sum_{j=1}^3 \frac{p_j^2}{2m} \quad (2.23)$$

By using this H in Eqs.2.21,2.22, we obtain the following equations of motion for the positions and momenta:

$$\frac{d}{dt} x_i = \left\{ x_i, \sum_{j=1}^3 \frac{p_j^2}{2m} \right\} = \frac{p_i}{m} \quad (2.24)$$

and

$$\frac{d}{dt} p_i = \left\{ p_i, \sum_{j=1}^3 \frac{p_j^2}{2m} \right\} = 0 \quad (2.25)$$

They agree with what was expected: $p_i = m\dot{x}_i$ and $\ddot{x}_i = 0$, where the dot indicates the time derivative. For another example, consider again the system of two point masses m_1, m_2 which are connected by a spring with spring constant k . Its Hamiltonian H was given in Eq.2.4. By using this H in Eqs.2.21,2.22 we should now be able to derive the system's equations of motion (as given in Eqs.2.1-2.3). Indeed:

$$\frac{d}{dt} x_i^{(r)} = \{x_i^{(r)}, H\} \quad (2.26)$$

$$= \frac{p_i^{(r)}}{m_r} \quad (2.27)$$

$$\frac{d}{dt} p_i^{(1)} = \{p_i^{(1)}, H\} \quad (2.28)$$

$$= -k(x_i^{(1)} - x_i^{(2)}) \quad (2.29)$$

$$\frac{d}{dt} p_i^{(2)} = \{p_i^{(2)}, H\} \quad (2.30)$$

$$= -k(x_i^{(2)} - x_i^{(1)}) \quad (2.31)$$

Let us omit the proof that Hamilton's formulation of classical mechanics always yields the same equations of motion as Newton's.

Exercise 2.8 Verify Eqs.2.24-2.31.

Exercise 2.9 Consider $f = gh$, where g and h are some polynomial expressions in the position and momentum variables. There are two ways to calculate df/dt : Either we use the Leibnitz rule, i.e., $\dot{f} = \dot{g}h + g\dot{h}$, and apply Eq.2.20 to both \dot{g} and \dot{h} , or we apply Eq.2.20 directly to gh and use the product rule (Eq.2.12) for Poisson brackets. Prove that both methods yield the same result.

This exercise shows that a property of the derivative on the left hand side of Eq.2.20 determines a rule for how the Poisson bracket had to be defined. In fact, such requirements of consistency are the main reason why the Poisson bracket is defined the way it is.

Exercise 2.10 Use Eq.2.13 to prove that:

$$\frac{d}{dt} \{f, g\} = \{\dot{f}, g\} + \{f, \dot{g}\} \quad (2.32)$$

2.3.4 Symmetries and Conservation laws

Our reason for reviewing the Hamiltonian formulation of mechanics is that it will be useful for the study of quantum mechanics. Before we get to that, however, let us ask why the Hamiltonian formulation of mechanics was useful for classical mechanics. It was, after all, developed more than half a century before quantum mechanics.

Sure, it was fine to be able to derive all the differential equations of motion from the one unifying equation:

$$\dot{f} = \{f, H\} \quad (2.33)$$

Ultimately, however, one obtained just the same equations of motion as Newton's methods would yield. Was there any practical advantage to using Hamilton's formulation of mechanics? Indeed, there is an important practical advantage: The main advantage of Hamilton's formulation of mechanics is that it gives us powerful methods for studying conserved quantities, such as the energy or angular momentum. To know conserved quantities usually significantly helps in solving the dynamics of complicated systems. This feature of Hamilton's formulation of mechanics will carry over to quantum mechanics, so studying it here will later help us also in quantum mechanics.

Consider a polynomial f in x and p 's with constant coefficients. Then, $\partial f/\partial t = 0$ and Eq.2.33 applies. We can easily read off from Eq.2.33 that *any such f is conserved in time if and only if its Poisson bracket with the Hamiltonian vanishes:*

$$\{f, H\} = 0 \quad \Rightarrow \quad \dot{f} = 0 \quad (2.34)$$

Consider, for example, a free particle. Its Hamiltonian is given in Eq.2.23. We expect of course that its momenta p_i are conserved. Indeed:

$$\dot{p}_i = \left\{ p_i, \sum_{j=1}^3 p_j^2/2m \right\} = 0 \quad (2.35)$$

For another example, consider a system whose Hamiltonian is any polynomial in x 's and p 's with constant coefficients. The proof that this system's energy is conserved is now fairly trivial:

$$\dot{H} = \{H, H\} = 0 \quad (2.36)$$

In order to be able to find solutions to the equations of motion of complicated real-life systems it is often crucial to find as many conserved quantities as possible.

For example, consider a 3-dimensional isotropic (i.e., rotation invariant) harmonic oscillator. Because of its symmetry under rotations, its angular momentum is conserved. But this oscillator has actually a much larger symmetry and therefore more conserved quantities. This is because a harmonic oscillator, being of the form $x^2 + p^2$ also possesses rotation symmetry in phase space. I will here only remark that this means that the 3-dimensional isotropic harmonic oscillator possesses $SO(3)$ rotational symmetry as well as a larger $SU(3)$ symmetry.

Powerful methods for discovering symmetries and constructing the implied conserved quantities for arbitrary systems have been developed on the basis of Eq.2.33 and the Poisson bracket. A key technique is that of so-called canonical transformations, i.e., of changes variables for which the Poisson brackets remain the same. You can find these methods in classical mechanics texts under the keywords "canonical transformations" and "Hamilton Jacobi theory".

In fact, Poisson bracket methods reveal a very deep one-to-one correspondence between conserved quantities and so-called symmetries. For example, the statement that an experiment on a system gives the same result no matter when we perform the experiment, is the statement of a "symmetry" which is called time-translation invariance symmetry. In practice, it means that the Hamiltonian of the system does not explicitly depend on time: $\partial H/\partial t = 0$. As we just saw, this implies energy conservation: $dH/dt = 0$.

Similarly, the statement that an experiment on a system gives the same result wherever we perform the experiment is the statement of space-translation symmetry. It implies and is implied by momentum conservation. Further, the statement that an experiment on a system gives the same result whatever the angular orientation of the experiment is the statement of rotation symmetry. It implies and is implied by angular momentum conservation.

These are examples of the so-called "Noether theorem", of Emmy Noether (1882-1935). Noether's theorem plays a crucial role both in practical applications, and in fundamental physics⁵. We will later come back to Noether's theorem.

⁵Mathematically, symmetries are described as groups (for example, the composition of two rotations yields a rotation and to every rotation there is an inverse rotation). In elementary particle physics, symmetry groups are one abstraction level higher than Hamiltonians: It has turned out that the complicated Hamiltonians which describe the fundamental forces, i.e., the electromagnetic, weak and strong force, are essentially derivable as being the simplest Hamiltonians associated with with three elementary symmetry groups.

Exercise 2.11 Show that in general, $dH/dt = \partial H/\partial t$ and give a nontrivial example.

Exercise 2.12 Consider the system with the Hamiltonian of Eq.2.4. **a)** Show that the total momentum is conserved. **b)** Prove the conservation of angular momentum about the center of mass by showing that its Poisson bracket with the Hamiltonian vanishes.

2.3.5 A representation of the Poisson bracket

In principle, we can evaluate any Poisson bracket $\{f, g\}$ by using the rules Eqs.2.6-2.12 if, as we assume, f and g are polynomials or well-behaved power series in the position and momentum variables. This is because the product rule allows us to break Poisson brackets that contain polynomials into factors of Poisson brackets that contain polynomials of lower degree. Repeating the process, we are eventually left with having to evaluate only Poisson brackets of linear terms, which can easily be evaluated using the first set of rules.

This is all good and fine but when f or g contain high or even infinite powers of the position and momentum variables, then the evaluation of the Poisson bracket $\{f, g\}$ can become rather tedious and cumbersome.

For practical purposes it is of interest, therefore, to have a shortcut for the evaluation of Poisson brackets. Indeed, for complicated f and g , the Poisson bracket $\{f, g\}$ can be evaluated usually faster by the following formula:

$$\{f, g\} = \sum_{r=1}^n \sum_{i=1}^3 \left(\frac{\partial f}{\partial x_i^{(r)}} \frac{\partial g}{\partial p_i^{(r)}} - \frac{\partial f}{\partial p_i^{(r)}} \frac{\partial g}{\partial x_i^{(r)}} \right) \quad (2.37)$$

Exercise 2.13 Evaluate $\{x^8 p^6, x^3 p^4\}$.

Exercise 2.14 Show that Eq.2.37 is indeed a representation of the Poisson bracket, i.e., that it always yields the correct answer. To this end, check that it obeys Eqs.2.6-2.12.

Exercise 2.15 Find the representation of the Hamilton equations Eq.2.19 and Eqs.2.21, 2.22 obtained by using Eq.2.37.

Remark: Some textbooks start with these representations of the Hamilton equations, along with the representation Eq.2.37 of the Poisson bracket - without reference to the Hamilton equations' more abstract origin in Eq.2.19 and Eqs.2.21, 2.22. This is unfortunate because those representations using Eq.2.37 do not carry over to quantum mechanics, while the more abstract equations Eq.2.19 and Eqs.2.21, 2.22 will carry over to quantum mechanics unchanged, as we will see.

2.4 Summary: The laws of classical mechanics

We already discussed that quantum mechanics must have strong similarities with classical mechanics, since it must reproduce all the successes of classical mechanics. This suggested that the laws of quantum mechanics might be a slight modification of Newton's laws which would somehow contain Planck's constant h . Since this did not work, we reformulated the laws of classical mechanics on a higher level of abstraction, namely in Hamilton's form. Before we now try to guess the laws of quantum mechanics, let us restate Hamilton's formulation of classical mechanics very carefully:

The starting point is the energy function H of the system in question. It is called the Hamiltonian, and it is an expression in terms of the position and momentum variables of the system. Then, assume we are interested in the time evolution of some quantity f which is also a polynomial in the x and p 's (say with constant coefficients). Then we can derive the equation of motion for f through:

$$\frac{d}{dt} f = \{f, H\} \quad (2.38)$$

In particular, f can be chosen to be any one of the position and momentum variables of the system, and we obtain their equations of motion as Eqs.2.21,2.22. In order to obtain explicit differential equations from Eqs.2.38,2.21,2.22 we evaluate the Poisson bracket on its right hand side. To this end, we use the definitions Eqs.2.6-2.13. The so-obtained differential equations are then solved to obtain the positions $x_i^{(r)}(t)$ and momenta $p_i^{(r)}(t)$ as functions of time.

We note that the Poisson bracket which is defined by the axioms Eqs.2.6-2.12 possesses an often convenient explicit representation through Eq.2.37. We need to keep in mind, however, that Eq.2.37 merely provides a convenient shortcut for evaluating the Poisson bracket. This shortcut only works in classical mechanics. In quantum mechanics, there will also be a representation of the Poisson bracket but it will look very different from Eq.2.37.

2.5 Classical field theory

This section is a mere comment. In classical mechanics, the dynamical variables are the positions of particles, together with their velocities or momenta. For each particle there are three degrees of freedom of position and momenta.

In a field theory, such as Maxwell's theory, positions (and momenta) are not dynamical variables. After all, unlike a particle that moves around, a field can be everywhere at the same time. In the case of a field theory, what is dynamical is its amplitude.

Consider say a scalar field ϕ . At every point x in space it has an amplitude $\phi(x, t)$ that changes in time with a ‘velocity’ of $\dot{\phi}(x, t)$ which we may call the canonically conjugate momentum field: $\pi(x, t) := \dot{\phi}(x, t)$. Unlike the three degrees of freedom that particle possesses, a field therefore possesses uncountably many degrees of freedom, one at each position x . Now one can define the Poisson brackets of the first kind for them in analogy to the Poisson brackets for particles:

$$\{\phi(x, t), \pi(x', t)\} = \delta^3(x - x') \quad (2.39)$$

$$\{\phi(x, t), \phi(x', t)\} = 0 \quad (2.40)$$

$$\{\pi(x, t), \pi(x', t)\} = 0 \quad (2.41)$$

Here, $\delta^3(x - x')$ is the three dimensional Dirac delta distribution. The second set of Poisson brackets is unchanged, i.e., it is still given by Eqs.2.9-2.13. The energy of the classical field, i.e., its Hamiltonian, is:

$$H(\phi, \pi) = \int d^3x \frac{1}{2} \left(\pi(x, t)^2 + \sum_{i=1}^3 (\partial_i \phi(x, t))^2 + m^2 \phi(x, t)^2 \right) \quad (2.42)$$

The Hamilton equation Eq.2.38 is unchanged.

Exercise 2.16 (*Bonus question*) *Derive the equations of motion for $\phi(x, t)$ and $\pi(x, t)$. Combine the two equations by eliminating $\pi(x, t)$.*

The combined equation is the so-called Klein Gordon equation. The Dirac equation and the Maxwell equations can be treated similarly, although with some small extra complications because the amplitudes of these fields are not scalar but are vectorial and spinorial respectively.

Chapter 3

Quantum mechanics in Hamiltonian form

We formulated the laws of classical mechanics on a higher level of abstraction, as summarized in Sec.2.4 because classical mechanics appeared to be not upgradeable when written in Newton's formulation. We are now ready to upgrade the more abstract Hamiltonian laws of classical mechanics to obtain quantum mechanics. A modification is needed which is small enough to preserve all the successes of classical mechanics while it must also introduce \hbar in order to correctly predict quantum mechanical effects.

For example, could it be that Eq.2.38 needs to be modified to obtain quantum mechanics? Could the correct equation be, say, $\frac{d}{dt} f = \{f, H\} + \hbar$ or $\frac{d}{dt} f = \{f + \hbar^2/f, H\}$, where \hbar is Planck's constant? Those two equations can't be right, of course, already because the units generally don't match. Could it be then that to obtain quantum mechanics we will have to change the definitions for the Poisson bracket? Could it be that the definition Eq.2.12 needs to be changed? This, of course, is unlikely too because the definitions for the Poisson bracket were fixed by consistency conditions (recall e.g. Ex.2 of Sec.2.9). The structure of our Poisson algebra is quite tightly constrained.

No, the necessary upgrade of the Hamiltonian formalism is actually much more subtle! Let us remember that when we defined the Poisson algebra structure in the previous section we did not make any assumptions about the mathematical nature of the functions $x(t)$ and $p(t)$ (let us omit writing out the indices). In particular, we did not make the assumption that these functions are number-valued. We can start with a Hamiltonian, i.e., a polynomial in the x and p and then by using the rules of the Poisson bracket we can derive the differential equations of motion. In the process, we never need to assume that the functions $x(t)$ and $p(t)$ are number valued. Could it be that the $x(t)$ and $p(t)$ need not be number valued and that this holds the key to upgrading classical mechanics to obtain quantum mechanics? Actually yes!

Before we get to this, we have to consider though that we actually did assume the x and p to be number valued at one specific point at the very end of the previous

chapter. There, we wrote down a convenient representation of the Poisson bracket in Eq.2.37, and there we needed the x and p to be number-valued - because to use this convenient representation we needed to be able to differentiate with respect to the x and p . We can conclude from this that if allowing the $x(t)$ and $p(t)$ to be something else than number valued is the key to upgrading to quantum mechanics, then the Poisson bracket will not be representable any more through Eq.2.37.

In fact, as we will see, this is how it will play out. Everything we did in the previous chapter, except for the representation Eq.2.37 will still exactly hold true in quantum mechanics. In particular, the differential equations of motion derived from the Hamiltonian will look exactly the same in quantum and classical mechanics. That's because they are derived from the same Hamiltonian polynomial in the x and p by using the same rules for the Poisson bracket. But then, if not in the equations of motion, how does the upgrade involve \hbar at all?

3.1 Reconsidering the nature of observables

At this point, let us reconsider the very basics: How do the symbols we write on paper relate to real systems? We measure a system with concrete measurement devices in the lab, for example, devices for the measurement of positions and devices for the measurement of momenta. As usual, we invent for each kind of measurement a symbol, say $x_i^{(r)}$ and $p_i^{(r)}$. At this stage we need to be careful not to over-interpret these symbols. At this stage, these symbols have nothing to do (yet) with numbers, vectors, matrices, operators or bananas. Instead, these symbols are merely names for kinds of measurement. We need to find out more about the nature of these $x_i^{(r)}$ and $p_i^{(r)}$.

Now according to our everyday experience, the operation of a position measurement device does not interfere with the operation of a momentum measurement device: it seems that we can always measure both, positions and momenta. For example, GPS units are able to tell both position and velocity at the same time to considerable accuracy. It is tempting to assume, therefore, that there is no limit, in principle, to how accurately positions and velocities can be determined. And that would mean that we can let each of the symbols $x_i^{(r)}(t)$ and $p_i^{(r)}(t)$ stand for its measurement devices's output number at time t .

It is at this very point, namely when we make the assumption that positions and momenta can be accurately measured simultaneously, that we make the assumption that the symbols $x_i^{(r)}(t)$ and $p_i^{(r)}(t)$ can be represented mathematically as number-valued functions of time. And number-valued functions have the property of being commutative:

$$x_i^{(r)} p_j^{(s)} - p_j^{(s)} x_i^{(r)} = 0 \quad (3.1)$$

Since measurement values cannot be just any number but always come out real, we

also have the law:

$$\left(x_i^{(r)}\right)^* = x_i^{(r)} \quad \text{and} \quad \left(p_j^{(s)}\right)^* = p_j^{(s)} \quad (3.2)$$

Similarly, we have $H^* = H$. Technically, the $*$ -operation is an example of what is called an involution.

Exercise 3.1 *Find and list the defining properties of an involution.*

The statements above, namely that position and momentum measurements are compatible and come out as real numbers are indeed a nontrivial part of the laws of classical mechanics. For completeness we should have included them in the summary of classical mechanics in Sec.2.4.

A reality property of the form of Eq.3.2 will still be true in quantum mechanics. But the commutativity property expressed in Eq.3.1 and its underlying assumption that the operation of position and momentum measurement devices do not interfere with another needs to be abandoned and upgraded. It turns out that position and momentum measurements are like taking a shower and doing sports. It matters in which sequence one does them.

3.2 The canonical commutation relations

For the remainder of this course, we will need a way to make it transparent in every equation whether a variable is number valued or not. To this end, we will decorate variables that may not be number valued with a hat, for example, \hat{H} , \hat{p} , \hat{x} , or more specifically $\hat{x}_i^{(r)}$ and $\hat{p}_i^{(r)}$ for each position and momentum measurement device. Now how can the interference of the measurement devices mathematically be expressed as properties of the symbols $\hat{x}_i^{(r)}$ and $\hat{p}_i^{(r)}$?

According to classical mechanics one would be able to operate all measurement devices all the time and they would not interfere with another. We could therefore choose the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ to stand for the number-valued outcomes of those measurements as functions of time. Crucially, the fact that we can't actually know positions and momenta simultaneously means that we can no longer choose the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ to stand simply for number-valued outcomes of those measurements as functions of time.

Mathematically, it was the commutativity law of Eq.3.1 which expressed that in classical mechanics the symbols $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ can be represented as number valued functions. Could it be that Eq.3.1 has to be modified to include \hbar so that the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ become non-commutative and therefore can no longer be number-valued functions?

Are position and momentum measurements non commuting similar to how doing sports and having a shower don't commute?

It was Dirac who first realized that all of the Poisson algebra structure that we defined above can be kept (and therefore the ability to derive the equations of motion), while changing one little thing: allowing the symbols $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(s)}(t)$ to be non-commutative, though only in a particular way. Consistency with the Poisson algebra structure imposes strict conditions on the form that this noncommutativity can take. Namely, following Dirac, let us consider the Poisson bracket

$$\{\hat{u}_1\hat{u}_2, \hat{v}_1\hat{v}_2\} \quad (3.3)$$

where $\hat{u}_1, \hat{u}_2, \hat{v}_1, \hat{v}_2$ are arbitrary polynomials in the variables $\hat{x}_i^{(r)}$ and $\hat{p}_j^{(s)}$. Expression Eq.3.3 can be evaluated in two ways and, of course, any noncommutativity of the $\hat{x}_i^{(r)}$ and $\hat{p}_j^{(s)}$ has to be such that both ways yield the same outcome:

$$\begin{aligned} \{\hat{u}_1\hat{u}_2, \hat{v}_1\hat{v}_2\} &= \hat{u}_1\{\hat{u}_2, \hat{v}_1\hat{v}_2\} + \{\hat{u}_1, \hat{v}_1\hat{v}_2\}\hat{u}_2 \\ &= \hat{u}_1(\hat{v}_1\{\hat{u}_2, \hat{v}_2\} + \{\hat{u}_2, \hat{v}_1\}\hat{v}_2) + (\hat{v}_1\{\hat{u}_1, \hat{v}_2\} + \{\hat{u}_1, \hat{v}_1\}\hat{v}_2)\hat{u}_2 \end{aligned} \quad (3.4)$$

This must agree with:

$$\begin{aligned} \{\hat{u}_1\hat{u}_2, \hat{v}_1\hat{v}_2\} &= \hat{v}_1\{\hat{u}_1\hat{u}_2, \hat{v}_2\} + \{\hat{u}_1\hat{u}_2, \hat{v}_1\}\hat{v}_2 \\ &= \hat{v}_1(\hat{u}_1\{\hat{u}_2, \hat{v}_2\} + \{\hat{u}_1, \hat{v}_2\}\hat{u}_2) + (\hat{u}_1\{\hat{u}_2, \hat{v}_1\} + \{\hat{u}_1, \hat{v}_1\}\hat{u}_2)\hat{v}_2 \end{aligned} \quad (3.5)$$

Thus:

$$\{\hat{u}_1, \hat{v}_1\}(\hat{v}_2\hat{u}_2 - \hat{u}_2\hat{v}_2) = (\hat{v}_1\hat{u}_1 - \hat{u}_1\hat{v}_1)\{\hat{u}_2, \hat{v}_2\} \quad (3.6)$$

Since this has to hold for all possible choices of $\hat{u}_1, \hat{u}_2, \hat{v}_1, \hat{v}_2$, we require all expressions \hat{u}, \hat{v} in the position and momentum variables to obey:

$$\hat{v}\hat{u} - \hat{u}\hat{v} = k\{\hat{u}, \hat{v}\} \quad (3.7)$$

Here, k must be independent of \hat{u} and \hat{v} and must be commuting with everything. But what value does k take?

The case $k = 0$ would be classical mechanics, because it implies that all expressions in the positions and momenta commute.

However, it turns out that in order to eventually yield the correct experimental predictions (we will later see how), we have to set $k = -i\hbar/2\pi$, i.e., we have

$$\hat{u}\hat{v} - \hat{v}\hat{u} = i\hbar\{\hat{u}, \hat{v}\} \quad (3.8)$$

where we used the convenient definition:

$$\hbar = \frac{h}{2\pi} \quad (3.9)$$

In particular, choosing for \hat{u} and \hat{v} the variables $x_i^{(r)}$ and $p_j^{(s)}$ using Eqs.2.6-2.8, we now obtain the quantum mechanical commutation relations for n particles:

$$\hat{x}_i^{(r)} \hat{p}_j^{(s)} - \hat{p}_j^{(s)} \hat{x}_i^{(r)} = i\hbar \delta_{i,j} \delta_{r,s} \quad (3.10)$$

$$\hat{x}_i^{(r)} \hat{x}_j^{(s)} - \hat{x}_j^{(s)} \hat{x}_i^{(r)} = 0 \quad (3.11)$$

$$\hat{p}_i^{(r)} \hat{p}_j^{(s)} - \hat{p}_j^{(s)} \hat{p}_i^{(r)} = 0 \quad (3.12)$$

Notice that we did not modify the rules of the Poisson bracket. We still have:

$$\{x_i, p_j\} = \delta_{i,j} \quad (3.13)$$

$$\{x_i, x_j\} = 0 \quad (3.14)$$

$$\{p_i, p_j\} = 0 \quad (3.15)$$

$$\{f, g\} = -\{g, f\} \quad \text{antisymmetry} \quad (3.16)$$

$$\{cf, g\} = c\{f, g\}, \quad \text{for any number } c \quad \text{linearity} \quad (3.17)$$

$$\{f, g+h\} = \{f, g\} + \{f, h\} \quad \text{addition rule} \quad (3.18)$$

$$\{f, gh\} = \{f, g\}h + g\{f, h\} \quad \text{product rule} \quad (3.19)$$

$$0 = \{f, \{g, h\}\} + \{h, \{f, g\}\} + \{g, \{h, f\}\} \quad \text{Jacobi id.} \quad (3.20)$$

Because the rules for the Poisson bracket did not change with the upgrade to quantum mechanics, one arrives in quantum mechanics at the same equations of motion as in classical mechanics. This is as long as one does not unnecessarily commute any variables.

The equations Eqs.3.10-3.12 are called the ‘‘Canonical Commutation Relations’’ (CCRs). The appearance of the imaginary unit i will be necessary to ensure that measurements are predicted as real numbers, as we will see below. Eqs.3.11,3.12 express that position measurements among another and momentum measurements among another do not interfere. Only positions and momenta of the same particle and in the same direction, i.e., for $i = j$ and $r = s$, are noncommutative.

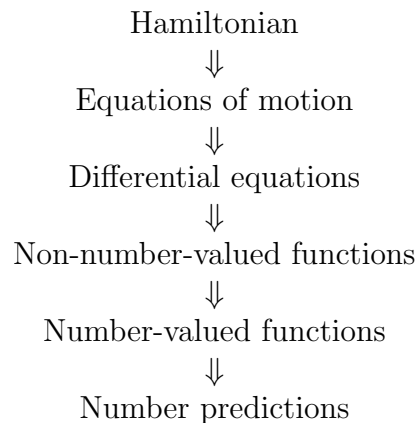
In conclusion, we upgrade classical mechanics to quantum mechanics by first formulating classical mechanics in Hamiltonian form to identify the Poisson algebra structure. Then, we realize that while keeping all the rules for the Poisson bracket intact, there is still the freedom to make the associative multiplication in the Poisson algebra non-commutative, parametrized by some constant k . Nature chose the modulus of k to be nonzero though very small, namely \hbar . The fact that the Poisson bracket stays the same when quantizing explains why quantum mechanics has the same equation of motion as does classical mechanics. The fact that \hbar is so small explains why it took long to discover quantum mechanics.

In spite of the tremendous similarity between classical and quantum mechanics from this perspective, quantum mechanical calculations will in practise look rather different

from classical calculations. This is because they will require representations of the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ variables as explicit non-number valued mathematical entities that obey the commutation relations. Even though there is only a slight noncommutativity in the Poisson algebra of quantum mechanics its representations will necessarily look quite different from the representation of the classical commutative Poisson algebra. This will explain why the Schrödinger equation looks rather different from Newton's equations.

3.3 From the Hamiltonian to the Equations of Motion

In quantum mechanics, as in classical mechanics, the energy function \hat{H} encodes all information about the system. It is still called the Hamiltonian and it is in general some polynomial (or well-behaved power series) in the positions and momenta $\hat{x}_i^{(r)}$ and $\hat{p}_i^{(r)}$ of the system. In quantum mechanics, the sequence of steps that lead from the Hamiltonian down to concrete number predictions for experiments can be drawn schematically in this form:



So far, we can perform the first step, namely the derivation of the equations of motion from the Hamiltonian: Assume that we are interested in the time evolution of some \hat{f} which is a polynomial in the \hat{x} and \hat{p} 's (say with constant coefficients). Then we can derive the equation of motion for \hat{f} through:

$$\frac{d}{dt} \hat{f} = \{\hat{f}, \hat{H}\} \quad (3.21)$$

where $\{, \}$ is the usual Poisson bracket, as defined in Eqs.2.6-2.12. In particular, \hat{f} can be chosen to be any one of the position and momentum variables of the system, so that

we obtain for their equations of motion, exactly as in Eqs.2.21,2.22:

$$\frac{d}{dt} \hat{x}_i^{(r)} = \{\hat{x}_i^{(r)}, \hat{H}\} \quad (3.22)$$

$$\frac{d}{dt} \hat{p}_i^{(r)} = \{\hat{p}_i^{(r)}, \hat{H}\} \quad (3.23)$$

By evaluating the Poisson bracket on the right hand side of Eqs.3.22,3.23 these equations of motion then become differential equations for the entities $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$. Clearly, the resulting equations of motion will be analogous to those of classical mechanics. The entities $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ must also still obey Eq.3.2, which in quantum mechanics is usually written as:

$$\left(\hat{x}_i^{(r)}\right)^\dagger = \hat{x}_i^{(r)} \quad \text{and} \quad \left(\hat{p}_j^{(s)}\right)^\dagger = \hat{p}_j^{(s)} \quad (3.24)$$

We will call any polynomial or well-behaved power series \hat{f} in the \hat{x} and \hat{p} an “observable”, if it obeys $\hat{f}^\dagger = \hat{f}$. As we will see later, the condition $\hat{f}^\dagger = \hat{f}$ will indeed imply that measurement outcomes are predicted as real numbers. In addition to the position variables $\hat{x}_i^{(r)}(t)$ and momentum variables $\hat{p}_j^{(s)}(t)$ also, e.g., the energy $\hat{H}(t)$ and the angular momentum variables $\hat{L}_i(t)$ are observables.

While classical mechanics requires the Poisson algebra to be commutative, quantum mechanics requires that the equations of motion be solved by entities $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ which are noncommutative:

$$\hat{x}_i^{(r)} \hat{p}_j^{(s)} - \hat{p}_j^{(s)} \hat{x}_i^{(r)} = i\hbar \delta_{i,j} \delta_{r,s} \quad (3.25)$$

$$\hat{x}_i^{(r)} \hat{x}_j^{(s)} - \hat{x}_j^{(s)} \hat{x}_i^{(r)} = 0 \quad (3.26)$$

$$\hat{p}_i^{(r)} \hat{p}_j^{(s)} - \hat{p}_j^{(s)} \hat{p}_i^{(r)} = 0 \quad (3.27)$$

Technically, we will, therefore, need to solve differential equations of motion with noncommutative entities. In practice, the task is then to start from the top level of abstraction, the Hamiltonian of a system, then working one’s way down by calculating the equations of motion, and then solving them to obtain something from which eventually predictions can be made of numbers that can be measured in experiments on the system. In the next section, we will investigate what kind of noncommutative mathematical objects, such as, for example, matrices, may represent the position and momentum variables.

Exercise 3.2 *For classical mechanics, formula Eq.2.37 provided a convenient representation of the Poisson bracket. However, Eq.2.37 is not a valid representation of the Poisson bracket in the case of quantum mechanics. In quantum mechanics, we have a (not so convenient) representation of the Poisson bracket through Eq.3.8:*

$$\{\hat{u}, \hat{v}\} = \frac{1}{i\hbar} (\hat{u}\hat{v} - \hat{v}\hat{u}) \quad (3.28)$$

Use this representation to evaluate the Poisson bracket $\{\hat{x}^2, \hat{p}\}$.

Let us introduce an often-used notation, called “the commutator”:

$$[A, B] := A B - B A \quad (3.29)$$

For simplicity, assume that \hat{H} and \hat{f} are polynomials in the positions and momenta which depend on time only through their dependence on the \hat{x} and \hat{p} . Then the Hamilton equation Eq.3.21 holds and takes the form:

$$i\hbar \frac{d}{dt} \hat{f}(t) = [\hat{f}(t), \hat{H}] \quad (3.30)$$

$$(3.31)$$

and, in particular:

$$i\hbar \frac{d}{dt} \hat{x}_i^{(r)}(t) = [\hat{x}_i^{(r)}(t), \hat{H}]$$

$$i\hbar \frac{d}{dt} \hat{p}_i^{(r)}(t) = [\hat{p}_i^{(r)}(t), \hat{H}] \quad (3.32)$$

These equations are called the Heisenberg equations of motion.

Remark: The particular method by which in the past few sections we upgraded classical mechanics to quantum mechanics is called canonical quantization. I covered it in some detail because of its importance: Essentially the same method was used to find quantum electrodynamics starting from Faraday and Maxwell’s electromagnetism. All the quantum field theories of elementary particles can be derived this way. Even string theory and most other modern attempts at finding the unifying theory of quantum gravity try to employ canonical quantization. I should mention too that the problem of canonical quantization for constrained classical systems was also pioneered by Dirac but is still not fully understood. A simple example of a constrained system would be a particle that is constrained to move on a curved surface. The most important constrained system is general relativity.

Exercise 3.3 *Reconsider the system with the Hamiltonian Eq.2.4, which consists of two particles which are attracted to another through a harmonic force (a force which is proportional to their distance). In practice, for example the force that binds diatomic molecules and the force that keeps nucleons (i.e., neutrons and protons) inside a nucleus are essentially harmonic. In those cases the effect of \hbar cannot be neglected. One obtains the correct quantum theoretic Hamiltonian from the classical Hamiltonian of Eq.2.4 by simply placing hats on the x and p ’s. Find explicitly all the equations which the $\hat{x}_i^{(r)}$ and $\hat{p}_j^{(r)}$ (where $r \in \{1, 2\}$) of this system must obey.*

Exercise 3.4 *To obtain the quantum Hamiltonian from the classical Hamiltonian and vice versa by placing or removing hats on the x and p 's is generally not as straightforward as in the previous exercise! Namely, there can occur so-called "ordering ambiguities": Consider the two Hamiltonians $\hat{H}_1 = \hat{p}^2/2m + a\hat{x}\hat{p}^2\hat{x}$ and $\hat{H}_2 = \hat{p}^2/2m + b\hat{p}\hat{x}^2\hat{p}$, where a and b are constants with suitable units. These Hamiltonians are identical in classical mechanics. Check whether or not they differ in quantum mechanics, i.e., when x and p no longer commute. Give an example of two quantum Hamiltonians that strongly differ but that are the same classically. Remark: In principle, experiments are needed to decide how to resolve ordering ambiguities in the process of quantization of a Hamiltonian.*

3.4 From the Hamiltonian to predictions of numbers

In the framework of classical mechanics we know how to descend from the most abstract level, where the system is described simply by giving its Hamiltonian H , down to the concrete level of predicting numbers for measurement outcomes. We will now have to develop methods for descending in quantum mechanics from the level of the Hamiltonian down to the concrete predictions of numbers in experiments.

As the first step, we are able use the Hamiltonian to derive the equations of the system. Since the Poisson brackets have not changed, the equations of motion resemble those of classical mechanics. Due to the requirements of noncommutativity, in Eqs.3.25-3.27, the equations of motion can no longer be interpreted as differential equations for number-valued functions. For the next steps, in order to be able to solve the equations of motion as explicit differential equations, the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ must be viewed as functions whose values are noncommutative mathematical objects.

What could those mathematical objects be? Let us recall that every mathematical object can be viewed as a map, if need be, as a trivial map. We may, therefore, restrict our search to maps. We should be able to represent the symbols $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as map-valued functions of time. A simple class of maps is formed by the linear maps. For example, matrices are linear maps and they are generally noncommutative. Also, for example, the derivative operator, D , acting on functions as $D : g(\lambda) \rightarrow \partial_\lambda g(\lambda)$ is a linear map, because it obeys $\partial_\lambda (c g(\lambda)) = c \partial_\lambda g(\lambda)$ for all numbers c and because $\partial_\lambda (g_1(\lambda) + g_2(\lambda)) = \partial_\lambda g_1(\lambda) + \partial_\lambda g_2(\lambda)$.

Exercise 3.5 *Verify that the multiplication operator, M , which maps $M : g(\lambda) \rightarrow \lambda g(\lambda)$ is linear.*

Here, we use the definition that the term operator is used for any linear maps on an infinite-dimensional vector space. Since spaces of functions are infinite dimensional vectors spaces, the multiplication and differentiation maps are called operators. Similarly, infinite by infinite matrices describe linear maps on infinite dimensional vector

spaces. They are, therefore, also called operators.

Exercise 3.6 Show that D and M are noncommutative by calculating $(DM - MD)g(\lambda)$.

Is it possible to represent the symbols $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as linear map-valued functions of time, say as matrix-valued functions in time? As we will see, the answer is yes. In fact, essentially¹ all linear representations are physically equivalent! This is the content of the Stone von Neumann theorem, which we will later cover more precisely. But could it be that one should use representations of the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as nonlinear maps instead? Non-linear representations have been considered in the literature, see, e.g., articles by S. Weinberg. According to current knowledge, however, nonlinear quantum theories generally lead to physically incorrect predictions and we will therefore here only consider linear representations.

3.4.1 A matrix representation

As an example, let us consider how $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ can be represented as matrix-valued functions in time, and how, therefore, the abstract equations of motion can be represented as explicit matrix differential equations for matrix-valued functions of time. For simplicity, we will restrict attention to the one-dimensional case with just one $\hat{x}(t)$ and one $\hat{p}(t)$.

The canonical commutation relations are of course to hold at all times. To begin with, let us ask whether it is possible to find two $N \times N$ matrices $\hat{x}(t_0)$ and $\hat{p}(t_0)$ so that at the starting time, t_0 , of the experiment the canonical commutation relations hold:

$$\hat{x}(t_0) \hat{p}(t_0) - \hat{p}(t_0) \hat{x}(t_0) = i\hbar \mathbf{1} \quad (3.33)$$

Here, $\mathbf{1}$ is the identity matrix. At this point it is useful to remember that the trace of matrices $\text{Tr}(A) = \sum_n A_{n,n}$ is linear and cyclic:

$$\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B) \quad \text{and} \quad \text{Tr}(AB) = \text{Tr}(BA) \quad (3.34)$$

Exercise 3.7 Verify Eqs.3.34.

We see that the trace of the left hand side of Eq.3.33 vanishes, while the trace of the right hand side is $i\hbar N$. Thus, there are in fact no $N \times N$ matrices, i.e., there are no finite-dimensional matrices $\hat{x}(t_0)$ and $\hat{p}(t_0)$ that obey the commutation relation Eq.3.33! For infinite dimensional matrices, however, the trace may be ill-defined on both sides, and our argument then does not apply. In fact, there exist infinite-dimensional matrices which do obey the commutation relation.

¹There are exceptions but nature does not appear to make use of those.

In order to find such matrices we start by defining the $\infty \times \infty$ dimensional matrix:

$$a = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & & \\ 0 & 0 & \sqrt{2} & 0 & 0 & & \\ 0 & 0 & 0 & \sqrt{3} & 0 & & \\ 0 & 0 & 0 & 0 & \sqrt{4} & & \\ 0 & 0 & 0 & 0 & 0 & & \\ & & & & & \ddots & \end{pmatrix} \quad (3.35)$$

The hermitean conjugate is:

$$a^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & & \\ \sqrt{1} & 0 & 0 & 0 & 0 & & \\ 0 & \sqrt{2} & 0 & 0 & 0 & & \\ 0 & 0 & \sqrt{3} & 0 & 0 & & \\ 0 & 0 & 0 & \sqrt{4} & 0 & & \\ & & & & & \ddots & \end{pmatrix} \quad (3.36)$$

Their commutation relation is:

$$aa^\dagger - a^\dagger a = \mathbf{1} \quad (3.37)$$

Since they are noncommutative we should decorate a and a^\dagger with hats but we will follow convention by not putting hats on them.

Exercise 3.8 *Verify Eq.3.37.*

Using a and a^\dagger , we can now represent $\hat{x}(t_0)$ and $\hat{p}(t_0)$ as matrices that obey the canonical commutation relation, namely by defining:

$$\hat{x}(t_0) = L(a^\dagger + a) \quad (3.38)$$

and

$$\hat{p}(t_0) = \frac{i\hbar}{2L}(a^\dagger - a) \quad (3.39)$$

Here, L is some arbitrary real number with units of length, which we need because \hat{x} has a unit of length while a and a^\dagger do not have units. The definitions are such that the realness conditions Eqs.3.24 are obeyed, i.e., such that the matrices are formally² hermitean: $\hat{x}^\dagger(t_0) = \hat{x}(t_0)$ and $\hat{p}^\dagger(t_0) = \hat{p}(t_0)$.

Exercise 3.9 *Verify that the two matrices defined in Eqs.3.38,3.39 are formally hermitean.*

²I am writing here “formally” hermitean, because the issue of whether a matrix is hermitean, symmetric or self-adjoint is quite subtle for infinite-dimensional matrices, as we will see later.

Exercise 3.10 Show that the hermitean conjugation of matrices reverses the order, i.e., that if A and B are linear maps, then $(AB)^\dagger = B^\dagger A^\dagger$.

Technically, † is a map from the Poisson algebra into itself which is called an involution because it is its own inverse. Because it also reverses the order it is called an “anti” algebra mapping: First multiplying and then applying † is the same as first applying † and then multiplying, up to the reversal of the order.

We can now understand the appearance of the imaginary unit i in the canonical commutation relations: If we apply † to the commutation relations $\hat{x}\hat{p} - \hat{p}\hat{x} = k\mathbf{1}$ we obtain $\hat{p}\hat{x} - \hat{x}\hat{p} = k^*\mathbf{1}$, i.e., we obtain $k = -k^*$. Thus, k has to be imaginary. And of course it is: $k = i\hbar$.

Exercise 3.11 Verify that the two matrices defined in Eqs.3.38,3.39 do obey the commutation relation Eq.3.33. You may use the results of Exercise 3.8.

3.4.2 Solving the matrix differential equations

In the case of the free particle which moves in one dimension, the Hamiltonian is $\hat{H} = \hat{p}^2/2m$. The Hamilton equations or, equivalently, the Heisenberg equations, yield the abstract equations of motion:

$$\frac{d}{dt}\hat{x}(t) = \frac{1}{m}\hat{p}(t) \quad (3.40)$$

$$\frac{d}{dt}\hat{p}(t) = 0 \quad (3.41)$$

Let us view these equations as matrix equations. Using the results of the previous section, it becomes clear that these equations are solved through

$$\hat{x}(t) = \hat{x}(t_0) + \frac{(t - t_0)}{m}\hat{p}(t_0) \quad (3.42)$$

and

$$\hat{p}(t) = \hat{p}(t_0), \quad (3.43)$$

where $\hat{x}(t_0)$ and $\hat{p}(t_0)$ are the matrices of Eqs.3.38,3.39. Concretely, by substituting in the matrices a and a^\dagger , we have:

$$\hat{x}(t) = \begin{pmatrix} 0 & \sqrt{1} \left(L - \frac{i\hbar(t-t_0)}{2Lm} \right) & 0 & & \\ \sqrt{1} \left(L + \frac{i\hbar(t-t_0)}{2Lm} \right) & 0 & \sqrt{2} \left(L - \frac{i\hbar(t-t_0)}{2Lm} \right) & & \\ 0 & \sqrt{2} \left(L + \frac{i\hbar(t-t_0)}{2Lm} \right) & 0 & & \\ & & & \ddots & \end{pmatrix} \quad (3.44)$$

$$\hat{p}(t) = \begin{pmatrix} 0 & -\sqrt{1}\frac{i\hbar}{2L} & 0 & & \\ \sqrt{1}\frac{i\hbar}{2L} & 0 & -\sqrt{2}\frac{i\hbar}{2L} & & \\ 0 & \sqrt{2}\frac{i\hbar}{2L} & 0 & & \\ & & & \ddots & \\ & & & & \ddots \end{pmatrix} \quad (3.45)$$

Exercise 3.12 Show that the matrices $\hat{x}(t)$ and $\hat{p}(t)$ obey at all times $t > t_0$ all the quantum mechanical conditions, i.e., the equations of motion, the hermiticity condition, and the commutation relation.

Remark: We had arranged for the commutation relation and the hermiticity condition to hold only at the initial time t_0 . Having solved the equations of motion we found that the commutation relation and the hermiticity conditions continue to hold at all times t . This is nontrivial but it is not a coincidence. As we will soon see, the quantum mechanical time evolution of all systems³ preserves the commutation relations and hermiticity. The preservation of the commutation relations is of course the preservation of the Poisson bracket. And we have in classical and quantum mechanics that the Poisson brackets between the positions and momenta are preserved by the dynamics through the Hamilton equation: $d/dt \{\hat{x}, \hat{p}\} = \{\{\hat{x}, \hat{p}\}, \hat{H}\} = \{1, \hat{H}\} = 0$.

Exercise 3.13 The vibrational degree of freedom of a diatomic molecule such as HF, CO or HCl can be described as a harmonic oscillator. If x is the deviation from the equilibrium distance between the two nuclei, then:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2$$

a) Use the ansatz (an ansatz is an educated guess)

$$\hat{x}(t) = \xi(t)a + \xi^*(t)a^\dagger,$$

where ξ is a to-be-determined complex-valued function of time, to calculate $\hat{x}(t)$, $\hat{p}(t)$ and $\hat{H}(t)$. Verify that your solution obeys the equations of motion, the CCRs and the hermiticity conditions.

b) The function ξ is not unique. There is a particularly important choice of ξ : Find that solution function ξ such that \hat{H} is a diagonal matrix. Remark: Making such choices wisely is particularly useful in quantum field theory where each wavelength (and therefore frequency ω) has its own harmonic oscillator degree of freedom.

3.4.3 From matrix-valued functions to number predictions

Let us assume now that we have solved a quantum mechanical problem in the sense that we have found explicit matrix-valued functions $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(j)}(t)$ which obey

³With the possible exception of systems that involve black hole horizons or other gravitational horizons.

the canonical commutation relations, the hermiticity conditions, and the equations of motion. For example, the quantum mechanical problem of the free particle in one dimension is solved by the matrix-valued functions given in Eqs.3.44,3.45.

How then are we to further descend the ladder of abstraction, down to the most concrete level, namely that of predictions of numbers that will be measured in experiments? How can we extract from those matrix-valued functions the information which will let us predict the outcome of say a position or a momentum measurement at some time t ?

To this end, assume that we have solved the dynamics, i.e., that we have calculated $\hat{x}(t)$ and $\hat{p}(t)$ as explicit matrix-valued functions of time. Then we can also immediately write down the time evolution of any polynomial $f(\hat{x}(t), \hat{p}(t))$ of the $\hat{x}(t)$ and $\hat{p}(t)$ that we may be interested in. For example, we can write down the time evolution of the Hamiltonian $\hat{H}(\hat{x}(t), \hat{p}(t))$ or say the third component of the angular momentum: $\hat{L}_3(t) = \hat{x}_2(t)\hat{p}_3(t) - \hat{p}_2(t)\hat{x}_3(t)$.

Mathematically, the problem now boils down to this question: Given such a matrix valued function $\hat{f}(\hat{x}(t), \hat{p}(t))$, let us pick a time t . Then $\hat{f}(\hat{x}(t), \hat{p}(t))$ at the time t is an explicit infinite by infinite matrix. How can we extract from that matrix a prediction for the number-valued outcome, $\bar{f}(t)$, of an experiment that measures \hat{f} ? For example, say we want to measure the position at time t . Given the matrix $\hat{x}(t)$ by what method can we extract from that matrix a prediction for the position $\bar{x}(t)$?

To find that method, let us start with the observation that the method by which we extract a number-valued prediction $\bar{f}(t)$ from a matrix $\hat{f}(t)$ should not depend on the basis in which we write down the matrix $\hat{f}(t)$. The reason is that a change of basis in a vector space yields merely another way to write down the same linear map. And physical predictions should not depend on any human choice of how (i.e., in which basis) to write down a map. This means that $\bar{f}(t)$ should be a scalar formed from the matrix $\hat{f}(t)$.

Now how can one get a scalar from a matrix? By using the scalar product of course. So assume that we are given two column vectors with coefficients ψ_i and ϕ_i . Then

$$\sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}(t) \phi_m \tag{3.46}$$

is scalar. Could this be the prediction for the measurement outcome? No this cannot be quite right because this quantity is generally complex while measurement outcomes are of course always real numbers. This leaves us with the conjecture that the predicted

value $\bar{f}(t)$ for a measurement of $\hat{f}(t)$ at time t is⁴:

$$\bar{f}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}(t) \psi_m \quad (3.47)$$

This number is guaranteed to be real for all quantities \hat{f} which are hermitean, $\hat{f}^\dagger = \hat{f}$, i.e., for all quantities that in classical mechanics are real-valued, as it should be.

Let us check that Eq.3.47 always comes out real if \hat{f} is what is called an observable, i.e., if \hat{f} obeys $\hat{f}^\dagger = \hat{f}$, i.e., if $\hat{f}_{s,r}^* = \hat{f}_{r,s}$. Indeed⁵:

$$\bar{f}(t)^* = \left(\sum_{r,s=1}^{\infty} \psi_r^* \hat{f}_{r,s}(t) \psi_s \right)^* = \sum_{r,s=1}^{\infty} \psi_r \hat{f}_{r,s}^*(t) \psi_s^* = \sum_{r,s=1}^{\infty} \psi_s^* \hat{f}_{s,r}(t) \psi_r = \bar{f}(t) \quad (3.48)$$

So this works! And, for example, the predictions for measurements at time t of the position, momentum, angular momentum or energy are, therefore:

$$\bar{x}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{x}_{n,m}(t) \psi_m \quad (3.49)$$

$$\bar{p}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{p}_{n,m}(t) \psi_m \quad (3.50)$$

$$\bar{L}^{(i)}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{L}_{n,m}^{(i)}(t) \psi_m \quad (3.51)$$

$$\bar{H}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{H}_{n,m}(t) \psi_m \quad (3.52)$$

⁴There is a more general possibility: clearly, \bar{f} should depend on \hat{f} linearly (they have the same units) and this leaves the possibility that $\bar{f}(t) = \text{Tr}(\rho \hat{f}(t)) = \sum_{i,j} \rho_{i,j} \hat{f}_{j,i}(t)$ where ρ is some hermitean matrix. As we will see, we will make use of this possibility when describing a system whose initial conditions we are not quite sure of, where we can only give probabilities of initial conditions, i.e., where we don't even know for sure even the initial expectation values of the various observables. In this case, we say that the system is described by a mixed state and the matrix ρ is called the mixed state matrix. The term "mixed" is used because the uncertainties in predictions then have mixed origins - both from quantum effects but also from our ignorance of the system to start with. What we consider so far are so-called pure states ψ , which are the special case where $\rho_{i,j} = \psi_i^* \psi_j$. For general ρ , the normalization condition $\sum_i \psi_i^* \psi_i = 1$ is replaced by $\text{Tr}(\rho) = 1$.

⁵This argument is correct for finite dimensional matrices only. Quantum mechanics requires infinite-dimensional matrices where the sums are infinite sums and analytic issues therefore arise. That there is a fundamental difference between finite and infinite dimensional vector spaces we saw earlier when we found that the canonical commutation relations do not possess finite dimensional representations. We will, therefore later revisit the issue of hermiticity.

3.5 Initial conditions

In order to obtain those scalar functions for predictions, $\bar{f}(t)$, we had to introduce a complex vector with infinitely many coefficients!

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{pmatrix} \quad (3.53)$$

We are free to choose ψ , and it is a convention to choose a vector ψ of unit length, i.e., for which $\sum_n \psi_n^* \psi_n = 1$. We call such vectors normalized. For example, the vector ψ could be given by:

$$\psi = \frac{1}{5} \begin{pmatrix} 4 \\ 3i \\ 0 \\ 0 \\ \vdots \end{pmatrix} \quad (3.54)$$

There are, of course, infinitely many choices for such vectors ψ . But what does such a vector ψ mean? What is the physics of it?

Obviously the choice of ψ determines the predictions that we make for all measurements at any time t . The choice of ψ in fact even determines what the expected outcome is for measurements at the initial time t_0 ! And this observation is important:

Let us remember that when we solved the equations of motion to obtain those matrix-valued functions $\hat{x}(t)$ and $\hat{p}(t)$, we did not have an opportunity to specify the initial conditions of the experimental setup. We did not have an opportunity to specify, for example, whether the particle was initially fast or slow, or where the particle was at the initial time t_0 .

Now we have an opportunity to specify how the system started off at time t_0 : The choice of ψ encodes our specification of the initial state of the system: by choosing a vector ψ we are choosing an experimental starting condition at time t_0 . Namely, by choosing ψ , we are choosing what measurement outcomes to expect if we measure right at t_0 . For example:

$$\bar{x}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{x}_{n,m}(t_0) \psi_m \quad (3.55)$$

$$\bar{p}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{p}_{n,m}(t_0) \psi_m \quad (3.56)$$

$$\bar{L}^{(i)}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{L}_{n,m}^{(i)}(t_0) \psi_m \quad (3.57)$$

$$\bar{H}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{H}_{n,m}(t_0) \psi_m \quad (3.58)$$

And of course also for any \hat{f} obeying $\hat{f}^\dagger = \hat{f}$, we have

$$\bar{f}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}(t_0) \psi_m \quad (3.59)$$

3.6 Emergence of probabilities

The fact that we have equations of motion and that initial conditions are to be specified is not unusual of course, because this is also what one does in classical mechanics. But what seems unusual here is that we have to specify so many initial conditions. In order to choose a vector ψ that describes the initial state of our quantum system, we get to choose its infinitely many coefficients ψ_i (with the only constraint being that ψ should be normalized). Why are there so many initial conditions? In classical mechanics, it sufficed to specify the initial position and the initial momentum and that determined the initial state completely! And from that initial condition you could then calculate $x(t)$ and $p(t)$. And, in classical mechanics, once you have $x(t)$ and $p(t)$ you automatically also have the predictions for any $f(x(t), p(t))$.

So let us ask: in quantum mechanics, does it really matter which values we choose for the infinitely many coefficients of ψ or do perhaps only two of these coefficients matter? Isn't it the case that once we can make a prediction $\bar{x}(t)$ and $\bar{p}(t)$ we can also predict any $\bar{f}(\hat{x}(t), \hat{p}(t))$? That would mean:

$$\bar{f}(\hat{x}(t), \hat{p}(t)) = f(\bar{x}(t), \bar{p}(t)) \quad (3.60)$$

Actually, this equation does not hold in quantum mechanics! Just because we have a prediction for positions and momentum values does not mean that we have a prediction for other measurements such as the energy or the momentum!

Exercise 3.14 Give a counter example for Eq.3.60. To this end, write out Eq.3.60 explicitly, i.e., in matrix form, for the case $\hat{f}(\hat{x}(t), \hat{p}(t)) = \hat{x}^2$. Then choose a suitable normalized ψ so that Eq.3.60 is seen to be violated. (It is not difficult to find such a ψ , almost every one will do.)

On one hand, this explains why, mathematically, we have to specify so many initial conditions in quantum mechanics, namely all those coefficients ψ_i . But what is the physics of this?

To see this, let us have a closer look at the observation that knowing the number-valued predictions $\bar{x}(t), \bar{p}(t)$ does not alone suffice to make predictions of the outcome of other measurements f . Namely, this means, in particular, that even if we have

a prediction for, say, the position, $\bar{x}(t)$, we actually don't automatically have also a prediction for the square of the position: $\overline{x^2(t)}$.

From the perspective of classical mechanics this is weird. If we have a prediction for the position shouldn't we have a prediction for the square of it too? Well yes. Except, if the prediction is not certain, if it has some statistical spread or uncertainty. Then, even in classical physics, the square of the expectation value of a measurement need not be the expectation value of the square. In fact, as is well known, the statistical variance $(\Delta(Q))^2$ of any variable Q is defined as the difference between the two

$$(\Delta(Q))^2 := \overline{(Q - \bar{Q})^2} = \overline{Q^2} - \bar{Q}^2 \quad (3.61)$$

which, as the middle term shows, is also the mean squared deviation from the mean. ΔQ is called the uncertainty in Q .

Exercise 3.15 *Spell out the step of the second equality in Eq.3.61.*

Now in quantum mechanics, if \hat{f} is an observable, i.e., if it is hermitean, so is \hat{f}^2 . This is because if $\hat{f}^\dagger = \hat{f}$ then $(\hat{f}^2)^\dagger = \hat{f}^2$. It is important that in quantum mechanics they are independent observables. For example, their initial values can be specified independently. This is because, as always in statistics, we generally have $\overline{f^2} \neq \bar{f}^2$. The average of some squared numbers is rarely the same as the square of the average of those numbers: generally, e.g. $(a_1^2 + a_2^2)/2 \neq ((a_1 + a_2)/2)^2$. Interestingly, this means that quantum mechanics also allows us to calculate the variance in the set of measurement outcomes of each observable \hat{f} , namely through this mean value:

$$(\Delta f(t))^2 = \overline{(f(t) - \bar{f}(t))^2} = \overline{f^2(t)} - \bar{f}^2 \quad (3.62)$$

For example, from Eq.3.62:

$$(\Delta x(t))^2 = \sum_{r,s,t=1}^{\infty} \psi_r^* \hat{x}_{r,s} \hat{x}_{s,t} \psi_t - \left(\sum_{u,v=1}^{\infty} (\psi_u^* \hat{x}_{u,v} \psi_v) \right)^2 \quad (3.63)$$

Here, the number $\hat{x}_{r,s}$ is the matrix element of the matrix \hat{x} with indices r, s . Similarly, given ψ , also all the higher moments of the probability distributions of positions and momenta are predictable, such as $\overline{x^n(t)}$ and $\overline{p^n(t)}$.

What we have found, therefore, is that in quantum mechanics, since the predictions generally (i.e., except for special cases) obey

$$\bar{f}(\hat{x}(t), \hat{p}(t)) \neq f(\bar{x}(t), \bar{p}(t)), \quad (3.64)$$

the predictions should come with uncertainty. They should be statistical. Our predictions for observables $\bar{f}(t)$ such as $\bar{x}(t), \bar{p}(t), \bar{H}(t), \bar{L}_i(t), \dots$ can only be predictions for expectation values. There will generally be a spread of outcomes, i.e., there will be

nontrivial variances. This crucial finding, namely that the predicted expectation values of observables \hat{f} are largely independent, also explains why it takes so many numbers, namely all the coefficients ψ_i to specify initial conditions in quantum mechanics. In effect, one has to specify all the initial expectation values of all the possible observables. And there are infinitely many polynomials $\hat{f}(\hat{x}, \hat{p})$ that obey $\hat{f}(\hat{x}, \hat{p})^\dagger = \hat{f}(\hat{x}, \hat{p})$. In effect, we need to specify so many initial conditions because we need to fix an entire probability distribution.

We therefore arrive at this interpretation: Assume we run an ensemble of experiments, each with the same initial experimental setup, i.e., all described by the same initial state vector ψ . Then, having calculated the solutions to the equations of motion as in Eqs.3.44,3.45, we can calculate the means of any observable $\bar{f}(t)$, such as, for example, position $\bar{x}(t)$ and momentum $\bar{p}(t)$ that will be measured in an ensemble of measurements by using Eqs.3.49,3.50. In fact, we can only predict means. But this also includes the ability to predict the variance of any variable, because the variance of an observable is a mean value too, as see Eq.3.62 shows.

Remark: Also for systems of many particles, such as a molecule, all observables $\hat{f}(t)$, such as $\hat{x}_i^{(r)}$ and $\hat{p}_j^{(s)}$, can be represented as matrices acting in the same vector space. The choice of ψ in this vector space determines how all the constituent particles start off, because all $\bar{f}(t_0)$ are determined, including, e.g., $\bar{x}_i^{(r)}(t_0) = \sum_i \psi_i^* \hat{x}_i^{(r)}(t_0) \psi_i$ etc.

Remark: We say that ψ is the so-called *state vector* of the system. It is clear from Eqs.3.47 that if two state vectors ψ and ϕ differ only by a phase, $\psi_n = e^{i\alpha} \phi_n$ for all n , then they yield the same predictions and are, therefore, describing the same state. The state vector of any system is defined only up to an overall phase.

Remark: Conversely, assume we prepare an experimental setup for which we know the ensemble mean values at initial time $\bar{f}(t)$ for all observables \hat{f} :

$$\bar{x}(t_0) = a_1, \quad \bar{p}(t_0) = a_2, \quad \bar{x^2}(t_0) = a_3, \quad \bar{p^2}(t_0) = a_4, \quad \bar{x^3}(t_0) = a_5, \quad \dots \quad (3.65)$$

There are, clearly, infinitely many observables \hat{f} (with $\hat{f}^\dagger = \hat{f}$) whose initial values can be specified. Which ψ describes a system with so-specified initial conditions? ψ can be calculated from Eqs.3.65, which are infinitely many equations for the unknown vector components $\{\psi_i\}$ in terms of the given coefficients $\{a_j\}$:

$$\sum_{i,j} \psi_i^* \hat{x}_{i,j}(t_0) \psi_j = a_1, \quad \sum_{i,j} \psi_i^* \hat{p}_{i,j}(t_0) \psi_j = a_2, \quad \sum_{i,j,k} \psi_i^* \hat{x}_{i,k} \hat{x}_{k,j}(t_0) \psi_j = a_3, \quad \dots \quad (3.66)$$

Mathematically, we are dealing with a so-called moment problem. We must ask, in particular, what conditions the coefficients $\{a_i\}$ must obey for there to exist a matching state ψ . Physically, this is the question which initial conditions can actually occur in an experimental setup. We anticipate, of course, that the $\{a_i\}$ cannot be chosen completely arbitrarily because some observables are interfering variables. This question will later lead us to Heisenberg's famous uncertainty relations.

Exercise 3.16 *Verify that ψ of Eq.3.54 is normalized. For this choice of ψ , calculate explicitly the expectation values $\bar{x}(t), \bar{p}(t)$ as well as the uncertainties in those predictions, i.e., the standard deviations $\Delta x(t)$ and $\Delta p(t)$ for the free particle. Your results should show that neither the position nor the momentum are predicted with certainty at any time, not even at the initial time t_0 . The fact that $\Delta x(t)$ grows in time expresses that a momentum uncertainty over time leads to increasing position uncertainty. $\Delta p(t)$ remains constant in time, expressing that the momentum of a free particle, no matter what value it has, remains unchanged.*

Finally, we also have to conclude that if we measure our system at time say t_1 then we gain information and we have to update our initial state vector accordingly to a new initial state vector ψ' which is such as to encode our knowledge of the initial state of the system at t_1 . We will later revisit the question of this so-called wave function collapse.

3.7 The Hilbert space and Dirac's notation

So far, we solved the equations of motion for matrix-valued functions $\hat{x}(t)_{ij}, \hat{p}(t)_{ij}$ from which for every observable $\hat{f} = \hat{f}^\dagger$, we get a matrix-valued function $\hat{f}(t)_{ij}$. And we imposed the initial conditions by choosing vector coefficients ψ_i .

Of course, the coefficients ψ_i themselves are numbers and not vectors. The ψ_i are the coefficients of a vector which, following Dirac, we will call a “ket” and denote by $|\psi\rangle$. The numbers ψ_i are the coefficients of $|\psi\rangle$ in a basis given by some kets $|b_n\rangle$:

$$|\psi\rangle = \sum_n \psi_n |b_n\rangle \quad (3.67)$$

Similarly, the matrix elements $\hat{x}(t)_{ij}$ themselves are numbers. They are the coefficients of a linear map $\hat{x}(t)$ from the vector space into itself, in the basis of the $|b_n\rangle$. If we change basis in the vector space then the coefficients of the matrices $\hat{x}(t)_{ij}, \hat{p}(t)_{ij}$ etc and the coefficients ψ_i of the vector ψ will change accordingly.

We say that the abstract vector $|\psi\rangle$ is an element of the vector space that is represented as a column of vector coefficients. And we say that the abstract $\hat{x}(t), \hat{p}(t), \dots$ are abstract maps of the vector space into itself which are represented by matrices. When the vector space in question is infinite dimensional, one also calls these maps of the vector space into itself “operators”. Every choice of basis⁶ gives us a representation of operators and vectors as matrices and columns of numbers respectively.

By design, the predictions of a quantum theory only depend on the abstract $\hat{x}(t), \hat{p}(t), \dots$ and on the abstract ψ . This is because the predictions, $\bar{x}(t), \bar{p}(t)$, etc are scalars, as we discussed in the previous section.

⁶Notice that this also means, because ψ has unit length, that we could always rotate whichever our starting basis may be so that in the new basis the vector $|\psi\rangle$ is the vector with the coefficients $\psi_i = \delta_{i,0}$. This is usually not convenient but it is possible.

For practical calculations, one usually chooses a convenient basis to work in and this will include also so-called continuous representations that one may loosely think of as working with matrices and columns that have continuous indices. But it is often also possible, and conceptually much clearer, to work directly with the vectors $|\psi\rangle$ and operators $\hat{x}(t), \hat{p}(t)$, etc rather than through their coefficients in a basis. In order to be able to do these abstract calculations, we will need to clarify exactly what the rules are for the manipulation of these vectors and the operators that act on them.

3.7.1 Hilbert spaces

We begin by reviewing the definition of a complex vector space. In brief, it is a set, \mathcal{H} , which is an abelian group over the complex numbers:

Definition: Any set \mathcal{H} is called a *complex vector space*, if it a) possesses an operation $\mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}$ called “addition” which obeys the rules for a commutative (i.e., also called abelian) group and if b) the set \mathcal{H} has a multiplication $\mathbb{C} \times \mathcal{H} \rightarrow \mathcal{H}$ obeying the following axioms for all $|v\rangle, |w\rangle \in \mathcal{H}$ and for all $\alpha, \beta \in \mathbb{C}$:

$$(\alpha + \beta)|v\rangle = \alpha|v\rangle + \beta|v\rangle \quad (3.68)$$

$$\alpha(|v\rangle + |w\rangle) = \alpha|v\rangle + \alpha|w\rangle \quad (3.69)$$

$$(\alpha\beta)|v\rangle = \alpha(\beta|v\rangle) \quad (3.70)$$

$$1|v\rangle = |v\rangle \quad (3.71)$$

Notice that *every* set obeying these axioms is a complex vector space. To illustrate this point, consider, for example, the set of 3×2 matrices with complex entries. We can add such matrices and we can multiply them with complex numbers. It is easy to check that the above axioms are obeyed, so this set is a complex vector space. Also, consider, for example, the set of complex-valued continuous functions on \mathbb{R}^4 , such as $g(x_1, x_2, x_3, x_4) = x_1 \cos(x_2 x_3^x) e^{i x_2 x_4}$ or $h(x_1, x_2, x_3, x_4) = x_1 + i(x_2 + x_4^3)$. We can add such functions and we can multiply them with complex numbers and we will always get yet another continuous function on \mathbb{R}^4 . It is easy to check that the set of complex-valued continuous functions on \mathbb{R}^4 is a complex vector space (which is infinite dimensional). Also, and this will be very important, given any complex vector space one can construct another complex vector space, called the dual vector space, \mathcal{H}^* .

Definition: For any complex vector space, \mathcal{H} , we define the complex vector space called its *dual space*, \mathcal{H}^* , as the set of continuous⁷ linear maps $\hat{v} \rightarrow \mathbb{C}$. We call the

⁷Aren't all linear maps continuous? Well no, not necessarily in infinite-dimensional spaces! Consider for example the map $\phi : \mathcal{H} \rightarrow \mathbb{C}$ that maps vectors into numbers through $|\psi\rangle \rightarrow \sum_n n! \psi_n$ in some basis. ϕ is linear but clearly arbitrarily small changes to the vector it acts on can change the image arbitrarily strongly.

elements of \mathcal{H}^* “bra” vectors and use the notation $\langle r| \in \mathcal{H}^*$. They linearly map elements of \mathcal{H} into complex numbers:

$$\langle r| : \mathcal{H} \rightarrow \mathbb{C} \quad (3.72)$$

$$\langle r| : |v\rangle \rightarrow \langle r|v\rangle \quad (3.73)$$

That they are linear maps means:

$$\langle r| : (\alpha|v\rangle + \beta|w\rangle) \rightarrow \alpha\langle r|v\rangle + \beta\langle r|w\rangle \quad (3.74)$$

$$(3.75)$$

Exercise 3.17 Verify that \mathcal{H}^* is a complex vector space.

Definition: A complex vector space is said to possess a *scalar product* or also called *inner product* if it possesses a map $\mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$. Among other conditions listed below, the map is required to be linear in the right argument. This means that a scalar product provides a so-called *hermitean conjugation map*, denoted $\dagger : \mathcal{H} \rightarrow \mathcal{H}^*$: Feed the scalar product the left argument and you obtain an element of \mathcal{H}^* . The hermitean conjugation maps every element $|v\rangle \in \mathcal{H}$ to an element in \mathcal{H}^* , which we choose to call by the same name, i.e., $\langle v| \in \mathcal{H}^*$:

$$\dagger : \mathcal{H} \rightarrow \mathcal{H}^* \quad (3.76)$$

$$\dagger : |v\rangle \rightarrow \langle v| \quad (3.77)$$

Using this definition, we can express the full list of requirements on a scalar product using the bra-ket notation: A scalar product is required to map any two vectors $|v\rangle, |w\rangle \in \mathcal{H}$ into a complex number denoted by $\langle v|w\rangle \in \mathbb{C}$, obeying the conditions:

$$\langle u|(\alpha|v\rangle + \beta|w\rangle) = \alpha\langle u|v\rangle + \beta\langle u|w\rangle \quad (3.78)$$

$$\langle v|w\rangle^* = \langle w|v\rangle \quad (3.79)$$

$$\langle v|v\rangle \geq 0 \quad (3.80)$$

$$\langle v|v\rangle = 0 \quad \text{only if } |v\rangle = 0 \quad (3.81)$$

Definition: A complex vector space equipped with a scalar product is called a unitary vector space or inner product space or also pre-Hilbert space.

Definition: The “length” or “norm”, $\| |v\rangle \|$, of a vector $|v\rangle \in \mathcal{H}$ is defined as $\| |v\rangle \| = \langle v|v\rangle^{1/2}$. We say that $|v\rangle$ is normalized if $\| |v\rangle \| = 1$.

Definition: The distance $d(|v\rangle, |w\rangle)$ between two vectors is defined as $\|(|v\rangle - |w\rangle)\|$.

Since quantum mechanics requires infinite-dimensional representations, we will have to be concerned with sequences of vectors and their convergence. For example, if we are to make sense of a series such as Eq.3.67 we need to define what it means to sum up an infinite number of basis vectors.

Definition: We say that a sequence of vectors converges, i.e., $\lim_{n \rightarrow \infty} |v_n\rangle = |v\rangle$, iff (i.e., if and only if) it is true that $\lim_{n \rightarrow \infty} d(|\psi_n\rangle, |\psi\rangle) = 0$.

Definition: A sequence $\{|v_n\rangle\}$ is called fundamental (or Cauchy), if and only if $\lim_{n, m \rightarrow \infty} d(|v_n\rangle, |v_m\rangle) = 0$.

Definition: A pre-Hilbert space is called a Hilbert space if all its fundamental sequences converge. In particular, all vectors in a Hilbert space have finite length.

How then, do these abstract concepts relate to the concrete vector components ψ_n and explicit matrix elements $\hat{x}_{r,s}, \hat{p}_{r,s}$ and more general $\hat{f}_{r,s}$?

To see this connection, we need the concept of Hilbert basis:

3.7.2 Hilbert bases

Definition: We say that a set of orthonormal vectors $\{|b_n\rangle\}$ (i.e., a set of vectors obeying $\langle b_n | b_m \rangle = \delta_{n,m}$) is a Hilbert basis for \mathcal{H} , if all $|\psi\rangle \in \mathcal{H}$ have a unique representation of the form:

$$|\psi\rangle = \sum_n \psi_n |b_n\rangle \quad (3.82)$$

Definition: A Hilbert space \mathcal{H} is called “separable”, if it possesses a countable Hilbert basis.

Since we have succeeded above in representing the canonical commutation relations in a representation of matrices and column vectors we know already that in quantum mechanics we can work with a separable Hilbert space. In fact, in quantum mechanics, it always suffices to consider a separable Hilbert space. Separable Hilbert spaces are the smallest Hilbert spaces if we disregard finite-dimensional cases. Recall that there are different kinds of infinity: following Cantor, we say that two sets have the same number of elements, or have the same “cardinality”, if their elements can be mapped bijectively into another. For example, there are just as many even numbers, as there are natural numbers, as there are rational numbers. There are, however, many more real numbers, i.e., their cardinality is higher. Quantum field theory does appear to require nonseparable Hilbert spaces whose bases have the cardinality of the real numbers. If nature possesses a shortest length scale then a separable Hilbert space could suffice for quantum field theory as well.

Theorem: All Hilbert bases of a given Hilbert space have the same cardinality.

This theorem implies for quantum mechanics, because we know its Hilbert space is separable, that all its Hilbert bases are countable. This means that whatever Hilbert basis we may choose, our vectors will always be represented as column vectors and our \hat{x} , \hat{p} and general \hat{f} will be represented as matrices. It also means that the Hilbert space of a 1-dimensional harmonic oscillator has the same dimension (namely countable infinity) as does the Hilbert space of an elephant⁸.

As we will see later, there is a way to use what amounts to a continuous basis, but these come at the cost of the “basis vectors” having infinite length and therefore not being in the Hilbert space.

3.7.3 Discrete wave functions and matrix representations

Following up on Eq.3.82, it is important now to note that the coefficients ψ_n of the vector $|\psi\rangle$ in the $\{|b_n\rangle\}$ basis can be calculated through the scalar products:

$$\psi_n = \langle b_n | \psi \rangle \quad (3.83)$$

Eq.3.83 is easily verified by applying $\langle b_m |$ from the left in Eq.3.82:

$$\begin{aligned} \langle b_m | \psi \rangle &= \langle b_m | \sum_n \psi_n | b_n \rangle \\ &= \sum_n \psi_n \langle b_m | b_n \rangle \\ &= \sum_n \psi_n \delta_{n,m} \\ &= \psi_m \end{aligned}$$

Definition: We call the set of coefficients $\psi_n = \langle b_n | \psi \rangle$ the “wave function” of the state $|\psi\rangle$ in the basis $\{|b_n\rangle\}$.

Remark: We have not yet introduced “continuous bases” such as the “position basis” $|x\rangle$ because they come with another set of issues to discuss. But it may be worth mentioning already that for them the corresponding definition will be: We call the set of coefficients $\psi(x) = \langle x | \psi \rangle$ the “wave function” of the state $|\psi\rangle$ in the basis of the $|x\rangle$.

⁸In quantum field theory, which supersedes quantum mechanics, the Hilbert space is, in principle, non-separable. That’s because every wave vector, of which there are continuously infinitely many, has its own harmonic oscillator. To avoid this problem, in practice we can consider a finite region of spacetime so that the set of wave vectors becomes discrete. If we further consider only wavelengths larger than some very small cutoff then only finitely many wave vectors remain and the Hilbert space is then separable. It is generally considered likely that there is a smallest length scale in nature, due to quantum gravity effects.

A very useful observation is that the elements $|b_n\rangle$ of any Hilbert basis can be used to provide a representation of the identity map in this way:

$$\mathbf{1} = \sum_n |b_n\rangle\langle b_n| \quad (3.84)$$

This is called the resolution of the identity in terms of the basis vectors $|b_n\rangle$. For example, using Eq.3.83:

$$\mathbf{1}|\psi\rangle = \sum_{n=1}^{\infty} |b_n\rangle\langle b_n|\psi\rangle = \sum_n \psi_n |b_n\rangle \quad (3.85)$$

Resolutions of the identity are commonly used for the important task of turning abstract equations into equations for concrete matrices and concrete vector coefficients. For example, let us insert the identity in the expression for the length of a vector:

$$\langle\psi|\psi\rangle = \langle\psi|\mathbf{1}|\psi\rangle \quad (3.86)$$

$$= \sum_n \langle\psi|b_n\rangle\langle b_n|\psi\rangle \quad (3.87)$$

$$= \psi_n^* \psi_n \quad (3.88)$$

Since all vectors $|\psi\rangle \in \mathcal{H}$ are of finite length, we conclude that the vector components must be square summable:

$$\sum_n \psi_n^* \psi_n < \infty \quad (3.89)$$

Further, if matrix elements $\hat{f}_{r,s}$ are given in the $\{|b_n\rangle\}$ basis, then they define a linear map \hat{f} :

$$\hat{f} = \sum_{r,s=1}^{\infty} |b_r\rangle \hat{f}_{r,s} \langle b_s| \quad (3.90)$$

Conversely, the matrix elements of a linear map \hat{f} in the basis of the $|b_n\rangle$ can be calculated as scalar products:

$$\hat{f}_{r,s} = \langle b_r|\hat{f}|b_s\rangle \quad (3.91)$$

This equation follows from Eq.3.90 by applying to it $\langle b_r|$ and $|b_m\rangle$ from the left and right respectively.

3.7.4 The domain of operators

We need to be careful because in infinite-dimensional Hilbert spaces linear maps normally cannot be allowed to act on all vectors of the Hilbert space! This is because if $|\psi\rangle \in \mathcal{H}$ and if therefore $\sum_n \psi_n^* \psi_n < \infty$, this does not imply that $\hat{f}|\psi\rangle \in \mathcal{H}$, as is clear because the coefficients $\phi_n = \sum_m \hat{f}_{n,m} \psi_m$ may not be square summable: $\sum_n \phi_n^* \phi_n =$

divergent, i.e., we may have $|\phi\rangle \notin \mathcal{H}$.

For example, consider the matrix:

$$\hat{f}_{n,m} = \begin{pmatrix} 1^1 & 0 & 0 & 0 & & \\ 0 & 2^2 & 0 & 0 & & \\ 0 & 0 & 3^3 & 0 & & \\ 0 & 0 & 0 & 4^4 & & \\ & & & & \ddots & \\ & & & & & \ddots \end{pmatrix}_{n,m} \quad (3.92)$$

While the column vector

$$\psi_n = \begin{pmatrix} 1/1 \\ 1/2 \\ 1/3 \\ 1/4 \\ \vdots \end{pmatrix}_n \quad (3.93)$$

is square summable and therefore defines a $|\psi\rangle \in \mathcal{H}$, the action of the matrix $(\hat{f})_{n,m}$ on this column vector would yield a column vector with components $\phi_n = n^{n-1}$, which is clearly not square summable and therefore does not correspond to a vector in the Hilbert space.

Definition: We call the set $D_f \subset \mathcal{H}$ of vectors on which \hat{f} is allowed to act the “domain of \hat{f} ”.

Note that we can allow the matrix $(\hat{f})_{n,m}$ of Eq.3.92 to act on all column vectors which possess only finitely many nonzero components, i.e., all the corresponding states are in the domain of the linear map $\hat{f} = \sum_{n,m} \hat{f}_{n,m} |b_n\rangle \langle b_m|$ that is defined by the matrix elements of Eq.3.92.

Questions regarding the domains of operators are very important, for example, for the uncertainty principle, as we will see. For now, we will postpone further discussion of the domain of operators until we arrive at the notion of self-adjoint operators.

3.7.5 Changes of basis

Using the Dirac bra-ket notation, we can now rewrite equations such as Eq.3.47 in a much more compact form:

$$\bar{f}(t) = \langle \psi | \hat{f}(t) | \psi \rangle \quad (3.94)$$

We can easily recover Eq.3.47 from Eq.3.94, simply by twice inserting the resolution of the identity:

$$\bar{f}(t) = \langle \psi | \hat{f}(t) | \psi \rangle \quad (3.95)$$

$$= \langle \psi | \mathbf{1} \hat{f}(t) \mathbf{1} | \psi \rangle \quad (3.96)$$

$$= \sum_n \sum_m \langle \psi | b_n \rangle \langle b_n | \hat{f} | b_m \rangle \langle b_m | \psi \rangle \quad (3.97)$$

$$= \sum_{n,m} \psi_n^* \hat{f}_{n,m} \psi_m \quad (3.98)$$

The beauty of Dirac's notation is that it does not refer to any particular choice of basis. If we do want to work in a particular basis we simply suitably insert resolutions of the identity in terms of the basis vectors. If we wish to change basis then we can implement this through a linear map

$$\hat{U} = \sum_{r,s} |b_r\rangle U_{r,s} \langle b_s| \quad (3.99)$$

which maps the old basis vectors $|b_n\rangle$ into new basis vectors $|c_n\rangle$

$$\hat{U} : |b_n\rangle \rightarrow |c_n\rangle \quad (3.100)$$

where

$$|c_n\rangle = \hat{U}|b_n\rangle = \sum_{r,s} |b_r\rangle U_{r,s} \langle b_s|b_n\rangle = \sum_r \hat{U}_{r,n}|b_r\rangle \quad (3.101)$$

Exercise 3.18 Show that if the new vectors $|c_n\rangle$ are to form an orthonormal basis then \hat{U} has to be unitary, i.e., that $\hat{U}^\dagger = \hat{U}^{-1}$.

The power of Dirac's notation is that it clearly shows that all the physical predictions $\bar{f}(t) = \langle \psi | \hat{f} | \psi \rangle$ only depend on $\hat{f}(t)$, which is a linear map, and on the choice of state vector $|\psi\rangle$. The predictions do not depend on the basis which we choose in the vector space: Crucially, the physically relevant numbers $\bar{f}(t)$ do not depend on the choice of basis, while the coefficients ψ_n of $|\psi\rangle$ as a column vector and the matrix elements $\hat{f}_{r,s}$ of $\hat{f}(t)$ as a matrix strongly depend on the choice of basis.

Recall, in particular, that in solving the equations of motion we encountered the arbitrary constant L . In fact, the very choice of matrices a and a^\dagger was merely convenient, but not unique. In general, there are infinitely many solutions to the quantum mechanical equations of motion, canonical commutation relations and hermiticity conditions in terms of matrix-valued functions. As we mentioned already, the Stone von Neumann theorem assures us that all those different matrix-valued function solutions to the quantum mechanical problem merely differ by a choice of basis. All physical predictions $\bar{f}(t)$ are obtained basis independently and are, therefore, identical.

Exercise 3.19 Assume that b, b^\dagger are linear maps on a Hilbert space and are obeying $[b, b^\dagger] = \mathbf{1}$, where $\mathbf{1}$ is the identity map. Assume that there is a normalized⁹ vector, which we denote by $|0\rangle$, which obeys $b|0\rangle = 0$. **a)** Calculate the norm of the vector

⁹Notice that $|0\rangle$ is normalized, i.e., that it is of length one. This means that, in spite of its misleading (but usual) name, $|0\rangle$ is certainly not the zero-length vector of the Hilbert space.

$(b^\dagger)^n|0\rangle$. **b)** Further, show that the vector $|z\rangle$ which is defined through $|z\rangle = e^{zb^\dagger}|0\rangle$ is an eigenvector of b if z is any complex number. These vectors are related to so-called coherent states which are of practical importance, for example, in quantum optics (light is often found in a similar quantum state). These states are also of importance regarding the problem of “decoherence” in quantum measurement theory, as we will discuss later. **c)** Calculate $\langle z_1|z_2\rangle$.

Chapter 4

Uncertainty principles

4.1 The Heisenberg uncertainty relations

To solve a quantum mechanical problem is to choose or determine the initial state vector $|\psi\rangle$ and to calculate the position and momentum operators $\hat{x}(t), \hat{p}(t)$ subject to the equations of motion, the canonical commutation relations and subject to the hermiticity conditions $\hat{x}(t) = \hat{x}(t)^\dagger$ and $\hat{p}(t) = \hat{p}(t)^\dagger$.

Once the $\hat{x}(t), \hat{p}(t)$ (usually with indices indicating dimensions and particle number) and the initial state vector $|\psi\rangle$ of the system are known, we can calculate everything: Consider any observable $\hat{f}(t)$, i.e., any polynomials or well-behaved power series in the positions and momenta obeying $\hat{f}^\dagger(t) = \hat{f}(t)$. Then we can calculate its ensemble expectation values $\bar{f}(t) = \langle \psi | \hat{f}(t) | \psi \rangle$. Here, the term ensemble expectation value means the average outcome of the measurement of the observable $\hat{f}(t)$ if the same experiment (i.e., with the same initial conditions) is either repeated many times, or is performed many times simultaneously. An ensemble of experimental runs can be an ensemble in time or an ensemble in space.

Now if $\hat{f}(t)$ is an observable, i.e., if it obeys $\hat{f}(t) = \hat{f}(t)^\dagger$ then also $\hat{f}(t)^2$ is an observable, i.e., it obeys $\hat{f}(t)^2 = (\hat{f}(t)^2)^\dagger$. Thus, for any observable $\hat{f}(t)$ we can predict not only its ensemble average $\bar{f}(t)$, but we can also predict the ensemble average value $\overline{f^2}(t)$ of the square of that observable.

In classical mechanics, this would not be surprising. For example, when the initial conditions for throwing an object (such as a ball or a particle) are known then the distance that the object will fly (say in vacuum) before hitting the ground is an observable that can be predicted. Let's say the prediction is $50m$. Well then clearly we can also predict the square of that distance: It will be $2500m^2$. Since in classical mechanics we are sure about the distance we are also sure about the square of the distance.

In quantum mechanics, we saw that the prediction for the squares of the distance measurements is generally not the square of the prediction for the distance measurements! Knowing the average outcome of the measurement of an observable does not

automatically tell us the average outcome of the square of that observable: In general (i.e., in most cases), we have $\overline{f^2} \neq \overline{f}^2$. This means that the predictions cannot be certain, there must be a spread of the measurement outcomes. For example, the measurement outcomes for the distance might be 49m, 50m and 51m with probabilities 1/3 each. Then the prediction is 50m but the predicted square of the distances is then $(49^2 + 50^2 + 51^2)/3 \text{ m}^2 = (2500 + 2/3)m^2$.

We conclude that quantum mechanics implies that when performing an ensemble of identical experiments, each time measuring \hat{f} at the end, then the measurement values for f must have a spread. Only this way can we have that $\overline{f^2} \neq \overline{f}^2$. The predictions of quantum mechanics are generally probabilistic. We can only predict the outcome of a measurement with certainty if the observable in question happens to obey $\overline{f^2} = \overline{f}^2$.

The extent to which the equation $\overline{f^2} = \overline{f}^2$ is violated quantifies how large the spread of the outcomes of the \hat{f} measurements will be in an ensemble of experimental runs.

In fact the difference between the left hand side and the right hand side of that equation coincides with the variance of the measurement outcomes. Remember that the variance of a statistical distribution is the average of the squared deviation from the average. Here, in Dirac notation:

$$(\Delta f(t))^2 = \overline{(\hat{f}(t) - \overline{f}(t))^2} \quad (4.1)$$

$$= \langle \psi | \left(\hat{f}(t) - \langle \psi | \hat{f}(t) | \psi \rangle \right)^2 | \psi \rangle \quad (4.2)$$

$$= \langle \psi | \hat{f}^2(t) | \psi \rangle - \langle \psi | \hat{f}(t) | \psi \rangle^2 \quad (4.3)$$

$$= \overline{f(t)^2} - \overline{f(t)}^2 \quad (4.4)$$

We can now derive Heisenberg's famous uncertainty relations for the variances and the square roots of the variances (the so called standard deviation) between any pairs of observables:

Proposition: Assume $\hat{f}(t)$ and $\hat{g}(t)$ are observables and assume that the system is in the state $|\psi\rangle$. Then:

$$\Delta f(t) \Delta g(t) \geq \frac{1}{2} |\langle \psi | [\hat{f}(t), \hat{g}(t)] | \psi \rangle| \quad (4.5)$$

In particular, we have for all states $|\psi\rangle$:

$$\Delta x(t) \Delta p(t) \geq \frac{1}{2} |\langle \psi | i\hbar | \psi \rangle| = \frac{\hbar}{2} \quad (4.6)$$

In this way, the noncommutativity of the positions and momenta directly imply that in a state in which the position $\bar{x}(t)$ is predicted sharply, i.e., with small standard deviation $\Delta x(t)$, the prediction $\bar{p}(t)$ of the momentum must come with a correspondingly large standard deviation $\Delta p(t)$, and vice versa. In general, we have of course:

$$\Delta x_i^{(r)}(t) \Delta p_j^{(s)}(t) \geq \frac{\hbar}{2} \delta_{i,j} \delta_{r,s} \quad (4.7)$$

Recall that initial conditions can be posed by specifying the mean values $\bar{f}(t_0), \bar{g}(t_0)$ etc. of all observables $\hat{f}(t_0), \hat{g}(t_0)$ including their arbitrary powers. We asked under which conditions so-posed initial conditions determine a state $|\psi\rangle$. We see now that the Heisenberg uncertainty relations impose restrictions on which so-described initial conditions can occur.

Proof of the Heisenberg uncertainty principle: Assume $|\psi\rangle$ is normalized and assume that \hat{f} and \hat{g} are observables at some time t . (To keep the formulas from getting too long we will simply write \hat{g}, \hat{f} instead of $\hat{g}(t), \hat{f}(t)$). We start by considering the vector

$$|\phi\rangle = \left(\hat{f} - \bar{f}\mathbf{1} + i\alpha(\hat{g} - \bar{g}\mathbf{1}) \right) |\psi\rangle \quad (4.8)$$

where α is an arbitrary real number. No vector's norm is negative. In particular, $\langle\phi|\phi\rangle \geq 0$, i.e.:

$$\langle\psi| \left((\hat{f} - \bar{f}\mathbf{1}) - i\alpha(\hat{g} - \bar{g}\mathbf{1}) \right) \left((\hat{f} - \bar{f}\mathbf{1}) + i\alpha(\hat{g} - \bar{g}\mathbf{1}) \right) |\psi\rangle \geq 0 \quad (4.9)$$

Thus:

$$\langle\psi|(\hat{f} - \bar{f}\mathbf{1})^2|\psi\rangle + \alpha^2\langle\psi|(\hat{g} - \bar{g}\mathbf{1})^2|\psi\rangle + \alpha\langle\psi|i(\hat{f}\hat{g} - \hat{g}\hat{f})|\psi\rangle \geq 0 \quad (4.10)$$

Therefore:

$$(\Delta f)^2 + \alpha^2(\Delta g)^2 + \alpha\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle \geq 0 \quad (4.11)$$

Thus, completing the squares for α :

$$(\Delta f)^2 + (\Delta g)^2 \left(\alpha + \frac{\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle}{2(\Delta g)^2} \right)^2 - \frac{\left(\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle \right)^2}{(2(\Delta g)^2)^2} (\Delta g)^2 \geq 0 \quad (4.12)$$

We note that $\left(i[\hat{f}, \hat{g}] \right)^\dagger = i[\hat{f}, \hat{g}]$, which implies that $\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle$ is a real number. We observe that if we were to choose α very large, then the big bracket is large and the inequality is trivially obeyed. Conversely, for any given $|\psi\rangle$, we obtain the most stringent inequality for the standard deviations by choosing α such that the big bracket vanishes, i.e., if we choose $\alpha = -\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle / (2(\Delta g)^2)$. We obtain:

$$(\Delta f)^2(\Delta g)^2 - \frac{\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle^2}{4} \geq 0 \quad (4.13)$$

and therefore, finally:

$$\Delta f(t) \Delta g(t) \geq \frac{1}{2} |\langle\psi|[\hat{f}(t), \hat{g}(t)]|\psi\rangle| \quad (4.14)$$

Exercise 4.1 *There are indications from studies of quantum gravity¹, that the uncertainty relation between positions and momenta acquire corrections due to gravity effects and should be of the form: $\Delta x \Delta p \geq \frac{\hbar}{2}(1 + \beta(\Delta p)^2 + \dots)$, where β is expected to be a small positive number. Show that this type of uncertainty relation arises if the canonical commutation relation is modified to read $[\hat{x}, \hat{p}] = i\hbar(1 + \beta\hat{p}^2)$. Sketch the modified uncertainty relation $\Delta x \Delta p \geq \frac{\hbar}{2}(1 + \beta(\Delta p)^2)$ in the Δp versus Δx plane. Show that this resulting uncertainty relation implies that the uncertainty in position can never be smaller than $\Delta x_{min} = \hbar\sqrt{\beta}$.*

Technical remark: In this case, the position observables \hat{x} cannot possess eigenvectors nor close approximations to eigenvectors, because they would have vanishing position uncertainty. Such \hat{x} therefore cannot be diagonalizable and therefore, by the so-called spectral theorem, they cannot be self-adjoint (i.e., the domains of x and x^\dagger do not coincide). Such position operators are what is called symmetric operators, i.e., they obey only the bare minimum condition on an observable namely that all its expectation values are real: \hat{f} is called symmetric iff $\langle \psi | \hat{f} | \psi \rangle \in \mathbb{R} \forall |\psi\rangle \in D_{\hat{f}}$. In linear algebra, i.e., when the Hilbert spaces are finite dimensional, the notions of self-adjoint operator and symmetric operator coincide. In infinite-dimensional Hilbert spaces, all self-adjoint operators are also symmetric but not all symmetric operators are self-adjoint. In the international mathematical literature, the definitions of self-adjoint operator and symmetric operator are generally agreed upon. Unfortunately, however, there is no agreement on the definition of the term hermitean operator, which can mean either self-adjoint or symmetric operator, depending on the author. In the physics literature, the term hermitean is often used but its definition is rarely specified. Here, we will use the hermitean operator as synonymous to symmetric operator. The term describes the bare minimum requirement on any observable: Its expectation values must be real. We will write $\hat{f} = \hat{f}^\dagger$ with the tacit understanding that the domains of \hat{f} and \hat{f}^\dagger may not coincide.

¹Gravity comes in this way: Momentum, just like energy, gravitates by curving space. Assume that we want to measure positions very precisely, i.e., we try to make Δx very small. This, however, leads to a large momentum uncertainty Δp and therefore ultimately to a large uncertainty in the curvature of space at the location where we try to resolve the position. It can be shown that if Δx were about $10^{-35}m$ or smaller, the resulting curvature uncertainty would significantly disturb the precision of the position predictions. Thus, in simple models, a finite lower bound $\approx 10^{-35}m$ to the uncertainty in positions is expected. The critical length $10^{-35}m$ is called the Planck length. In my thesis, I happened to find the first Hilbert space representations for such minimal-length uncertainty relations. This then led to a lot of follow-up papers that calculated the impact on atoms, black holes and in cosmology.

4.2 The impact of quantum uncertainty on the dynamics

The fact that, in general, $(\bar{f})^n \neq \overline{f^n}$ is also important for the dynamics of quantum systems. Namely, notice that, in spite of all the subtleties of quantum mechanics, the mean values $\bar{x}_i^{(r)}, \bar{p}_j^{(s)}$ of the positions and momenta sometimes obey exactly the same equations of motion as the those of Newton. Consider, for example, the harmonic oscillator $\hat{H} = \frac{\hat{p}^2}{2m} + \frac{k}{2}\hat{x}^2$. Using the Heisenberg equations, we obtain

$$\frac{d\hat{x}}{dt} = \frac{\hat{p}}{m} \quad \text{and} \quad \frac{d\hat{p}}{dt} = -k\hat{x} \quad (4.15)$$

and, therefore:

$$\langle \psi | \frac{d\hat{x}}{dt} | \psi \rangle = \langle \psi | \frac{\hat{p}}{m} | \psi \rangle \quad \text{and} \quad \langle \psi | \frac{d\hat{p}}{dt} | \psi \rangle = -k \langle \psi | \hat{x} | \psi \rangle \quad (4.16)$$

Since the state $|\psi\rangle$ does not depend on time, we obtain these Ehrenfest equations:

$$\frac{d\bar{x}}{dt} = \frac{\bar{p}}{m} \quad \text{and} \quad \frac{d\bar{p}}{dt} = -k\bar{x} \quad (4.17)$$

Thus, the mean values $\bar{x}(t), \bar{p}(t)$ obey exactly Newton's equations. Without even solving the canonical commutation relations we can quickly solve Eqs.4.17 to obtain that $\bar{x}(t)$ and $\bar{p}(t)$ oscillate with frequency $\omega = \frac{\sqrt{k}}{2\pi\sqrt{m}}$.

Do the mean values always obey Newton's equations? No! Consider, for example, the non-harmonic oscillator $\hat{H} = \frac{\hat{p}^2}{2m} + \alpha\hat{x}^4$, which leads to

$$\frac{d\hat{x}}{dt} = \frac{\hat{p}}{m} \quad \text{and} \quad \frac{d\hat{p}}{dt} = -4\alpha\hat{x}^3 \quad (4.18)$$

and, therefore:

$$\frac{d}{dt} \langle \psi | \hat{x} | \psi \rangle = \langle \psi | \frac{\hat{p}}{m} | \psi \rangle \quad \text{and} \quad \frac{d}{dt} \langle \psi | \hat{p} | \psi \rangle = -4\alpha \langle \psi | \hat{x}^3 | \psi \rangle \quad (4.19)$$

We obtain these equations for the expectation values:

$$\frac{d\bar{x}}{dt} = \frac{\bar{p}}{m} \quad \text{and} \quad \frac{d\bar{p}}{dt} = -4\alpha\bar{x}^3 \quad (4.20)$$

We remember now that, in general, $\langle \psi | \hat{x}^3 | \psi \rangle \neq \langle \psi | \hat{x} | \psi \rangle^3$, i.e., $\overline{x^3} \neq \bar{x}^3$. Therefore, the equations Eqs.4.20 do not match Newton's equations, which would be: $\frac{d\bar{x}}{dt} = \frac{\bar{p}}{m}$ and $\frac{d\bar{p}}{dt} = -4\alpha\bar{x}^3$.

We now observe that the equations for the mean values $\bar{x}_i^{(r)}(t)$ and $\bar{p}_j^{(s)}(t)$ of positions

and momenta obey Newton's equation if and only if the Hamiltonian is a polynomial of degree at most two in the positions and momenta.

The proof is simple. The Hamilton equations hold true also in quantum mechanics:

$$\frac{d\hat{x}_i^{(r)}(t)}{dt} = \{\hat{x}_i^{(r)}(t), \hat{H}\} = \hat{f}(\hat{x}, \hat{p}) \quad \text{and} \quad \frac{d\hat{p}_i^{(r)}(t)}{dt} = \{\hat{p}_i^{(r)}(t), \hat{H}\} = \hat{g}(\hat{x}, \hat{p}) \quad (4.21)$$

Here, \hat{x}, \hat{p} stand for all position and momentum variables. If \hat{H} is a polynomial of degree ≤ 2 in the positions and momenta, then the right hand sides $\hat{f}(\hat{x}, \hat{p}), \hat{g}(\hat{x}, \hat{p})$ of these equations are linear functions in the positions and momenta. This implies that $\overline{f(\hat{x}, \hat{p})} = f(\bar{x}, \bar{p})$ and $\overline{g(\hat{x}, \hat{p})} = g(\bar{x}, \bar{p})$, so that we obtain that the expectation values obey equations whose form is identical to Newton's equations:

$$\frac{d\bar{x}_i^{(r)}}{dt} = f(\bar{x}, \bar{p}) \quad \text{and} \quad \frac{d\bar{p}_j^{(s)}}{dt} = g(\bar{x}, \bar{p}) \quad (4.22)$$

Remark: Examples of such systems are free particles, particles in harmonic oscillator potentials as well as particles exposed to constant magnetic fields.

Remark: For Hamiltonians of the form $\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x})$ where V is not a polynomial of degree ≤ 2 , the dynamics of the mean values of positions and momenta is generally quite different from Newton's dynamics. However, if the particle is well-localized, then the variance $(\Delta x)^2$ is small, i.e., we have $(\Delta x)^2 = \overline{x^2} - \bar{x}^2 \approx 0$ and, more generally: $\overline{x^n} \approx \bar{x}^n$. We conclude that as long as such a particle is well localized, its position and momentum expectation values $\bar{x}(t)$ and $\bar{p}(t)$ approximately obey the Newton equations.

4.3 The time and energy uncertainty relation

We have seen that in quantum mechanics the position coordinates are observables whose uncertainties can be calculated through $\Delta x = \langle \psi | (\hat{x} - \langle \hat{x} | \psi \rangle)^2 | \psi \rangle^{1/2}$. In contrast, quantum mechanics treats the time coordinate t as a mere parameter. Therefore, if we are looking to derive an uncertainty relation that involves time, we first need to clarify what we mean by an uncertainty Δt in time.

To this end, consider an observable $\hat{f}(t)$. Its expectation value changes over time: $\bar{f}(t)$. Starting from some time t , how much time, Δt do we have to wait until this expectation value has changed appreciably, i.e., until the change exceeds the quantum uncertainty? Namely, how much time does it take for the expectation value $\bar{f}(t)$ to change by one standard deviation $\Delta f(t)$? Clearly, Δt obeys:

$$\Delta t \left| \frac{d\bar{f}}{dt} \right| = \Delta f \quad (4.23)$$

Interestingly, quantum mechanics implies a stringent relation between the time Δt that it takes for **any** observable \hat{f} to change by a noticeable amount (i.e., by more than

Δf) on one hand and the uncertainty in the energy ΔH on the other. Namely, from

$$\Delta f \Delta H \geq \frac{1}{2} \left| \langle \psi | [\hat{f}, \hat{H}] | \psi \rangle \right| \quad (4.24)$$

and

$$\frac{d\hat{f}}{dt} = \frac{1}{i\hbar} [\hat{f}, \hat{H}] \quad (4.25)$$

we obtain:

$$\Delta f \Delta H \geq \frac{\hbar}{2} \left| \langle \psi | \frac{d\hat{f}}{dt} | \psi \rangle \right| = \frac{\hbar}{2} \left| \frac{d}{dt} \langle \psi | \hat{f} | \psi \rangle \right| \quad (4.26)$$

We used that the initial state vector $|\psi\rangle$ does not depend on time. We obtain:

$$\Delta f \Delta H \geq \frac{\hbar}{2} \left| \frac{d\bar{f}}{dt} \right| \quad (4.27)$$

With Eq.4.23, we finally obtain the time-energy uncertainty relation:

$$\Delta t \Delta H \geq \frac{\hbar}{2} \quad (4.28)$$

What this means is that if the energy of a system has uncertainty ΔH , defined as usual for an observable such as the Hamiltonian, then for every observable \hat{f} it takes at least an amount of time Δt that obeys Eq.4.28 for the mean value \bar{f} of the observable to change appreciably, i.e., for it to change by at least the amount Δf . If we know the energy of a system precisely, then none of its observables possesses any time variation. Consider, for example, a system in its lowest energy state, say a hydrogen atom. If we know that the hydrogen atom is in its lowest energy state and if we are sure, therefore, what its energy is, $\Delta H = 0$, then none of the observables \hat{f} of the hydrogen atom changes over time! In particular, there is no rotation of the electron about the proton in the sense that all observables' expectation values are constant. Conversely, if a system possess any observables \hat{f} which do change appreciably on a short time scale Δt , say if we have built a fast quantum computer, then the system must be in a state in which its energy is uncertain by a large amount ΔH , obeying Eq.4.28.

Exercise 4.2 *There are quantum mechanical systems that have a bounded energy spectrum $E_1 \leq E_2 \leq E_3 \leq \dots \leq E_{max}$. For systems with a bounded energy spectrum, the uncertainty in the energy cannot be larger than the spread of the spectrum, i.e., for all physical states $|\psi\rangle$ we have $\Delta E \leq E_{max} - E_{min}$. One can use small quantum systems as a clock (for example, in a quantum computer, the precessing spin of a nucleus, electron, atom or molecule). Assume that the quantum system of such a clock has a bounded energy spectrum with $E_n - E_1 = 1eV$ ($1eV=1\text{electronvolt}$). Calculate the maximally achievable accuracy for this clock. I.e., what is shortest time interval within which any observable property of the clock could change its expectation value by a standard deviation?*

Chapter 5

Pictures of the time evolution

5.1 The time-evolution operator

A system of N particles possesses $3N$ degrees of freedom in the sense that it has $3N$ pairs of position and momentum observables $\hat{x}_i^{(r)}, \hat{p}_j^{(s)}$. These obey the $6N$ equations of motion Eqs.3.32,3.32, which are in practice $6N$ coupled differential equations for $3N$ matrix-valued functions of time. Obviously it becomes very difficult to solve all those equations if N is large. Fortunately, there is a technique that allows us to avoid having to solve all that many differential equations: In fact, it suffices to solve just one differential equation of motion for just one matrix-valued function $\hat{U}(t)$!

Definition: The solution, $\hat{U}(t)$, to the equations

$$i\hbar \frac{d}{dt} \hat{U}(t) = \hat{U}(t) \hat{H}(t) \quad (5.1)$$

and

$$\hat{U}(t_0) = \mathbf{1} \quad (5.2)$$

where $\mathbf{1}$ is the identity matrix (or identity map) is called the system's time-evolution operator.

Proposition: Assume we have found matrices $\hat{x}_i^{(r)}(t_0), \hat{p}_j^{(s)}(t_0)$ which obey the canonical commutation relations and hermiticity conditions at the initial time t_0 . Then, the solutions $\hat{x}_i^{(r)}(t), \hat{p}_j^{(s)}(t)$ can easily be obtained from the time-evolution operator:

$$\hat{x}_i^{(r)}(t) = \hat{U}^\dagger(t) \hat{x}_i^{(r)}(t_0) \hat{U}(t) \quad (5.3)$$

$$\hat{p}_j^{(s)}(t) = \hat{U}^\dagger(t) \hat{p}_j^{(s)}(t_0) \hat{U}(t) \quad (5.4)$$

Proof: The proof is straightforward. For example, let us check that the $\hat{x}(t)$ defined in terms of the initial $\hat{x}(t_0)$ and the time evolution operator in equation Eq.5.3 does

obey the correct equation of motion:

$$\begin{aligned}
i\hbar \frac{d}{dt} \hat{x}(t) &= i\hbar \frac{d}{dt} \left(\hat{U}^\dagger(t) \hat{x}(t_0) \hat{U}(t) \right) \\
&= i\hbar \left(\dot{\hat{U}}^\dagger \hat{x}(t_0) \hat{U}(t) + \hat{U}^\dagger(t) \hat{x}(t_0) \dot{\hat{U}}(t) \right) \\
&= -\hat{H}(t) \hat{U}^\dagger(t) \hat{x}(t_0) \hat{U}(t) + \hat{U}^\dagger(t) \hat{x}(t_0) \hat{U}(t) \hat{H}(t) \\
&\quad \text{(we used that } i\hbar \dot{\hat{U}} = \hat{U} \hat{H} \text{ implies } -i\hbar \dot{\hat{U}}^\dagger = \hat{H} \hat{U}^\dagger) \\
&= -\hat{H}(t) \hat{x}(t) + \hat{x}(t) \hat{H}(t) \\
&= [\hat{x}(t), \hat{H}(t)]
\end{aligned} \tag{5.5}$$

The proof for $\hat{p}(t)$ is similar.

Exercise 5.1 Assume that $\hat{f}(t)$ is any observable which does not explicitly depend on time (i.e., which is a polynomial or a well-behaved power series in the position and momentum operators with constant coefficients). Show that the time evolution of any such $\hat{f}(t)$ is given by:

$$\hat{f}(t) = \hat{U}^\dagger(t) \hat{f}(t_0) \hat{U}(t) \tag{5.6}$$

In Sec.5.1.2, we will see that the time evolution also automatically conserves the hermiticity conditions and the canonical commutation relations.

Remark: Not only does the Hamiltonian determine the time-evolution operator $\hat{U}(t)$, but conversely $\hat{U}(t)$ also determines the Hamiltonian. From Eq.5.1, we obtain:

$$\hat{H}(t) = i\hbar \hat{U}^{-1}(t) \frac{d}{dt} \hat{U}(t) \tag{5.7}$$

Finally, let us remember that the Hamiltonian encodes all there is to know about a given system. Once we know the Hamiltonian, all equations of motion can be derived. Eq.5.7 shows that also the time-evolution operator encodes all information about a quantum system. This observation is the starting point in Feynman's formulation of quantum mechanics which we will discuss later in the course.

5.1.1 Calculating $\hat{U}(t)$

We are left with having to solve Eqs.5.1,5.2. For systems whose Hamiltonian \hat{H} does not depend on time we can immediately write down the solution! Namely:

$$\hat{U}(t) = e^{\frac{1}{i\hbar} \hat{H}(t-t_0)} \tag{5.8}$$

Of course, to make sense of this formula, we need to define what we mean by the exponentiation of a matrix or operator. This is easy. We exponentiate matrices the

same way that we exponentiate numbers, namely through the power series: $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. In our case:

$$\hat{U}(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{(t-t_0)}{i\hbar} \right)^n \hat{H}^n \quad (5.9)$$

Here, we also defined that the zero'th power of the matrix \hat{H} is the identity matrix: $\hat{H}^0 = \mathbf{1}$. This ensures that the initial condition Eq.5.2 holds true for the $\hat{U}(t)$ defined by Eq.5.9. Let us check that also Eq. 5.1 is obeyed:

$$i\hbar \partial_t \hat{U}(t) = \frac{i\hbar}{i\hbar} e^{\frac{1}{i\hbar} \hat{H}(t-t_0)} \hat{H} = \hat{U}(t) \hat{H} \quad (5.10)$$

In the main step, we differentiated the power series Eq.5.9 term by term.

We must ask whether there is any chance that we could actually sum up this power series for a given Hamiltonian matrix \hat{H} . The answer is yes: First of all, if we are given a Hamiltonian which happens to be a diagonal matrix then its exponentiation is easy to obtain $\hat{U}(t)$! That is because if

$$\hat{H} = \begin{pmatrix} E_1 & 0 & 0 & 0 & & \\ 0 & E_2 & 0 & 0 & & \\ 0 & 0 & E_3 & 0 & & \\ 0 & 0 & 0 & E_4 & & \\ & & & & \ddots & \end{pmatrix} \quad (5.11)$$

then

$$\hat{H}^n = \begin{pmatrix} E_1^n & 0 & 0 & 0 & & \\ 0 & E_2^n & 0 & 0 & & \\ 0 & 0 & E_3^n & 0 & & \\ 0 & 0 & 0 & E_4^n & & \\ & & & & \ddots & \end{pmatrix} \quad (5.12)$$

and, therefore:

$$\hat{U}(t) = \begin{pmatrix} e^{\frac{(t-t_0)}{i\hbar} E_1} & 0 & 0 & 0 & & \\ 0 & e^{\frac{(t-t_0)}{i\hbar} E_2} & 0 & 0 & & \\ 0 & 0 & e^{\frac{(t-t_0)}{i\hbar} E_3} & 0 & & \\ 0 & 0 & 0 & e^{\frac{(t-t_0)}{i\hbar} E_4} & & \\ & & & & \ddots & \end{pmatrix} \quad (5.13)$$

Of course, if \hat{H} is given as a non-diagonal matrix, then the calculation of its arbitrary powers to obtain $\hat{U}(t)$ may not be doable. However, as we will later see, Hamiltonians are self-adjoint operators and for those there is always a basis in the vector space in which they are in fact diagonal. The problem of finding a basis in which a time-independent Hamiltonian is diagonal is, therefore, of great practical importance and

various methods to this end have been devised.

In practice, Hamiltonians often possess an explicit time-dependence. For example, when working with nuclei, atoms or molecules in a quantum computer, the experimenter may want to be able to turn a knob to change and control the energy levels of nuclei, atoms or molecules so that, for example, the gaps between certain energy levels can be made at will to go in or out of resonance with incoming photons. To manipulate the energy levels of a nucleus, atom or molecule requires that the Hamiltonian has parameters in it that can be tuned externally. This can be achieved by applying from the outside, for example, a suitable magnetic or electric field whose strengths can be changed at will. For example, for a free charged particle exposed to a classical electromagnetic field the Hamiltonian $\hat{H} = \hat{p}^2/2m$ becomes:

$$H = \frac{1}{2m} \left(\vec{p} + \frac{e}{c} \vec{A}(\vec{x}, t) \right)^2 - e\Phi(\vec{x}, t) \quad (5.14)$$

Here, the vector potential \vec{A} and the potential Φ can be made to suitably change over time. When the application of an external electric field changes the energy levels of an atom or molecule, it is called the Starck effect. When the application of an external magnetic field changes the energy levels of an atom or molecule then it is called a Zeman effect¹. We have to keep in mind, however, that there are limitations to the validity of Eq.5.14. In particular, the electromagnetic field is itself a quantum system and therefore the \vec{A} and ϕ should obey suitable field commutation relations and be operator valued². We will later see why it is that the electromagnetic field often behaves approximately as if it were a classical field, justifying that Eq.5.14 is then a good approximation.

Given that time-dependent Hamiltonians are important, for example, for the control of quantum systems, the question must be addressed if one can give a closed formula for the time evolution operator $\hat{U}(t)$ also for systems whose Hamiltonian, $\hat{H}(t)$, is time dependent. The answer is yes, but it is complicated because the Hamiltonian $\hat{H}(t_1)$ at a time t_1 and the Hamiltonian $\hat{H}(t_2)$ at time t_2 are then generally quite different and have no reason to commute with another! The time-evolution operator is then:

$$\hat{U}(t) = T e^{\frac{1}{i\hbar} \int_{t_0}^t \hat{H}(t) dt} \quad (5.15)$$

Its simple looks are deceiving. Here, T is the so-called time-ordering operator. Applied to a product of Hamiltonians it orders them with respect to time:

$$T \left(\hat{H}(t_1) \hat{H}(t_2) \dots \hat{H}(t_n) \right) = \hat{H}(\tilde{t}_1) \hat{H}(\tilde{t}_2) \dots \hat{H}(\tilde{t}_n) \quad (5.16)$$

¹The Zeman effect is used to measure, for example, magnetic fields in far away galaxies: these fields change the absorption and transmission spectra of light that passes through gases in the galaxies.

²The full quantum theory of the electromagnetic field is called quantum electrodynamics.

The set of times $\{\tilde{t}_i\}_{i=1}^n$ is the same as the set of times $\{t_i\}_{i=1}^n$, but it is ordered: $\tilde{t}_1 \leq \tilde{t}_2 \leq \dots \leq \tilde{t}_n$. So, for example:

$$T\left(\hat{H}(3.4s)\hat{H}(4.1s)\hat{H}(2.7s)\right) = \hat{H}(2.7s)\hat{H}(3.4s)\hat{H}(4.1s). \quad (5.17)$$

The time ordering operator T is needed because $\hat{H}(t)$ and $\hat{U}(t)$ generally don't commute when $\hat{H}(t)$ depends on time. Explicitly, Eq.5.15 reads:

$$\hat{U}(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{1}{i\hbar}\right)^n \int_{t_0}^t \int_{t_0}^t \dots \int_{t_0}^t T\left(\hat{H}(t_1)\hat{H}(t_2)\dots\hat{H}(t_n)\right) dt_1 dt_2 \dots dt_n \quad (5.18)$$

Exercise 5.2 Show that Eq.5.18 solves Eq.5.1, i.e., show that, in this way, we achieve that differentiation of $\hat{U}(t)$ yields a factor of $\hat{H}(t)$ to the right of $\hat{U}(t)$.

Of course, if \hat{H} does not depend on time, we recover the formula Eq.5.9 for $\hat{U}(t)$. Notice that if \hat{H} does not depend on time the Hamiltonian commutes with $\hat{U}(t)$, because the Hamiltonian \hat{H} commutes with any power of itself (as does every matrix: $[A, A^n] = 0$).

5.1.2 Significance of $\hat{U}(t)$

The general expression for the time-evolution operator $\hat{U}(t)$ given in Eq.5.18 is of course difficult to use in practical calculations. But it can be very useful for abstract studies. For example, it can be used to show that the time-evolution operator is what is called “unitary”, i.e., that it obeys:

$$\hat{U}(t)^\dagger = \hat{U}(t)^{-1} \quad (5.19)$$

In the simpler case of Eq.5.9 this is easy to see:

$$\hat{U}(t)^\dagger = \left(\sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{(t-t_0)\hat{H}}{i\hbar} \right)^m \right)^\dagger \quad (5.20)$$

$$= \sum_{m=0}^{\infty} \frac{1}{m!} \left(-\frac{(t-t_0)\hat{H}}{i\hbar} \right)^m \quad (5.21)$$

$$= \hat{U}(t)^{-1} \quad (5.22)$$

The last step is justified because even for power series of matrices we have $e^A e^{-A} = 1$. The reason is that there is only one matrix A involved, i.e., noncommutativity does not come into play and the power series in the matrix A therefore behaves just like a power series in a number. The fact that the time evolution operator $\hat{U}(t)$ is unitary³

³If a number obeys $u^* = u^{-1}$ the number must be a phase, i.e., it must be of the form $e^{i\alpha}$ for a real α . Unitary operators behave in many ways like phases. In particular, there is always a basis in which a given unitary operator is diagonal. Then, its diagonal elements are all phases. We saw an example of this in Eq.5.13.

is in many ways of great mathematical and physical significance⁴. In particular, this fact allows us to prove:

Proposition: Assume that $\hat{x}_i^{(r)}(t_0)$ and $\hat{p}_j^{(s)}(t_0)$ obey the canonical commutation relations and the hermiticity conditions at the initial time t_0 . Then, the quantum mechanical time evolution operator $\hat{U}(t)$ yields $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ which obey the canonical commutation relations and the hermiticity conditions at all subsequent times.

Proof: In the case of the hermiticity conditions we have to show, for example, that $(\hat{x}_i^{(r)}(t))^\dagger = \hat{x}_i^{(r)}(t)$. Indeed:

$$(\hat{x}_i^{(r)}(t))^\dagger = (\hat{U}^\dagger(t)\hat{x}_i^{(r)}(t_0)\hat{U}(t))^\dagger \quad (5.23)$$

$$= \hat{U}^\dagger(t) (\hat{x}_i^{(r)}(t_0))^\dagger \hat{U}(t) \quad (5.24)$$

$$= \hat{U}^\dagger(t)\hat{x}_i^{(r)}(t_0)\hat{U}(t) \quad (5.25)$$

$$= \hat{x}_i^{(r)}(t) \quad (5.26)$$

Similarly, any \hat{f} obeying $\hat{f}(t_0) = \hat{f}^\dagger(t_0)$ will also obey $\hat{f}(t) = \hat{f}^\dagger(t)$ for all subsequent t .

Exercise 5.3 (a) Use the time evolution operator to prove that the canonical commutation relations are conserved, i.e., that, for example, $[\hat{x}(t_0), \hat{p}(t_0)] = i\hbar$ implies $[\hat{x}(t), \hat{p}(t)] = i\hbar$ for all t . (b) Consider the possibility that (due to quantum gravity effects) at some time t_0 the xp commutation relations take the form $[\hat{x}(t_0), \hat{p}(t_0)] = i\hbar(1 + \beta\hat{p}(t_0)^2)$ (where β is a small positive constant). Assume that the Hamiltonian is self-adjoint, i.e., that the time evolution operator is still unitary. Will these commutation relations be conserved under the time evolution?

Exercise 5.4 Consider a system with a Hamiltonian that has no explicit time dependence. Assume that we prepare the system in a state so that its energy at the initial time t_0 is known precisely. **a)** Show that the energy of the system will stay sharp, i.e., without uncertainty, at that value. **b)** Consider now the specific example of a harmonic oscillator system. Its positions and momenta evolve according to Eqs.4.17. Given the time-energy uncertainty relations, what more can you conclude for the time-evolution of $\bar{x}(t)$ and $\bar{p}(t)$ if the system is in a state with sharp energy?

⁴Possible exceptions to the unitarity of time evolution are being considered for the case of black holes horizons. There is a conundrum because unitary matrices are always invertible, but a fall into a black hole appears to be nonreversible. I think it could be that the resolution of the conundrum will involve time evolution that is described not by a unitary operator but by a merely isometric operator. Isometric operators preserve the hermiticity and canonical commutation relations without being invertible. In finite dimensional Hilbert spaces, isometry and unitarity are the same. Unitary and isometric operators are closely related (namely via the so-called Cayley transform) to self-adjoint and symmetric operators respectively, which we discussed before.

5.2 The pictures of time evolution

5.2.1 The Heisenberg picture

We had found that to solve a quantum mechanical problem of N particles we can solve the $6N$ equations of motion Eqs.3.32,3.32 for its $6N$ basic observables $\hat{x}_i^{(r)}(t), \hat{p}_j^{(s)}(t)$. In practice, this requires solving $6N$ coupled differential equations for infinite by infinite matrices. This method goes back all the way to when Heisenberg first discovered quantum mechanics. When working this way, we say we are working in the ‘‘Heisenberg picture’’.

In the last section we saw that the number of matrix differential equations that need to be solved in the Heisenberg picture, namely $6N$, can be reduced to a single differential equation, namely Eq.5.1 for the time-evolution operator, $\hat{U}(t)$. The time evolution of the $6N$ observables $\hat{x}_i^{(r)}(t), \hat{p}_j^{(s)}(t)$ is then immediately obtained through Eqs.5.3,5.4.

In fact, the computational effort in solving a quantum mechanical problem can be further reduced. Namely, instead of solving this one matrix differential equation, it actually suffices to solve just one vector-differential equation:

5.2.2 The Schrödinger picture

A key observation about the mathematics of quantum mechanics is that we never directly measure either the matrix elements $\hat{f}(t)_{n,m}$ of an observable $\hat{f}(t)$ nor do we ever directly measure the vector components ψ_n of the system’s state vector $|\psi\rangle$. We can only measure the scalar expectation values $\bar{f}(t) = \langle\psi|\hat{f}(t)|\psi\rangle = \sum_{n,m} \psi_n^* \hat{f}(t)_{n,m} \psi_m$. The deeper reason for this is that physical predictions cannot depend on the basis which we choose in the vector space. Of course, only scalars are basis independent, while vector and matrix components depend on the choice of basis in the vector space.

Therefore, in order to make physical predictions, our primary goal is to find the measurable functions $\bar{f}(t)$. And there is a shortcut to calculating these! To see this, first we use

$$\hat{f}(t) = \hat{U}^\dagger(t) \hat{f}(t_0) \hat{U}(t) \quad (5.27)$$

to write $\bar{f}(t)$ in the form:

$$\bar{f}(t) = \langle\psi| \left(\hat{U}^\dagger(t) \hat{f}(t_0) \hat{U}(t) \right) |\psi\rangle \quad (5.28)$$

$$= \sum_{i,j,n,m} \psi_i^* \left(\hat{U}_{i,j}^\dagger(t) \hat{f}_{j,n}(t_0) \hat{U}_{n,m}(t) \right) \psi_m \quad (5.29)$$

Since the multiplication of matrices and their action on vectors is associative, i.e., $a(bc) = (ab)c$, we can place the brackets also differently:

$$\bar{f}(t) = \left(\langle \psi | \hat{U}^\dagger(t) \right) \hat{f}(t_0) \left(\hat{U}(t) | \psi \rangle \right) \quad (5.30)$$

$$= \sum_{i,j,n,m} \left(\psi_i^* \hat{U}_{i,j}^\dagger(t) \right) \hat{f}_{j,n}(t_0) \left(\hat{U}_{n,m} \psi_m \right) \quad (5.31)$$

This suggests to define time-dependent states $|\psi(t)\rangle$:

$$|\psi(t)\rangle = \hat{U}(t) |\psi\rangle \quad (5.32)$$

These states are called ‘‘Schrödinger states’’, as opposed to the time-independent states $|\psi\rangle$ that we have dealt with so far and that are called ‘‘Heisenberg states’’. From Eq.5.2 we have that at the initial time t_0 the Schrödinger state starts out as identical to the Heisenberg state: $|\psi(t_0)\rangle = |\psi\rangle$.

Using the Schrödinger states we have:

$$\bar{f}(t) = \langle \psi(t) | \hat{f}(t_0) | \psi(t) \rangle \quad (5.33)$$

So we have now reformulated the calculation of $\bar{f}(t)$ so that we no longer need to know the time evolution of any observable $\hat{f}(t)$. It suffices to know the operators $\hat{f}(t_0)$ of an observable only at the initial time.

Now, however, we need instead to calculate the time-dependent vectors $|\psi(t)\rangle$. Have we really gained any advantage? Don’t we still first have to calculate $\hat{U}(t)$ to then obtain $|\psi(t)\rangle$ through Eq.5.32? Actually, no, there is a way to calculate $|\psi(t)\rangle$ without calculating $\hat{U}(t)$ first. To see this, let us rewrite the differential equation Eq.5.1 for $\hat{U}(t)$, using $\hat{U}^\dagger(t)\hat{U}(t) = 1$, to obtain:

$$i\hbar \frac{d}{dt} \hat{U}(t) = \hat{U}(t) \hat{H}(t) \hat{U}^\dagger(t) \hat{U}(t) \quad (5.34)$$

Applying this equation to the Heisenberg state $|\psi\rangle$ we obtain:

$$i\hbar \frac{d}{dt} \hat{U}(t) |\psi\rangle = \hat{U}(t) \hat{H}(t) \hat{U}^\dagger(t) \hat{U}(t) |\psi\rangle \quad (5.35)$$

This yields for the Schrödinger state $|\psi(t)\rangle$:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{U}(t) \hat{H}(t) \hat{U}^\dagger(t) |\psi(t)\rangle \quad (5.36)$$

This suggests to define:

$$\hat{H}_S(t) = \hat{U}(t) \hat{H}(t) \hat{U}^\dagger(t) \quad (5.37)$$

The operator $H_s(t)$ is called the ‘‘Hamilton operator in the Schrödinger picture’’. With this definition, we conclude that the time-dependent Schrödinger state $|\psi(t)\rangle$ can be calculated by solving the differential equation:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H}_S(t) |\psi(t)\rangle \quad (5.38)$$

This is the famous Schrödinger equation. It is a differential equation for a vector-valued function, namely $|\psi(t)\rangle$. It is, therefore, a good deal simpler than the differential equations for matrix-valued functions that we had dealt with so far. Choosing a basis, the Schrödinger equation reads in components:

$$i\hbar \frac{d}{dt} \psi_n(t) = \sum_m \hat{H}_S(t)_{n,m} \psi_m(t) \quad (5.39)$$

Once we obtain the Schrödinger state $|\psi(t)\rangle$, the prediction for the mean value of any observable \hat{f} follows from:

$$\bar{f}(t) = \langle \psi(t) | \hat{f}(t_0) | \psi(t) \rangle \quad (5.40)$$

$$= \sum_{n,m} \psi_n(t) \hat{f}_{n,m}(t_0) \psi_m(t) \quad (5.41)$$

Of course, in order to be able to solve the Schrödinger equation we first need know the Hamiltonian $\hat{H}_S(t)$ in the Schrödinger picture. And we found in Eq.5.37 that the Schrödinger picture Hamiltonian $\hat{H}_S(t)$ generally differs from the Heisenberg picture Hamiltonian $\hat{H}(t)$. If quantization of classical mechanics primarily yields the Heisenberg operator, then we now have to ask how we can find the Schrödinger Hamiltonian. Will we have to first calculate the unitary time evolution operator $\hat{U}(t)$ so that we can then use Eq.5.37? Having to first calculate the time evolution operator would of course defeat the purpose - because the whole point of using the Schrödinger picture is to avoid having to calculate an operator-valued function $\hat{U}(t)$ and instead only having to calculate a vector-valued function $|\psi(t)\rangle$.

Indeed, there is a direct way to obtain the Schrödinger Hamiltonian from the Heisenberg Hamiltonian: In Eq.5.37, the Heisenberg Hamiltonian is, as always, a polynomial or suitable power series of the position and momentum operators (with generally time-dependent coefficients):

$$\hat{H}_S(t) = \hat{U}(t) \hat{H}(\hat{x}(t), \hat{p}(t), t) \hat{U}^\dagger(t) \quad (5.42)$$

Since $\hat{U}^\dagger(t) \hat{U}(t) = 1$, we can also write:

$$\hat{H}_S(t) = \hat{H} \left(\hat{U}(t) \hat{x}(t) \hat{U}^\dagger(t), \hat{U}(t) \hat{p}(t) \hat{U}^\dagger(t), t \right) \quad (5.43)$$

For example, $\hat{U}(t) (\hat{p}(t)\hat{p}(t)) \hat{U}^\dagger(t) = \left(\hat{U}(t)\hat{p}(t)\hat{U}^\dagger(t) \right) \left(\hat{U}(t)\hat{p}(t)\hat{U}^\dagger(t) \right)$.

Now from Eqs.5.3,5.4 we have

$$\hat{U}(t)\hat{x}(t)\hat{U}^\dagger(t) = \hat{x}(t_0) \quad \text{and} \quad \hat{U}(t)\hat{p}(t)\hat{U}^\dagger(t) = \hat{p}(t_0) \quad (5.44)$$

so that we finally obtain:

$$\hat{H}_S(t) = \hat{H}(\hat{x}(t_0), \hat{p}(t_0), t) \quad (5.45)$$

We conclude that the Schrödinger Hamiltonian is the exact same polynomial or power series in the position and momentum operators as the Heisenberg Hamiltonian, i.e., its polynomial or power series has the same generally time-dependent coefficients. The only difference is that the position and momentum operators in the Schrödinger Hamiltonian are frozen at the initial time.

Finally, let us recall the equation of motion for the time evolution operator:

$$i\hbar \frac{d}{dt} \hat{U}(t) = \hat{U}(t) \hat{H}(t) \quad (5.46)$$

Using Eq.5.37, which is also $\hat{H}(t) = \hat{U}^\dagger(t) \hat{H}_S(t) \hat{U}(t)$, in Eq.5.46, we find that $\hat{U}(t)$ can also be calculated directly from the Schrödinger Hamiltonian, namely through:

$$i\hbar \frac{d}{dt} \hat{U}(t) = H_S(t) \hat{U}(t) \quad (5.47)$$

Exercise 5.5 Eq.5.37 shows that, in general, $\hat{H} \neq \hat{H}_S$ because in general the Heisenberg Hamiltonian does not commute with the time evolution operator. And this is because time dependent Heisenberg Hamiltonians generally don't even commute with themselves at different times. Show that if the Heisenberg Hamiltonian \hat{H} does not explicitly depend on time then it coincides with the Schrödinger Hamiltonian.

Exercise 5.6 Show that the following equation holds true in the Schrödinger picture and in the Heisenberg picture:

$$i\hbar \frac{d}{dt} \langle \psi | \hat{f} | \psi \rangle = \langle \psi | [\hat{f}, \hat{H}] | \psi \rangle \quad (5.48)$$

5.2.3 The Dirac picture

Working in the “Heisenberg picture”, we calculate predictions $\bar{f}(t)$ through:

$$\bar{f}(t) = \langle \psi | \hat{f}(t) | \psi \rangle = \langle \psi | \left(\hat{U}^\dagger(t) \hat{f}(t_0) \hat{U}(t) \right) | \psi \rangle \quad (5.49)$$

To this end, we have to solve the Heisenberg equation $i\hbar \frac{d}{dt} \hat{f}(t) = [\hat{f}(t), \hat{H}(t)]$ (at least for the basic position and momentum operators) to obtain the time-dependent

operators $\hat{f}(t)$. The state $|\psi\rangle$ is time independent. In contrast, when working in the “Schrödinger picture”, we calculate predictions $\bar{f}(t)$ through:

$$\bar{f}(t) = \langle\psi(t)|\hat{f}(t_0)|\psi(t)\rangle = \left(\langle\psi|\hat{U}^\dagger(t)\right) \hat{f}(t_0) \left(\hat{U}(t)|\psi\rangle\right) \quad (5.50)$$

Here, we have to solve the Schrödinger equation $i\hbar\frac{d}{dt}|\psi(t)\rangle = \hat{H}_S(t)|\psi(t)\rangle$, to obtain the time-dependent states $|\psi(t)\rangle$. The position and momentum operators are time independent.

One might think that it is always easiest to work in the Schrödinger picture, because in it we have to solve merely a vector differential equation rather than a matrix differential equation. Actually, under certain circumstances, it is possible to further simplify the calculation of the predictions, $\bar{f}(t)$. This is, when the system possesses a Schrödinger Hamiltonian $H_s(t)$ which consists of two parts

$$\hat{H}_S(t) = \hat{H}_S^{(e)}(t) + \hat{H}_S^{(d)}(t) \quad (5.51)$$

where $\hat{H}_S^{(e)}(t)$ is an easy-to-handle Hamiltonian (hence the superscript (e)) and $\hat{H}_S^{(d)}(t)$ is a Hamiltonian that is difficult to handle (hence the superscript (d)). For example, $\hat{H}_S^{(e)}(t)$ might be the Hamiltonian that contains only the kinetic energy terms of a bunch of particles, and $\hat{H}_S^{(d)}(t)$ could contain terms that describe complicated interactions of these particles. Or, $\hat{H}_S^{(e)}(t)$ might describe the time evolution of a gate within a quantum computer. The additional term $\hat{H}_S^{(d)}(t)$ might describe a special kind of interaction that the gate has with its environment. In such cases one would often call $\hat{H}_S^{(e)}(t)$ a “free” Hamiltonian while calling $\hat{H}_S^{(d)}(t)$ an “interaction Hamiltonian”. So if the interaction Hamiltonian $\hat{H}_S^{(d)}(t)$ were absent we could easily solve the dynamics of the system in either the Heisenberg or the Schrödinger picture. Only the presence of $\hat{H}_S^{(d)}(t)$ makes it difficult to calculate $\bar{f}(t)$.

This raises the question: Is there a way that we can make use of the fact that $\hat{H}_S^{(e)}(t)$ is easy, i.e., that we can easily obtain the time-evolution operator $\hat{U}^{(e)}(t)$ that solves:

$$i\hbar\frac{d}{dt}\hat{U}^{(e)}(t) = \hat{H}_S^{(e)}(t)\hat{U}^{(e)}(t) \quad (5.52)$$

The answer is yes: As Dirac first recognized, we can choose to work in what is called the “Dirac picture” which is also called the “interaction picture”. Just like the Heisenberg and the Schrödinger pictures, also the Dirac picture is obtained by a clever bracketing in the expression:

$$\bar{f}(t) = \langle\psi|\hat{U}^\dagger(t)\hat{f}(t_0)\hat{U}(t)|\psi\rangle \quad (5.53)$$

(For simplicity, let us continue to consider only observables $\hat{f}(t)$ which do not possess an explicit manually-introduced time dependence). Namely, to obtain the Dirac picture, the time evolution operator in this expression is first written as the product of two

evolution operators. One will be bracketed so that it goes with the state and one will be bracketed so that it goes with the observables.

The overall idea for Dirac picture is that the simple time evolution according to $\hat{H}^{(e)}(t)$ is given to the operators through an equation similar to the Heisenberg equation. The remaining, more difficult time evolution is then given to the state vectors through an equation similar to the Schrödinger equation.

We begin by defining the unitary operator $\hat{U}'(t)$ that expresses the difference between the full time evolution and the simple time evolution of only the easy-to-handle part of the Hamiltonian:

$$\hat{U}'(t) := \hat{U}^{(e)\dagger}(t)\hat{U}(t) \quad (5.54)$$

Notice that in the special case where the full Hamiltonian consists of only the easy-to-handle Hamiltonian, i.e., if $\hat{H}_S(t) = \hat{H}_S^{(e)}(t)$, then $\hat{U}(t) = \hat{U}^{(e)}(t)$ and therefore $\hat{U}'(t) = 1$.

Exercise 5.7 Show that $\hat{U}'(t)$ is unitary.

From Eq.5.54, we have:

$$\hat{U}(t) = \hat{U}^{(e)}(t)\hat{U}'(t) \quad (5.55)$$

Inserted into Eq.5.53, we obtain:

$$\bar{f}(t) = \langle \psi | \hat{U}'^\dagger(t) \hat{U}^{(e)\dagger}(t) \hat{f}(t_0) \hat{U}^{(e)}(t) \hat{U}'(t) | \psi \rangle \quad (5.56)$$

Now in order to obtain the Dirac picture, we choose to place brackets this way:

$$\bar{f}(t) = \left(\langle \psi | \hat{U}'^\dagger(t) \right) \left(\hat{U}^{(e)\dagger}(t) \hat{f}(t_0) \hat{U}^{(e)}(t) \right) \left(\hat{U}'(t) | \psi \rangle \right) \quad (5.57)$$

Accordingly, we define the operators and states in the Dirac picture through:

$$\hat{f}_D(t) = \hat{U}^{(e)\dagger}(t) \hat{f}(t_0) \hat{U}^{(e)}(t) \quad (5.58)$$

and

$$|\psi(t)\rangle_D = \hat{U}'(t) |\psi\rangle \quad (5.59)$$

In the Dirac picture, the time evolution operator $\hat{U}^{(e)}(t)$ solves Eq.5.52 and is by assumption easy to obtain. Therefore, $\hat{f}_D(t)$ is by assumption easy to obtain via Eq.5.58. Similar to the Heisenberg picture, the $\hat{f}_D(t)$ obey this equation:

$$i\hbar \frac{d}{dt} \hat{f}_D(t) = [\hat{f}_D(t), \hat{H}_D^{(e)}(t)] \quad (5.60)$$

as is easy to show. Notice that, in $H_D^{(d)}(t)$, the position and momentum operators possess the time evolution of the Dirac picture, i.e., they evolve according to Eq.5.58, as do all observables in the Dirac picture.

Now, the remaining part of the time evolution, namely the by assumption more difficult to handle part is described through the time evolution of the Dirac states. As we will now see, this time evolution is governed by an equation that is similar to a Schrödinger equation:

Proposition: The time evolution of the states $|\psi(t)\rangle_D$ in the Dirac picture obeys the equations

$$i\hbar \frac{d}{dt} |\psi(t)\rangle_D = \hat{H}_D^{(d)} |\psi(t)\rangle_D \quad (5.61)$$

$$|\psi(t_0)\rangle_D = |\psi\rangle \quad (5.62)$$

where $\hat{H}_D^{(d)}$ is given by:

$$\hat{H}_D^{(d)}(t) = \hat{U}^{(e)\dagger}(t) \hat{H}_S^{(d)}(t) \hat{U}^{(e)}(t) \quad (5.63)$$

In $\hat{H}_D^{(d)}$, (as we saw for $\hat{H}_D^{(e)}(t)$ above) the position and momentum operators evolve according to the Dirac picture, i.e., according to the time evolution generated by the easy-to-handle part of the Hamiltonian: From Eq.5.58 we see that $\hat{H}_D^{(d)}(t)$ is the same polynomial or power series in the positions and momenta as is $\hat{H}_S^{(d)}(t)$ - except that in $\hat{H}_D^{(d)}(t)$ the position and momentum operators evolve according to the easy time evolution operator $\hat{U}^{(e)}$. Since in the Dirac picture all observables evolve according the easy time evolution operator $\hat{U}^{(e)}$, we can also say that $\hat{H}_D^{(d)}(t)$ is $\hat{H}_S^{(d)}(t)$ after writing it in the Dirac picture.

Proof: We begin with:

$$i\hbar \frac{d}{dt} \hat{U}(t) = \hat{H}_S(t) \hat{U}(t) \quad (5.64)$$

Thus, with the dot denoting the time derivative:

$$\begin{aligned} i\hbar \left(\dot{\hat{U}}^{(e)} \hat{U}' + \hat{U}^{(e)} \dot{\hat{U}}' \right) &= \left(\hat{H}_S^{(e)} + H_S^{(d)} \right) \hat{U}^{(e)} \hat{U}' \\ i\hbar \dot{\hat{U}}^{(e)} \hat{U}' + i\hbar \hat{U}^{(e)} \dot{\hat{U}}' &= i\hbar \dot{\hat{U}}^{(e)} \hat{U}^{(e)\dagger} \hat{U}^{(e)} \hat{U}' + \hat{H}_S^{(d)} \hat{U}^{(e)} \hat{U}' \\ i\hbar \hat{U}^{(e)} \dot{\hat{U}}' &= \hat{H}_S^{(d)} \hat{U}^{(e)} \hat{U}' \\ i\hbar \dot{\hat{U}}' &= \hat{U}^{(e)\dagger} \hat{H}_S^{(d)} \hat{U}^{(e)} \hat{U}' \\ i\hbar \dot{\hat{U}}' &= \hat{H}_D^{(d)} \hat{U}' \\ i\hbar \dot{\hat{U}}' |\psi\rangle &= \hat{H}_D^{(d)} \hat{U}' |\psi\rangle \\ i\hbar \frac{d}{dt} |\psi(t)\rangle_D &= \hat{H}_D^{(d)} |\psi(t)\rangle_D \end{aligned}$$

To summarize, in the Dirac picture, we first split the Hamiltonian as in Eq.5.51 into an “easy” and a “difficult” Hamiltonian. We let the operators $\hat{f}_D(t)$ evolve according

to the easy Hamiltonian as given by Eqs.5.52,5.58 and we let the states $|\psi(t)\rangle_D$ evolve with respect to the difficult Hamiltonian according to Eqs.5.61,5.63. All predictions are then obtained through

$$\bar{f}(t) = {}_D\langle\psi(t)|\hat{f}_D(t)|\psi(t)\rangle_D \quad (5.65)$$

where ${}_D\langle\psi(t)|$ denotes the dual (i.e., bra-) vector to $|\psi(t)\rangle_D$.

For example, consider the Schrödinger Hamiltonian

$$\hat{H}_S(t) := \frac{\hat{p}(t_0)^2}{2m} + \frac{m\omega^2}{2} \hat{x}(t_0)^2 + e^{-\alpha t^2} \hat{p}(t_0)\hat{x}(t_0)^3\hat{p}(t_0) \quad (5.66)$$

In this case, the first two terms amount to a harmonic oscillator, which, on its own, is solvable. We could, therefore, work with these definitions for the Dirac picture:

$$\hat{H}_S^{(e)}(t) := \frac{\hat{p}(t_0)^2}{2m} + \frac{m\omega^2}{2} \hat{x}(t_0)^2 \quad (5.67)$$

$$\hat{H}_S^{(d)}(t) := e^{-\alpha t^2} \hat{p}(t_0)\hat{x}(t_0)^3\hat{p}(t_0) \quad (5.68)$$

In practice, the Dirac picture is in fact used ubiquitously. For example, when studying the interaction between an atom and the quantized electromagnetic field, the easy-to-handle part of the total Hamiltonian consists of the Hamiltonian of the atom and the Hamiltonian of the electromagnetic field. The difficult-to-handle part of the total Hamiltonian is the part that describes the interaction between the photon field and the atom.

In fact, notice that our treatment of the Heisenberg, Schrödinger and Dirac pictures of time evolution in this chapter is not restricted to quantum mechanics. We obtained those pictures of time evolution by shifting around brackets in formula for the calculation of any quantum theoretic prediction:

$$\begin{aligned} \bar{f}(t) &= \langle\psi|\hat{U}^\dagger(t)\hat{f}(t_0)\hat{U}(t)|\psi\rangle \\ &= \langle\psi|\left(\hat{U}^\dagger(t)\hat{f}(t_0)\hat{U}(t)\right)|\psi\rangle && \text{(Heisenberg picture)} \\ &= \left(\langle\psi|\hat{U}^\dagger(t)\right)\hat{f}(t_0)\left(\hat{U}(t)|\psi\rangle\right) && \text{(Schrödinger picture)} \\ &= \langle\psi|\hat{U}'^\dagger(t)\hat{U}^{(e)\dagger}(t)\hat{f}(t_0)\hat{U}^{(e)}(t)\hat{U}'(t)|\psi\rangle \\ &= \left(\langle\psi|\hat{U}'^\dagger(t)\right)\left(\hat{U}^{(e)\dagger}(t)\hat{f}(t_0)\hat{U}^{(e)}(t)\right)\left(\hat{U}'(t)|\psi\rangle\right) && \text{(Dirac picture)} \end{aligned}$$

This formalism applies, for example, unchanged also to quantum field theories. There, the Dirac picture is used, with the free evolution of all particles described by an “easy” Hamiltonian while the particle interactions are described by a difficult Hamiltonian.

Chapter 6

Measurements and state collapse

6.1 Ideal measurements

In this chapter, it will not matter which picture of the time evolution we choose, because we will not consider any dynamics. Instead, we will only consider the state and operators at a fixed point in time, t_m , at which we perform a measurement.

So let us assume that at the time t_m the state of a quantum system is given by a vector $|\psi\rangle$ and that we measure an observable \hat{f} . Let us further assume that the state $|\psi\rangle$ is an eigenvector to \hat{f} with an eigenvalue that we call $f \in \mathbb{R}$. This means that we assume that $|\psi\rangle = |f\rangle$ with:

$$\hat{f}|f\rangle = f|f\rangle \quad \text{and} \quad \langle f|f\rangle = 1 \quad (6.1)$$

We already saw that, in this case, we can predict the outcome of the measurement of the observable \hat{f} at the measurement time t_m with certainty. The predicted value is:

$$\bar{f} = \langle \psi | \hat{f} | \psi \rangle = \langle f | \hat{f} | f \rangle = \langle f | f | f \rangle = f \langle f | f \rangle = f \quad (6.2)$$

The uncertainty in this prediction is indeed vanishing:

$$(\Delta f)^2 = \langle \psi | \left(\hat{f} - \langle \psi | \hat{f} | \psi \rangle \right)^2 | \psi \rangle = \langle f | \left(\hat{f} - \langle f | \hat{f} | f \rangle \right)^2 | f \rangle = \langle f | (f - f)^2 | \psi \rangle = 0 \quad (6.3)$$

More generally now, assume that at the time t_m , i.e., when we measure the observable f , our quantum system is some general state $|\psi\rangle$ which is not necessarily an eigenstate of \hat{f} . In a single run of the experiment, the measurement of the observable f at time t_m will now give us some number f (that we cannot be sure about ahead of time). Let us further assume that our measurement apparatus is ideal in the sense that it measures the observable \hat{f} with absolute accuracy. This means that if, right away, we measure the observable \hat{f} again then the prediction for that measurement is the same value $\bar{f} = f$ again, and with vanishing uncertainty, $\Delta f = 0$.

Let us stress something important here: The outcome of an ideal measurement allows us to predict the outcome of an immediate repeat measurement. This means that the ideal measurement at time t_m changed the state of the system to a new state $|\psi_{after}\rangle$. Namely, the state $|\psi_{after}\rangle$ must be such that the outcome of an immediate re-measurement could be predicted with certainty. We say that at the time t_m of the measurement the previous state $|\psi\rangle$ of the quantum system *collapsed* into the new state $|\psi_{after}\rangle$. In spite of the term collapse, of course not only $|\psi\rangle$ but also $|\psi_{after}\rangle$ is normalized.

But which is the vector $|\psi_{after}\rangle$? Let us now show that if the measurement outcome was the number f then the new state $|\psi_{after}\rangle$ is eigenvector to \hat{f} with eigenvalue f . To see this, we recall that $|\psi_{after}\rangle$ must be such that the uncertainty in \hat{f} vanishes:

$$0 = (\Delta f)^2 \quad (6.4)$$

$$= \langle \psi_{after} | \left(\hat{f} - \langle \psi_{after} | \hat{f} | \psi_{after} \rangle \right)^2 | \psi_{after} \rangle \quad (6.5)$$

$$= \left[\langle \psi_{after} | \left(\hat{f} - \langle \psi_{after} | \hat{f} | \psi_{after} \rangle \right) \right] \left[\left(\hat{f} - \langle \psi_{after} | \hat{f} | \psi_{after} \rangle \right) | \psi_{after} \rangle \right] \quad (6.6)$$

$$= \langle \phi | \phi \rangle \quad (6.7)$$

where

$$|\phi\rangle = \left(\hat{f} - \langle \psi_{after} | \hat{f} | \psi_{after} \rangle \right) | \psi_{after} \rangle \quad (6.8)$$

Since the length of this vector vanishes, $\| |\phi\rangle \| = 0$, we can conclude that $|\phi\rangle = 0$ and therefore that:

$$\hat{f} | \psi_{after} \rangle = f | \psi_{after} \rangle \quad (6.9)$$

where $f = \langle \psi_{after} | \hat{f} | \psi_{after} \rangle$. Therefore, due to the measurement, the state $|\psi\rangle$ of the system collapses into a state $|\psi_{after}\rangle$ which is eigenstate of the measured observable with the eigenvalue being the value that was found in the measurement.

Corollary: A very important conclusion that we can draw from the above is that an ideal measurement of an observable \hat{f} can only ever produce an outcome that is among the eigenvalues of the operator f . For example, the set of eigenvalues of the quantized harmonic oscillator, which is its spectrum, is known to consist of the values $E_n = \hbar\omega(n + 1/2)$ with $n = 0, 1, 2, \dots$. An ideal measurement of the energy of the harmonic oscillator can therefore only ever find any one of these values.

Proposition (spectral theorem in finite dimensions): Assume \hat{f} is an N -dimensional matrix obeying $\hat{f}^\dagger = \hat{f}$. Then it possesses an eigenbasis of vectors $\{|\hat{f}_n\rangle\}_{n=1}^N$. While we saw earlier that quantum mechanics requires its Hilbert spaces to be finite dimensional, the finite-dimensional case is still important because it frequently applies to subsystems. In infinite dimensions, the spectral theorem is more complicated, as we will see.

Exercise 6.1 (a) From $\hat{f}^\dagger = \hat{f}$, show that the eigenvalues of any observable \hat{f} are real. (b) Assume that $|f_1\rangle$ and $|f_2\rangle$ are eigenvectors of \hat{f} with eigenvalues f_1, f_2 respectively. Show that if $f_1 \neq f_2$ then the two vectors are orthogonal, i.e., $\langle f_1|f_2\rangle = 0$.

6.2 The state collapse

Let us now assume that the measurement of \hat{f} yielded an eigenvalue f which is non-degenerate, i.e., that the corresponding eigenspace of the operator \hat{f} is one-dimensional. Our aim is to calculate the map that describes the collapse of the wave function:

$$|\psi_{after}\rangle = C(|\psi\rangle) \quad (6.10)$$

To this end, let us introduce the notion of projection operator:

Definition: An operator \hat{P} is called a projection operator, if it obeys:

$$\hat{P} = \hat{P}^\dagger \quad \text{and} \quad \hat{P}^2 = \hat{P} \quad (6.11)$$

Exercise 6.2 Since projection operators are self-adjoint, they can be diagonalized. (a) Calculate which eigenvalues a projection operator can possess. (b) Schematically, how does a projection operator look like as a matrix in its eigenbasis?

Using the projection operator onto the eigenspace with eigenvalue f , we can now write:

$$|\psi_{after}\rangle = \frac{1}{\|\hat{P}_f|\psi\rangle\|} \hat{P}_f|\psi\rangle \quad (6.12)$$

This means that $C(|\psi\rangle) = \left(\|\hat{P}_f|\psi\rangle\|\right)^{-1} \hat{P}_f|\psi\rangle$. The denominator is needed to ensure that $|\psi_{after}\rangle$ is normalized. Notice that, because the denominator contains $|\psi\rangle$, the map C is a nonlinear map. Sometimes it is possible to express \hat{P}_f directly in terms of \hat{f} :

Exercise 6.3 Consider a self-adjoint operator \hat{f} in an N -dimensional Hilbert space. Assume that its eigenvalues are f_1, \dots, f_N and that they are all different, i.e., assume that they are non-degenerate. In the eigenbasis $\{|f_n\rangle\}_{n=1}^N$ of \hat{f} , the operator \hat{f} takes the form:

$$\hat{f} = \sum_{n=1}^N f_n |f_n\rangle\langle f_n|. \quad (6.13)$$

We are interested in the projector \hat{P}_r onto the eigenvector $|f_r\rangle$, i.e., in the operator which obeys: $\hat{P}_r|f_n\rangle = \delta_{n,r}|f_n\rangle$. Show that \hat{P}_r can be expressed in terms for \hat{f} through:

$$\hat{P}_r = \prod_{j=1, \dots, N, j \neq r} \frac{\hat{f} - f_j \hat{1}}{f_r - f_j} \quad (6.14)$$

Here, $\hat{1}$ is the $N \times N$ identity matrix.

6.3 Simultaneous measurements

Partial measurements: Let us now consider the case that the measurement value $f \in \mathbb{R}$ that was measured is an eigenvalue of f which is degenerate. In this case, we conclude by the above arguments that the state collapses into a vector in the eigenspace of \hat{f} of eigenvalue f . But now that eigenspace is multi-dimensional. Which state in this multi-dimensional eigenspace does the system collapse into in this case? The answer is still that the state $|\psi\rangle$ collapses into the state:

$$|\psi_{after}\rangle = \frac{1}{\|\hat{P}_f|\psi\rangle\|} \hat{P}_f|\psi\rangle \quad (6.15)$$

Here, \hat{P}_f is the projector onto the eigenspace of \hat{f} for the eigenvalue f and the denominator is needed to make sure that $|\psi_{after}\rangle$ is normalized. The prescription Eq.6.15 expresses that the state is collapsed to the eigenspace but not within the eigenspace. This is because, by assumption, nothing has been measured that could distinguish the eigenstates within that eigenspace. The measurement has only been partial.

Complete measurements: We saw that if we measure an observable, $\hat{f}^{(1)}$, that possesses degenerate eigenvalues then the wave function will collapse onto the eigenspace of the measured value. But we could have measured more. We could have measured another observable, say $\hat{f}^{(2)}$, that further collapses the wave function, within the eigenspace. In general, we may be able to measure a whole number of observables $\hat{f}^{(i)}$ simultaneously, namely if they commute:

$$[\hat{f}^{(i)}, \hat{f}^{(j)}] = 0 \quad (6.16)$$

Intuitively, this is because then the uncertainty principle poses no constraint to having states that are simultaneously of zero uncertainty for both observables $\Delta\hat{f}^{(i)}\Delta\hat{f}^{(j)} \geq 0$. Concretely:

Proposition: Self-adjoint matrices can be simultaneously diagonalized if and only if they commute.

Proof: Assume that they can be jointly diagonalized. In this case, they are simultaneously diagonal in a basis (their joint eigenbasis) and they do, therefore, commute. This proves one direction of the proposition. To see the other direction, assume they commute. First, diagonalize $\hat{f}^{(1)}$. Pick the degenerate subspace V to an eigenvalue, say λ . I.e., we have that $\hat{f}^{(1)}|\phi\rangle = \lambda|\phi\rangle$ for all $|\phi\rangle \in V$. We now show that $\hat{f}^{(2)}$ maps this eigenspace into itself: $\hat{f}^{(2)} : V \rightarrow V$. This is important because it means that $\hat{f}^{(2)}$ is represented on V as a self-adjoint matrix. We have to show that if $|\phi\rangle \in V$ then also $\hat{f}^{(2)}|\phi\rangle \in V$, i.e., we have to show that $\hat{f}^{(1)}(\hat{f}^{(2)}|\phi\rangle) = \lambda(\hat{f}^{(2)}|\phi\rangle)$. But this is the case because $\hat{f}^{(1)}$ and $\hat{f}^{(2)}$ commute:

$$\hat{f}^{(1)}\hat{f}^{(2)}|\phi\rangle = \hat{f}^{(2)}\hat{f}^{(1)}|\phi\rangle = \hat{f}^{(2)}\lambda|\phi\rangle = \lambda(\hat{f}^{(2)}|\phi\rangle) \quad (6.17)$$

Now that we know that $\hat{f}^{(2)}$ maps V into itself and is therefore a self-adjoint matrix on V , we can choose a basis in that subspace that diagonalizes $\hat{f}^{(2)}$. If there is still a degenerate subspace, we proceed by also diagonalizing a third commuting observable $\hat{f}^{(3)}$, and so on. This completes the proof of the proposition.

State collapse: Let us now assume that we measure r commuting observables simultaneously. What is the state collapse? Each measurement contributes a projection operator $\hat{P}_{f^{(i)}}$. The state $|\psi\rangle$ then collapses into the state

$$|\psi_{after}\rangle = \frac{1}{\|\prod_{i=1}^r \hat{P}_{f^{(i)}}|\psi\rangle\|} \prod_{i=1}^r \hat{P}_{f^{(i)}}|\psi\rangle \quad (6.18)$$

Notice that the sequence of operators in the product does not matter because the projectors $\hat{P}_{f^{(i)}}$ commute. And this is because the projectors are diagonal in the same joint eigenbasis in which the operators $\hat{f}^{(i)}$ are diagonal.

Notation: After simultaneously measuring a sufficiently large set of r commuting observables $\hat{f}^{(1)}, \hat{f}^{(2)}, \dots, \hat{f}^{(r)}$ and obtaining r measurement values $f^{(1)}, f^{(2)}, \dots, f^{(r)}$, the state $|\psi_{after}\rangle$ that the system collapses into is fully characterized by the r measured eigenvalues, i.e., joint eigenspaces of all these observables are all one-dimensional. These joint eigenvectors are then commonly denoted by their joint eigenvalues:

$$|\psi_{after}\rangle = |f^{(1)}, f^{(2)}, \dots, f^{(r)}\rangle \quad (6.19)$$

Remark: In the quantum mechanics of a finite number of particles, there always exist finite maximal sets of commuting observables. When such a set of observables is measured simultaneously, any state collapses onto one-dimensional subspaces. Measuring more commuting observables would not bring more information in this case: In Eq.6.18, we would be multiplying more projectors but any additional projectors would merely act as the identity on the already one-dimensional space on which the other projectors are already projecting.

Exercise 6.4 *In quantum mechanics, in principle, one suitably-chosen observable \hat{f} always suffices to perform a complete measurement, i.e., a measurement that projects the state onto a one dimensional subspace. Write down such an observable \hat{f} in terms of its eigenvectors and eigenvalues.*

In the quantum mechanics of N particles in three dimensions, a maximal set of commuting observables is, for example, given by the set of their $3N$ momentum operators. For an electron in the Hydrogen atom (omitting the spin for now), a maximal set is also $(\hat{H}, \hat{L}^2, \hat{L}_3)$ and their joint eigenvectors are then denoted by $|f^{(1)}, f^{(2)}, f^{(3)}\rangle = |E_n, l, m\rangle$.

6.4 States versus state vectors

Comment regarding the uniqueness of $|\psi_{after}\rangle$: The vector $|\psi_{after}\rangle$, because it has to be normalized, $\langle\psi_{after}|\psi_{after}\rangle = 1$, is unique only up to a phase: If $|\psi_{after}\rangle$ is a

normalized eigenvector to the eigenvalue f , then so is $e^{i\alpha}|\psi_{after}\rangle$ for any $\alpha \in \mathbb{R}$. It does not matter which of these vectors we pick to represent the state after the measurement because these vectors all yield the same predictions for every observable, \hat{g} :

$$\bar{g} = \langle \psi_{after} | \hat{g} | \psi_{after} \rangle = \langle \psi_{after} | e^{-i\alpha} \hat{g} e^{i\alpha} | \psi_{after} \rangle \quad (6.20)$$

We say that these vectors represent the same state, i.e., a state is actually an equivalence class of normalized vectors that differ only by a phase.

Chapter 7

Quantum mechanical representation theory

We saw early on in this course that quantum mechanics requires its Hilbert spaces to be infinite dimensional. The reason was that if commutation relations of the form $[\hat{x}, \hat{p}] = i\hbar 1$ had a representation in terms of N -dimensional matrices then the trace could be taken on both sides of the equation, yielding the contradiction $0 = i\hbar N$.

Exercise 7.1 *Are there any values of β for which the commutation relation $[\hat{x}, \hat{p}] = i\hbar(1 + \beta\hat{p}^2)$ may possess a finite-dimensional Hilbert space representation?*

We will now study in more detail the properties that operators in infinite dimensional Hilbert spaces can possess. In particular, we know that every self-adjoint matrix \hat{Q} in a finite-dimensional Hilbert space can be diagonalized, i.e., that there exists a basis of eigenvectors of \hat{Q} and that these eigenvectors are in the Hilbert space. The full spectral theorem for self-adjoint operators on general Hilbert spaces shows that the situation is more subtle in infinite dimensions. Physicists speak of self-adjoint operators that possess non-normalizable eigenvectors which are not in the Hilbert space. What is the underlying mathematics?

7.1 Self-adjointness

What we have to look out for when defining self-adjointness in the infinite dimensional case is that the domain $D_{\hat{Q}}$ of an operator \hat{Q} may be somewhat smaller than entire the Hilbert space, because an operator can maximally only act on those Hilbert space vectors that it maps into the Hilbert space.

Remember that it happens easily that the domain of an operator is smaller than the Hilbert space. All it takes for a Hilbert space vector to be excluded from the domain is that the operator would map it into a non-normalizable vector and therefore out into the Hilbert space. For example, consider the vector with the coefficients

$(1, 1/2, 1/3, 1/4, \dots)$. Its coefficients are square-summable and so the vector is in the Hilbert space. But this vector is not in the domain of the matrix which is diagonal with the diagonal entries being $1, 2^5, 3^5, 4^5, \dots$. That's because the image would be the vector with the coefficients $(1, 2^4, 3^4, 4^4, \dots)$ and this vector is not normalizable and therefore not in the Hilbert space.

As a consequence, in the case of infinite-dimensional Hilbert spaces we have to carefully define the domain of all operators. Here is the definition of what we mean by the adjoint operator of an operator \hat{Q} , which is careful enough to be applicable also for infinite-dimensional Hilbert spaces:

Definition (adjoint operator): Assume that \hat{Q} is an operator on a Hilbert space \mathcal{H} with domain $D_{\hat{Q}}$, the domain being the maximal set of Hilbert space vectors that \hat{Q} maps into the Hilbert space. Then the domain of the adjoint operator, denoted \hat{Q}^\dagger is:

$$D_{\hat{Q}^\dagger} = \left\{ |\phi\rangle \in \mathcal{H} \mid \exists |\varphi\rangle \text{ so that } \langle \varphi | \psi \rangle = \langle \phi | \hat{Q} | \psi \rangle \quad \forall |\psi\rangle \in D_{\hat{Q}} \right\} \quad (7.1)$$

Then, \hat{Q}^\dagger is defined to act on its domain as:

$$\hat{Q}^\dagger |\phi\rangle = |\varphi\rangle \quad (7.2)$$

Definition (self-adjoint operator): An operator \hat{Q} is called self-adjoint if it is the same operator as its adjoint, and this includes that the requirement that their domains agree: $D_{\hat{Q}} = D_{\hat{Q}^\dagger}$.

7.2 The spectrum of an operator

Going back to the basics, we begin by reviewing the concept of eigenvector.

Definition (eigenvector): For a self-adjoint operator \hat{Q} on a Hilbert space \mathcal{H} (finite or infinite dimensional), an eigenvector is any vector $|\psi\rangle \in \mathcal{H}$ for which there is a number λ , called an eigenvalue, such that:

$$\hat{Q}|\psi\rangle = \lambda|\psi\rangle \quad (7.3)$$

The notions of eigenvector and eigenvalue are useful for infinite-dimensional Hilbert spaces too but we need the more general notions of spectrum and spectral resolution. To arrive at these notions, it will be useful to rewrite Eq.7.3 in this form:

$$(\hat{Q} - \lambda 1)|\psi\rangle = 0 \quad (7.4)$$

This equation shows us that the eigenvector $|\psi\rangle$ is a vector that the operator $(\hat{Q} - \lambda 1)$ maps into the zero-vector. As a consequence, the operator $(\hat{Q} - \lambda 1)$ maps any set of vectors that differ only by multiples of the vector $|\psi\rangle$ to the same vector. Therefore,

we can draw the important conclusion that for any eigenvalue λ , the operator $(\hat{Q} - \lambda 1)$ is not invertible¹.

This is a very useful observation because this property of eigenvalues, namely that they make $(\hat{Q} - \lambda 1)$ non-invertible, is a defining property that does not require writing down an eigenvector in the Hilbert space. Indeed, we are now ready to define the spectrum of an operator on a Hilbert space \mathcal{H} .

Definition (spectrum): For any operator \hat{Q} on a Hilbert space \mathcal{H} , any number $\lambda \in \mathbb{C}$ is said to be in the spectrum of \hat{Q} , if the operator $(\hat{Q} - \lambda 1)$ does not possess an inverse that is defined on the entire Hilbert space \mathcal{H} .

According to this definition, eigenvalues clearly are part of the spectrum because for them, as we just saw, $(\hat{Q} - \lambda 1)$ does not possess an inverse. We define:

Definition (point spectrum): The set of eigenvalues forms the subset of the spectrum that is called the point spectrum.

So what other values can there possibly be in the spectrum, besides eigenvalues? The answer is that - in infinite-dimensional Hilbert spaces - it can happen that there are values λ for which $(\hat{Q} - \lambda 1)$ does have an inverse $(\hat{Q} - \lambda 1)^{-1}$, but it cannot be defined on the entire Hilbert space \mathcal{H} , i.e., its domain is smaller than the Hilbert space. By the way, $(\hat{Q} - \lambda 1)^{-1}$ is also called the resolvent.

Definition (continuous spectrum): The set of $\lambda \in \mathbb{C}$ for which (a) the resolvent $(\hat{Q} - \lambda 1)^{-1}$ exists but (b) the domain of $(\hat{Q} - \lambda 1)^{-1}$ is smaller than \mathcal{H} and (c) no Hilbert space vector is orthogonal to the domain of the resolvent, forms the subset of the spectrum that is called the continuous spectrum.

Definition (residual spectrum): In principle, for arbitrary operators \hat{Q} on a Hilbert space, there is also the possibility that (a) $(\hat{Q} - \lambda 1)^{-1}$ exists but (b) the domain of $(\hat{Q} - \lambda 1)^{-1}$ is smaller than \mathcal{H} (c) there are Hilbert space vectors orthogonal to its domain (in this case, we say that the domain of $(\hat{Q} - \lambda 1)^{-1}$ is not dense in \mathcal{H}). Such λ for what is called the residual spectrum of \hat{Q} .

Proposition: The residual spectrum of self-adjoint and unitary operators is the empty set.

Therefore, we won't be concerned much with residual spectra.

Remark: It can be shown that the values in the continuous spectrum never arise as isolated points (unlike in the case of the point spectrum) but that they arise in continuous intervals, hence the naming.

¹Technically, the operator is not injective. Recall the relevant definitions: A map is called injective (or one-to-one) if every element of the target set is hit at most once. Injectivity is the condition needed for invertibility. A map is called surjective (or onto) if every element of the target set is hit at least once. A map is called bijective if it is both injective and surjective, i.e., if every element of the target set is hit exactly once.

It is clear that the values of the continuous spectrum are not eigenvalues because eigenvalues possess eigenvectors in the Hilbert space. Instead, we have:

Definition (approximate eigenvalues): The elements, λ of the continuous spectrum are called approximate eigenvalues.

This terminology is justified because it can be shown that:

Proposition: If λ is in the continuous spectrum, i.e., if it is an approximate eigenvalue, then:

$$\forall \epsilon > 0 \quad \exists |\psi\rangle \in \mathcal{H} \quad \text{with} \quad \|\psi\rangle\| = 1 \quad \text{so that} \quad \|(\hat{Q} - \lambda 1)|\psi\rangle\| < \epsilon \quad (7.5)$$

Therefore, there will always be Hilbert space vectors that approximate what would be eigenvectors arbitrarily closely, namely in the sense of Eq.7.5.

In quantum mechanics, there are plenty of operators which possess a continuous spectrum, such as the position and momentum operators. We see here already that these operators will not have eigenvectors for their continuous spectrum. However the fact that the continuous spectrum consists of approximate eigenvalues will translate into the important statement that their “eigenfunctions”, such as Dirac deltas, can always be approximated with square integrable functions. For example, the Dirac delta can be approximated in this way:

$$\int_a^b f(x) \delta(x) dx = \lim_{\epsilon \rightarrow 0} \int_a^b f(x) \frac{\epsilon}{\pi x^2 + \epsilon^2} dx \quad (7.6)$$

Terminology (bound and scattering states): In quantum mechanics, one often encounters operators that possess both a point spectrum and a continuous spectrum. For example, the Hamilton operators for systems such as an atom or a molecule will have both a point spectrum and a continuous spectrum. This is because these systems have bound states and scattering states. For example, the electron of a Hydrogen atom is not able to escape a nucleus if it is in a low energy eigenstate. Its wave function decays away from the nucleus and is normalizable. The energy levels of bound states belong to the point spectrum and are discrete, therefore. However, we can also shoot an electron very fast by a proton. Since we can do so with arbitrary increments of the energy we expect that this is the case of the continuous spectrum. And yes, in this case, the wave function is coming in from infinitely far, getting scattered off the proton and also spreading away infinitely far. Such wave functions are not normalizable. If they are energy eigenstates they can therefore only be approximate eigenvectors and belong to the continuous spectrum.

7.3 Stieljes Integration

In the case of self-adjoint operators on finite-dimensional Hilbert spaces, there is only a point spectrum. The spectral theorem in finite dimensions guarantees that the corresponding eigenvectors span the Hilbert space. For example, consider a self-adjoint

operator \hat{Q} in an N -dimensional Hilbert space and its eigenbasis $\{|q_n\rangle\}_{n=1}^N$ and the corresponding (point) spectrum of eigenvalues q_1, \dots, q_N . In this case, it is easy to write down a resolution of the identity by summing over the values of the spectrum:

$$1 = \sum_{n=1}^N |q_n\rangle\langle q_n| \quad (7.7)$$

Similarly, we have the spectral representation of \hat{Q} :

$$\hat{Q} = \sum_{n=1}^N q_n |q_n\rangle\langle q_n| \quad (7.8)$$

On infinite-dimensional Hilbert spaces, however, the spectra will generally contain both continuous and discrete parts. Resolutions of the identity and spectral representations of operators should therefore contain also integrals. In practice, most physicists simply write down sums and integrals as needed with a tacit understanding that there may be subtleties.

But there is a mathematically rigorous way to treat these sums and integrals in a unified way, namely through the notion of Stieltjes integral. We will here not treat all the details of how to Stieltjes integrate over Hilbert space operators but let us have a look at the key ideas.

Definition (Stieltjes integral): The Stieltjes integral

$$\int_a^b f(x) dm(x) \quad (7.9)$$

requires an integrand, i.e., a function $f(x)$ to be integrated over and a function $m(x)$, called the integrator (or measure). The integration is performed by considering partitions $x_0 < x_1 < x_2 < \dots < x_n$ of the interval $[a, b]$, i.e., $x_0 = a$ and $x_n = b$. We then define

$$\int_a^b f(x) dm(x) = \lim_{\epsilon \rightarrow 0} \sum_i f(\tilde{x}_i) (m(x_{i+1}) - m(x_i)) \quad (7.10)$$

Here, $f(\tilde{x}_i) \in [x_i, x_{i+1}]$ and the limit is a limit in which the maximum spacing, ϵ occurring in a sequence of partitionings goes to zero. How does this definition relate to ordinary Riemann integration?

Proposition: Assume that the derivative $m'(x)$ of $m(x)$ is continuous. Then the Stieltjes integral can be expressed in terms of the ordinary Riemann integral:

$$\int_a^b f(x) dm(x) = \int_a^b f(x) \frac{dm(x)}{dx} dx \quad (7.11)$$

So in particular, if we choose the integrator function to be $m(x) = x$ then we have simply the ordinary Riemann integral over f .

But what then is the advantage of the Stieltjes integration method? The key advantage is that the integrator function m need not have a continuous derivative. For example, consider the case of m being a Heaviside step function:

$$m(x) = \theta(x) \quad (7.12)$$

A little thought shows that the Stieltjes integral in this case yields:

$$\int_a^b f(x) dm(x) = f(0) \quad (7.13)$$

Exercise 7.2 Show Eq. 7.13.

Similarly, we can generate a sum by using an integrator function with multiple steps. For example, $m(x) = \sum_i \theta(x - x_i)$ yields.

$$\int_a^b f(x) dm(x) = \sum_i f(x_i) \quad (7.14)$$

More generally, by using an integrator function that is in some parts of the real line smooth and in some parts part of the real line with steps of various sizes, one single Stieltjes integral can express arbitrarily complicated combinations of sums and Riemann integrals.

Exercise 7.3 Plot an integrator function $m(x)$ which integrates over the intervals $[3, 6]$ and $[9, 11]$ and sums over the values of the integrand at the points $x = 5$ and $x = 6$.

Notice that the use of the Stieltjes integral made it unnecessary to use the Dirac Delta. Using the Dirac Delta, we could formally write:

$$\int_a^b f(x) dm(x) = \int_a^b f(x) \frac{dm(x)}{dx} dx = \int_a^b f(x) \sum_i \delta(x - x_i) dx = \sum_i f(x_i) \quad (7.15)$$

Here, we used that the derivative of the Heaviside function is the Dirac Delta.