Advanced Quantum Theory

AMATH473/673, PHYS454

Achim Kempf

Department of Applied Mathematics
University of Waterloo
Canada

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. All structure in the universe, and ultimately therefore also we, appear to have originated in quantum fluctuations.

The aim of this course is to explain the mathematical structure underlying all quantum theories and to apply it to the relatively simple case of nonrelativistic quantum mechanics. Nonrelativistic quantum mechanics is the quantum theory that replaces Newton's mechanics. 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 prepares for a number of graduate courses, for example, the graduate course Quantum Field Theory for Cosmology (AMATH872/PHYS785) that I normally teach every other year.

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. Also, for example, heat had been understood as a form of energy. Together, the then known 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. At that time, young Max Planck, for example, was an undergraduate student at the University of Munich. One of his instructors there was Philipp von Jolly. In the spirit of the time then, von Jolly advised Planck against a career in physics. Planck stuck with it though and became one of those whose work overthrew classical physics.

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 whose motion makes them act like little antennas that emit and absorb electromagnetic waves.

This of course also means that even relatively cold objects emit and absorb electromagnetic radiation. Their heat radiation is not visible to us because it too weak and mostly of too long wavelength for our eyes to see, but the heat radiatioin even of cold objects is easily measurable with instruments.

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 heat radiation of that frequency range most readily. For that reason, for example, tea in a tea pot that is black cools down faster than tea in a white or reflecting tea pot.

Now in the late 1800s, when researchers had completed the theorioes of classical physics, they were finally ready to try to calculate the spectrum of the radiation emitted by black bodies. To everybody's surprise, the calculations, which were first performed by Rayleigh and Jeans, predicted far more emission of waves of short wavelengths (such as ultraviolet and shorter wavelengths) than what experimental measurements seemed to indicate.

Actually, this was not a small 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. Roughly speaking, this was because, according to classical physics and its equipartition theorem, as a system (such as a cup of tea) is left alone and starts to thermally equilibrate, every physical degree of freedom, such as the kinetic energy of an atom or molecule - or such as the energy of an electromagnetic wave of a given wavelength - should acquire an average energy of $kT/2$. Here, k is the Boltzmann constant and T the temperature in Kelvin. This prediction, however, was problematic because there are infinitely many wavelengths and each was supposed to acquire $kT/2$. Even if we put the hot cup of tea in a metal box of finite size that there is a limit to how large the wavelengths of electromagnetic field can be, there is no limit to how short the wavelengths can be. Hence, the prediction of classical physics that even the just the heat of a simple hot cup of tea should excite electromagnetic waves, i.e., that it should emit an infinite amount of heat radiation of arbitrarily short wavelengths.

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". Planck decided to study 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. By that time, Planck had so much experience with the calculations of 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 strange 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 discrete packets of energy, with each packet, or quantum, carrying the energy $E_q = hf$. Planck found that the value of this constant, h, must be about 6.6 10^{-34}Kg m^2/s for the prediction of the heat radiation spectrum to come out right. The reason why this cures the ultraviolet catastrophe is that it means that electromagnetic waves of very short wavelengths are energetically expensive: Unlike in classical physics where a wave can have a very small amplitude and therefore small energy, now we have that to excite a wave of frequency f, at least the amount of energy hf must be invested to create at least one quantum (i.e., photon) of that wavelength. The equipartition law of classical physics is broken thereby:

We said that the charges that make up the cup of tea have random motions and therefore act like currents in an antenna, creating electromagnetic waves. Now, if these charges, acting like antennas, want to emit a photon of frequency f, they must muster the energy hf. But the charges only possess a typical energy, due to their thermal motion, of the order of kT. The charges that make up the cup of tea are, therefore, statistically unlikely to muster enough energy to create photons whose energy hf exceeds kT. This is why the spectrum of a lit candle of a few hundred degrees peaks in the red and then falls of quickly for higher frequencies. It's also why the spectrum of sunlight, at 6000 degrees, peaks in the green and then falls of quickly towards the ultraviolet.

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 patterns[1]. The quantum hypothesis was perceived so radical, therefore, that most researchers, including Planck himself, still expected to find an explanation of the quantum hypothesis within classical physics.

1.4 Mounting evidence for the fundamental importance of h

The significance of Planck's constant was at first controversial and even Planck was hesitant to view his quantum hypothesis as anything more than a mathematical trick that would presumably find some explanation within a continuum theory eventually. Einstein, however, was prepared to take Planck's finding at face value and in 1906, Einstein used it to quantitatively explain the photoelectric effect: Light can kick electrons

[1]It 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.

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 f comes in quanta, i.e., in packets of energy hf. 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. 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. According to classical physics, therefore, matter would not be stable and we could not exist.

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" from the action principle), 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. They were looking for a new kind of physics, a 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 absorbtion and emission spectra of atoms.

This task took more than twenty years of intense experimental and theoretical research by numerous researchers. Finally, in 1925, it was Werner 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, Erwin Schrödinger found a 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 finally 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 electromagnetic 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, with its assumptions of an absolute space and time, 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 relatively new theory of electromagnetism as superior 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 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 $v_1 = 5$km/h in a train travelling at $v_2 = 100$km/h has a speed of $v_3 = v_1 + v_2 = 105$km/h relative to the ground. In fact, she does not. Her speed to the ground is $v_3 = (v_1 + v_2)/(1 + v_1 v_2/c^2) = 104.9999994...$km/h. Today, the nonadditivity of velocities is an easy-to-measure everyday phenomenon which is built into GPS systems, for example. 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, Einstein developed the theory further into general relativity, which superseded 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[2]". Dirac's analysis, in particular, led him to correctly predict surprising magnetic properties of electrons, and it led him to correctly 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.

[2]This 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.

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. In quantum mechanics, the normalization of the wave function to norm 1 means that there is exactly one particle, somewhere. In second quantization with its uncertainty principles, one could generally no longer 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 particles[3].

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 particle annihilation and creation processes, it did yield much more than one had bargained for. The reason was that, since now the particle number was no longer conserved, the time-energy uncertainty principle made it possible for 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

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

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! All of this happens under the veil of the time-energy uncertainty principle.

It turned out that the quantum field theory's predictions was in good agreement with experimental results for the very simplest interactions. It was hoped that when taking into account also the predicted virtual processes hidden under the veil of the time-energy uncertainty principle, the accuracy of the predictions would increase even further.

Not so! These calculations typically yielded divergent integrals and so, at first, one only obtained seemingly meaningless predictions of infinite numbers. It took the combined efforts of numerous scientists, such as Feynman, Tomanaga, 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 under the veil of 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 okay, as long as the predicted measured values come out right.

1.8 Renormalization

Technically, if you like to know the gist of renormalization already, renormalization consists of the following steps: First, artificially render all predictions finite, say by cutting off 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 $\epsilon \to 0$, 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 $\epsilon \to 0$ - 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.

1.9 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 $\epsilon \to 0$ 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 about $10^{-35}m$.

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.

1.10 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 many orders of magnitude away 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. See, for example, this paper: https://doi.org/10.1103/PhysRevLett.119.031301. 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/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 repaired by suitably introducing h 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 course[1].

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 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:

Differential equations
$\Downarrow$
Functions
$\Downarrow$
Variables
$\Downarrow$
Numbers

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

[1]In his paper introducing the Schrödinger equation, Schrödinger 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. But, another part of Schrödinger's intuition was right on: His intuition was that the discreteness in quantum mechanics (e.g., of the energy levels of atoms and molecules) has its mathematical origin in the usual discreteness of the resonance frequencies of wave phenomena.

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 h 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:

Hamiltonians
⇓
Differential equations
⇓
Functions
⇓
Variables
⇓
Numbers

As a very simple example, let us consider a system of two point[2] 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:

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

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

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

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:

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

Here, $\left(\vec{p}^{(1)}\right)^2 = \sum_{i=1}^{3}\left(p_i^{(1)}\right)^2$ 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.

[2] In 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:

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

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:

$$\{x_i, p_j\} = \delta_{i,j} \tag{2.6}$$
$$\{x_i, x_j\} = 0 \tag{2.7}$$
$$\{p_i, p_j\} = 0 \tag{2.8}$$

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:

$$\{f, g\} = -\{g, f\} \qquad \text{antisymmetry} \tag{2.9}$$
$$\{cf, g\} = c\,\{f, g\}, \quad \text{for any number } c \qquad \text{linearity} \tag{2.10}$$
$$\{f, g + h\} = \{f, g\} + \{f, h\} \qquad \text{addition rule} \tag{2.11}$$

[3]Remark: 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.)

$$\{f, gh\} = \{f, g\}h + g\{f, h\} \qquad \text{product rule} \tag{2.12}$$
$$0 = \{f, \{g, h\}\} + \{h, \{f, g\}\} + \{g, \{h, f\}\} \qquad \text{Jacobi id.} \tag{2.13}$$

Let us postpone the explanation for why these definitions had to be chosen in exactly this way[4]. For now, note that an immediate consequence of these rules is that the Poisson bracket of a number always vanishes:

$$\{c, f\} = 0 \qquad \text{if } c \text{ is a number} \tag{2.14}$$

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:

$$\{x_3, p_3^2\} = \{x_3, p_3\}p_3 + p_3\{x_3, p_3\} = 1p_3 + p_3 1 = 2p_3 \tag{2.15}$$

Here now is the first set of exercises. These exercises, and all exercises up until Sec.2.3.5 are to be solved using only the above axioms for the Poisson bracket. In Sec.2.3.5, we will introduce a representation of the Poisson bracket in terms of derivatives[5]. You can use this representation only for the exercises from Sec.2.3.5 onward.

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. Evaluate* $\{x_1, p_1^n\}$*.*

Exercise 2.4 *Verify Eq.2.5.*

Exercise 2.5 *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 $r \in \{1, 2, ..., n\}$. How can we evaluate the Poisson brackets of

[4] If 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.

[5] The abstract, axiomatically-defined Poisson bracket that we defined above may sometimes be a bit tedious to evaluate but it is very powerful because of its abstractness. As Dirac first showed, this abstract, axiomatically-defined Poisson bracket is the exact same in classical and quantum mechanics. As we will see in Sec.2.3.5, there is a representation of this abstract Poisson bracket in terms of derivatives, which is more convenient to work with but this representation only works in classical mechanics.

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

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

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

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

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

Exercise 2.6 *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. Give your source(s). The source is to be a textbook, and not a site such as Wikipedia. Wikipedia is anonymous and therefore often inaccurate. Write out the answer and make sure you understand the answer. Such material is considered examinable.*

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:

$$\frac{df}{dt} = \{f, H\} + \frac{\partial f}{\partial t} \tag{2.19}$$

Here, the notation $\partial f/\partial t$ denotes differentiation of f with respect to only its explicit time dependence. In the example above it is the derivative of the time dependence in the term $\cos(4t^2)$.

Eq.2.19 is a famous equation which is called the Hamilton equation. Why is it famous? 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:

$$\frac{d}{dt}\, f \;=\; \{f, H\} \tag{2.20}$$

In the remainder of this book, unless otherwise specified, we will always choose our functions f, g, h to have no explicit dependence on time, i.e., they will depend on time only implicitly, namely through the time dependence of 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:

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

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

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

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

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

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

and

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

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:

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

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

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

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

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

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

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

Exercise 2.7 *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 2.8 *Use Eq.2.13 to prove that:*

$$\frac{d}{dt}\, \{f, g\} = \{\dot{f}, g\} + \{f, \dot{g}\} \tag{2.32}$$

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. Namely, if for simplicity, we are now assuming that $f(x, p)$

is chosen not to have any explicit dependence on time, the universal Hamilton equation which expresses the differential equation of all variables f, is:

$$\dot{f} \;=\; \{f, H\} \tag{2.33}$$

Ultimately, however, one obtains this way 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. In general, for complicated systems (such as planetary systems, for example), there can be much more complicated conserved quantities than these. 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:*

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

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:

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

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 to see (using Axiom Eq.2.9):

$$\dot{H} = \{H, H\} = 0 \tag{2.36}$$

I mentioned that 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. Here is an example. Consider a 3-dimensional isotropic (i.e., rotation invariant) harmonic oscillator. Because of its symmetry under rotations, it 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 Noether theorem, of Emmi Noether. Her theorem plays a crucial role both in practical applications, and in fundamental physics[6]. We will later come back to Noether's theorem.

Exercise 2.9 *Show that if H is a polynomial in the positions and momenta with arbitrary (and possibly time-dependent) coefficients, it is true that $dH/dt = \partial H/\partial t$.*

Exercise 2.10 *Consider the system with the Hamiltonian of Eq.2.4. Show that the total momentum is conserved, i.e., that $p_i^{(1)} + p_i^{(2)}$ is conserved for all i.*

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.

[6]Mathematically, 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.

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:

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

Exercise 2.11 *Use Eq.2.37 to evaluate* $\{x^8 p^6, x^3 p^4\}$.

Exercise 2.12 *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.2.9-2.13 and Eqs.2.16-2.18 (except: no need to check the Jacobi identity as that would be straightforward but too tedious).*

Exercise 2.13 *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:

$$\frac{d}{dt} f = \{f, H\} \tag{2.38}$$

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:

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

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

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

Here, $\delta^3(x - x')$ is the three dimensional Dirac delta distribution. The second set of axioms for the Poisson bracket is unchanged, i.e., it is still given by Eqs.2.9-2.13. There is also a representation of the Poisson bracket for fields in terms of derivatives

but we won't cover it here (can you guess it?). The energy of the classical field, i.e., its Hamiltonian, is:

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

The Hamilton equation Eq.2.38 is unchanged.

Exercise 2.14 *(Bonus question) Derive the equations of motion for $\phi(x,t)$ and $\pi(x,t)$, i.e., choose $f = \phi(x,t)$ or $f = \pi(x,t)$ in Eq.2.38. Combine these two differential equations by eliminating $\pi(x,t)$ to obtain one differential equation for $\phi(x,t)$ which contains up to the second time derivatives.*

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 a small enough to preserve all the successes of classical mechanics while it must also introduce h 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, $\frac{d}{dt}\, f = \{f, H\} + h$ or $\frac{d}{dt}\, f = \{f + h^2/f, H\}$, where h 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., Exercise 2.7). 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 h 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:

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

Since measurement values cannot be just any number but always come out real, we also have the law:

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

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

Exercise 3.1 *Find and list the defining property of an involution and give your source. The source is to be a textbook, and not a site such as Wikipedia. That's because Wikipedia is anonymous and therefore often inaccurate. Write out the answer and make sure you understand the answer. Such material is considered examinable.*

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 working out at the gym. It matters in which sequence one does them:

$$\text{gym}\cdot\text{shower - shower}\cdot\text{gym} = \text{sweat} \tag{3.3}$$

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 h 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 noncommuting 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 just one little thing: allowing the symbols $\hat{x}_i^{(r)}(t)$ and $\hat{p}_i^{(r)}(t)$ to be noncommutative in a very particular way.

Before we can follow Dirac's argument, let us first reconsider the product rule for the Poisson bracket:

$$\{f, gh\} = \{f, g\}h + g\{f, h\} \tag{3.4}$$

Using the antisymmetry rule, $\{f, g\} = -\{g, f\}$, the product rule can be rewritten in this form:

$$\{gh, f\} = \{g, f\}h + g\{h, f\} \tag{3.5}$$

Using Eqs.3.4,3.5, we can now follow Dirac's argument for why the Poisson algebra structure imposes strict conditions on the form that any noncommutativity can take. Dirac considered the Poisson bracket

$$\{\hat{u}_1\hat{u}_2, \hat{v}_1\hat{v}_2\} \tag{3.6}$$

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.3.6 can be decomposed into simpler Poisson brackets in two ways, namely using first Eq.3.4 and then Eq.3.5, or vice versa. 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:

$$\begin{aligned} \{\hat{u}_1\hat{u}_2, \hat{v}_1\hat{v}_2\} &= \hat{u}_1\{\hat{u}_2, \hat{v}_1\hat{v}_2\} + \{\hat{u}_1, \hat{v}_1\hat{v}_2\}\hat{u}_2 \\ &= \hat{u}_1(\hat{v}_1\{\hat{u}_2, \hat{v}_2\} + \{\hat{u}_2, \hat{v}_1\}\hat{v}_2) + (\hat{v}_1\{\hat{u}_1, \hat{v}_2\} + \{\hat{u}_1, \hat{v}_1\}\hat{v}_2)\hat{u}_2 \end{aligned} \tag{3.7}$$

must agree with:

$$\begin{aligned} \{\hat{u}_1\hat{u}_2, \hat{v}_1\hat{v}_2\} &= \hat{v}_1\{\hat{u}_1\hat{u}_2, \hat{v}_2\} + \{\hat{u}_1\hat{u}_2, \hat{v}_1\}\hat{v}_2 \\ &= \hat{v}_1(\hat{u}_1\{\hat{u}_2, \hat{v}_2\} + \{\hat{u}_1, \hat{v}_2\}\hat{u}_2) + (\hat{u}_1\{\hat{u}_2, \hat{v}_1\} + \{\hat{u}_1, \hat{v}_1\}\hat{u}_2)\hat{v}_2 \end{aligned} \tag{3.8}$$

We can, therefore, conclude that, independently of whether or not we have commutativity, it must always be true that:

$$\{\hat{u}_1, \hat{v}_1\}(\hat{v}_2\hat{u}_2 - \hat{u}_2\hat{v}_2) = (\hat{v}_1\hat{u}_1 - \hat{u}_1\hat{v}_1)\{\hat{u}_2, \hat{v}_2\} \tag{3.9}$$

And this equation has to be true for all possible choices of $\hat{u}_1, \hat{u}_2, \hat{v}_1, \hat{v}_2$. How can we ensure this? As is easy to check, Eq.3.9 will be true[1] if we require all expressions $\hat{u}, \hat{v}$ in the position and momentum variables obey:

$$\hat{v}\hat{u} - \hat{u}\hat{v} = k\{\hat{u}, \hat{v}\} \tag{3.10}$$

[1]We will not show here that, vice versa, *only* the condition Eq.3.10 ensures that Eq.3.9 holds true. If you are interested, there is plenty of literature on the topic of "quantization".

with k being some constant that commutes with everything. This is because in this case, the left and right hand sides of Eq.3.9 are automatically identical. But what value does k take? Of course, 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 $k = -ih/2\pi$, i.e., we have

$$\hat{u}\hat{v} - \hat{v}\hat{u} = i\hbar\{\hat{u}, \hat{v}\} \tag{3.11}$$

where we used the convenient definition:

$$\hbar = \frac{h}{2\pi} \tag{3.12}$$

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

$$\hat{x}_i^{(r)}\hat{p}_j^{(s)} - \hat{p}_j^{(s)}\hat{x}_i^{(r)} = i\hbar\ \delta_{i,j}\delta_{r,s} \tag{3.13}$$
$$\hat{x}_i^{(r)}\hat{x}_j^{(s)} - \hat{x}_j^{(s)}\hat{x}_i^{(r)} = 0 \tag{3.14}$$
$$\hat{p}_i^{(r)}\hat{p}_j^{(s)} - \hat{p}_j^{(s)}\hat{p}_i^{(r)} = 0 \tag{3.15}$$

Let us keep in mind that we did not modify the rules of the Poisson bracket. We still have:

$$\{\hat{x}_i, \hat{p}_j\} = \delta_{i,j} \tag{3.16}$$
$$\{\hat{x}_i, \hat{x}_j\} = 0 \tag{3.17}$$
$$\{\hat{p}_i, \hat{p}_j\} = 0 \tag{3.18}$$

$$\{f, g\} = -\{g, f\} \quad \text{antisymmetry} \tag{3.19}$$
$$\{cf, g\} = c\,\{f, g\}, \quad \text{for any number } c \quad \text{linearity} \tag{3.20}$$
$$\{f, g + h\} = \{f, g\} + \{f, h\} \quad \text{addition rule} \tag{3.21}$$
$$\{f, gh\} = \{f, g\}h + g\{f, h\} \quad \text{product rule} \tag{3.22}$$
$$0 = \{f, \{g, h\}\} + \{h, \{f, g\}\} + \{g, \{h, f\}\} \quad \text{Jacobi id.} \tag{3.23}$$

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.3.13-3.15 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.3.14,3.15 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 noncommutative, 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:

Hamiltonian
$\Downarrow$
Equations of motion
$\Downarrow$
Differential equations
$\Downarrow$
Non-number-valued functions
$\Downarrow$
Number-valued functions

$$\Downarrow$$

Number predictions

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:

$$\frac{d}{dt}\,\hat{f} \;=\; \{\hat{f}, \hat{H}\} \tag{3.24}$$

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:

$$\frac{d}{dt}\,\hat{x}_i^{(r)} \;=\; \{\hat{x}_i^{(r)}, \hat{H}\} \tag{3.25}$$

$$\frac{d}{dt}\,\hat{p}_i^{(r)} \;=\; \{\hat{p}_i^{(r)}, \hat{H}\} \tag{3.26}$$

By evaluating the Poisson bracket on the right hand side of Eqs.3.25,3.26 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:

$$\left(\hat{x}_i^{(r)}\right)^\dagger = \hat{x}_i^{(r)} \qquad \text{and} \qquad \left(\hat{p}_j^{(s)}\right)^\dagger = \hat{p}_j^{(s)} \tag{3.27}$$

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:

$$\hat{x}_i^{(r)}\hat{p}_j^{(s)} - \hat{p}_j^{(s)}\hat{x}_i^{(r)} \;=\; i\hbar\;\delta_{i,j}\delta_{r,s} \tag{3.28}$$

$$\hat{x}_i^{(r)}\hat{x}_j^{(s)} - \hat{x}_j^{(s)}\hat{x}_i^{(r)} \;=\; 0 \tag{3.29}$$

$$\hat{p}_i^{(r)}\hat{p}_j^{(s)} - \hat{p}_j^{(s)}\hat{p}_i^{(r)} \;=\; 0 \tag{3.30}$$

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.3.11:*

$$\{\hat{u}, \hat{v}\} = \frac{1}{i\hbar}(\hat{u}\hat{v} - \hat{v}\hat{u}) \tag{3.31}$$

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

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

$$[A, B] \ := \ A\ B \ - \ B\ A \tag{3.32}$$

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.3.24 holds and takes the form:

$$i\hbar \frac{d}{dt}\ \hat{f}(t) \ = \ [\hat{f}(t), \hat{H}] \tag{3.33}$$

$$\tag{3.34}$$

and, in particular:

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

$$i\hbar \frac{d}{dt}\ \hat{p}_i^{(r)}(t) \ = \ [\hat{p}_i^{(r)}(t), \hat{H}] \tag{3.35}$$

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 that 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 approximately harmonic for small oscillations. 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 of motion 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 $\hat{H}_1 = \hat{p}^2/2m + a(\hat{x}^2\hat{p}\hat{x} - \hat{x}\hat{p}\hat{x}^2)$ and $\hat{H}_2 = \hat{p}^2/2m + b(\hat{p}\hat{x}\hat{p}^2 - \hat{p}^2\hat{x}\hat{p})$ where a and b are constants with appropriate units. Check whether or not these two Hamiltonians are the same in classical mechanics. Also find all choices of a, b for which the two Hamiltonians are the same in quantum mechanics.*

Exercise 3.5 *Find a Hamiltonian which contains at least one thousand powers of $\hat{x}$ and which also agrees with the Hamiltonian $\hat{H}_1$ of the previous exercise in classical mechanics. Make sure that your Hamiltonian is formally hermitean, i.e., that it obeys $\hat{H}^\dagger = \hat{H}$. Help: To ensure hermiticity, you can symmetrize. For example, $\hat{x}\hat{p}^2$ is not hermitean but $(\hat{x}\hat{p}^2 + \hat{p}^2\hat{x})/2$ is hermitean.*

Remark: In quantum theory, the choice of Hamiltonian always has an ordering ambiguity because one could always add to the Hamiltonian any extra terms that are proportional to $(\hat{x}\hat{p} - \hat{p}\hat{x})$ because those terms don't affect what the Hamiltonian is in classical mechanics. In principle, experiments are needed to decide which Hamiltonian is the correct one. In practice, the simplest choice is usually the correct choice. The simplest choice is obtained by symmetrizing the given classical Hamiltonian and then not adding any extra terms that are proportional to $\hat{x}\hat{p} - \hat{p}\hat{x}$. This is called the Weyl ordered 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. Now we will have to develop methods for descending in quantum mechanics from the level of the Hamiltonian down to the concrete predictions of numbers in experiments.

In the previous section, we already took the first step: we found that we can use the Hamiltonian to derive the differential equations of motion of the system. Since the Poisson brackets have not changed as we went from classical to quantum mechanics, the equations of motion are the same as those of classical mechanics[2].

The big change compared to classical mechanics, is that now the position and momentum variables $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ are noncommutative according to Eqs.3.28-3.30, and that they, therefore, can no longer be represented by number-valued functions of time. This means that the equations of motion can no longer be interpreted as differential equations for number-valued functions!

But we need to find a way to descend the ladder of abstractions all the way from the Hamiltonian on the top of the ladder down to concrete predictions of numbers at the bottom of the ladder. To that end, in order now 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 of time whose values are some kind of noncommutative mathematical objects. What kind of mathematical objects could these be?

3.4.1 Linear maps

Actually, *every* mathematical object can be viewed as a map, if need be, as a trivial map. For example the number 5 can be identified with the map that maps everything to 5. So let us look at maps. Let's try to represent the symbols $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as some kind of explicit map-valued functions of time. A simple kind of maps is the linear maps. And they can be noncommutative! So this looks promising.

For example, any square matrix that acts on a finite-dimensional vector space of column vectors represents a linear map. And square matrices are generally noncommutative! In principle, we need the matrices to be square matrices so that they map back into the same vector space, so that we can multiply any two matrices on the same vector space. As we will see later, with some precautions, we can also consider infinite-by-infinite matrices that act as linear maps on infinite-dimensional vector spaces.

[2]Except for the ordering ambiguity: when going from the classical to the quantum Hamiltonian we could (if we had any experimental reason to do so) add to the quantum Hamiltonian any hermitean terms that are proportional to $\hbar$, such as terms like $ig(\hat{x})(\hat{x}\hat{p} - \hat{p}\hat{x})\hat{g}(\hat{x})$ where g is some polynomial in $\hat{x}$.

There are actually many kinds of linear maps and they may not act on vector spaces of column vectors at all!

Let us consider, for example, the infinite-dimensional vector space $V := C^7(\mathbb{R})$ of seven times continuously differentiable functions on the real line. The set V forms a vector space because it obeys the defining axioms of a vector space: in brief, one can suitably add any two elements of V and get an element of V and one can suitably multiply any element of V with a number and get an element of V.

Exercise 3.6 *Find and list the precise axioms that a set has to obey to be called a vector space.*

Definition: *A map on a vector space that is infinite dimensional is called an operator.*

For example, the derivative operator, D, acts on functions in V in this way:

$$D : g(\lambda) \to \frac{d}{d\lambda} g(\lambda) \tag{3.36}$$

The operator D is a linear operator, i.e., it is a linear map, because it obeys $\partial_\lambda\,(c\,g(\lambda)) = c\;\partial_\lambda g(\lambda)$ for all numbers c and because $\partial_\lambda(g_1(\lambda) + g_2(\lambda)) = \partial_\lambda g_1(\lambda) + \partial_\lambda g_2(\lambda)$. Here, in order to simplify the notation, we introduced the notation: $\partial_\lambda := \frac{d}{d\lambda}$

Exercise 3.7 *Check whether or not the multiplication operator, M, which maps $M : g(\lambda) \to \lambda g(\lambda)$ is a linear operator.*

Exercise 3.8 *Show that the two operators D and M on V do not commute, namely by calculating $(DM - MD)g(\lambda)$.*

Exercise 3.9 *Check whether or not the operator Q which acts on functions in V as $Q : g(\lambda) \to \lambda^5 g(\lambda)$ is a linear operator.*

3.4.2 Choices of representation

We have just seen examples of linear maps and, since they generally do not commute, they may be useful for representing the variables $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as explicit mathematical objects. And this is what we need to be able to descend further down the ladder of abstractions, down to predictions of numbers for measurement outcomes.

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. There are articles by Steven Weinberg, for example, on this topic. This work has shown, however, that any attempt at using nonlinear spaces or nonlinear operators to define quantum theories generally leads to physically incorrect predictions. We will, therefore, here only consider linear representations.

Now that we have settled on representations of the variables $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as linear operators, we still have plenty of choice, because there are so many vector spaces and so many linear operators on them. And this leads to a worry: could it happen that we invest great effort in developing one particular kind of representation of the variables $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as operators, say as matrices, and then it turns out that we have bet on the wrong horse? Maybe, we should have instead developed a representation of the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as, for example, multiplication and differentiation operators?

Fortunately, essentially[3] all linear representations of variables $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ that obey the canonical commutation relations are equivalent, i.e., they lead to the exact same predictions! This is the content of the Stone von Neumann theorem, which we will later cover more precisely. Technically, as we will see, all linear representations are the same, up to a change of basis in the vector space. It may seem strange that, for example, a space of column vectors with countably infinitely many entries could be isomorphic to some space of functions on the real line. But this is what will turn out to be the case[4]!

So to recapitulate: our task is to solve the equations of motion, the hermiticity conditions and the canonical commutation relations for $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$ as linear-map-valued (instead of number-valued) functions of time.

We now know that the choice of which kind of linear representation we use will ultimately not matter when calculating physical predictions.

As our first choice, let us, therefore, use the most concrete kind of linear maps to represent the $\hat{x}_i^{(r)}(t)$ and $\hat{p}_j^{(s)}(t)$, namely, let us try to represent them as matrix-valued functions in time. Historically, quantum mechanics was actually first written down in terms of matrix-valued functions, back in June 1925 when the young Heisenberg had some quiet time while escaping his hay fever on the island of Helgoland in the North Sea.

3.4.3 A matrix representation

Let us now find out how the variables $\hat{x}_i^{(r)}(t)$ 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. To keep the number of indices in check, we will restrict ourselves here to the case of just one $\hat{x}(t)$ and one $\hat{p}(t)$ operator.

[3] There is a small subtlety, arising from the fact that, as we'll see later, the $\hat{x}$ and $\hat{p}$ are what are called unbounded operators. This leaves some loopholes, in principle, but nature does not appear to make use of those.

[4] If you want to know the essence already: the space of functions will be the set of equivalence classes of square-integrable functions, two functions being in the same equivalence class if their difference has vanishing integral over its norm squared. In this space of equivalence classes one can find bases of countably infinitely many basis vectors.

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:

$$\hat{x}(t_0)\ \hat{p}(t_0)\ -\hat{p}(t_0)\ \hat{x}(t_0)\ =\ i\hbar\ \mathbf{1} \tag{3.37}$$

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

$$\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B) \quad \text{and} \quad \mathrm{Tr}(AB) = \mathrm{Tr}(BA) \tag{3.38}$$

Exercise 3.10 *Verify Eqs.3.38 by writing the matrices in an orthonormal basis, i.e., with indices, and by then evaluating the trace by summing up the matrix elements on the diagonal.*

We see that the trace of the left hand side of Eq.3.37 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.3.37! 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.

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

$$a_{n,m} = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \\ 0 & 0 & \sqrt{2} & 0 & 0 & \\ 0 & 0 & 0 & \sqrt{3} & 0 & \\ 0 & 0 & 0 & 0 & \sqrt{4} & \\ 0 & 0 & 0 & 0 & 0 & \\ & & & & & \ddots \end{pmatrix}_{n,m} \tag{3.39}$$

The hermitean conjugate is:

$$a^\dagger_{n,m} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \\ \sqrt{1} & 0 & 0 & 0 & 0 & \\ 0 & \sqrt{2} & 0 & 0 & 0 & \\ 0 & 0 & \sqrt{3} & 0 & 0 & \\ 0 & 0 & 0 & \sqrt{4} & 0 & \\ & & & & & \ddots \end{pmatrix}_{n,m} \tag{3.40}$$

Their commutation commutation relation is:

$$aa^\dagger - a^\dagger a\ =\ \mathbf{1} \tag{3.41}$$

Since they are not numbers, we should decorate a and $a^\dagger$ with hats but traditionally one doesn't put hats on these operators.

Remark: *In case you are wondering because you feel that you have seen similar things before: fundamentally, these operators a and $a^\dagger$ have absolutely nothing to do with harmonic oscillators. What we are currently doing will be good for any choice of system, not just harmonic oscillators. We are currently developing a representation of the variables $\hat{x}(t)$ and $\hat{p}(t)$ as matrices and this representation will, of course, be good for any arbitrary choice of Hamiltonian*[5]*.*

Exercise 3.11 *Verify Eq.3.41.*

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:

$$\hat{x}(t_0) = L(a^\dagger + a) \tag{3.42}$$

and

$$\hat{p}(t_0) = \frac{i\hbar}{2L}(a^\dagger - a) \tag{3.43}$$

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.3.27 are obeyed, i.e., such that the matrices are formally[6] hermitean: $\hat{x}^\dagger(t_0) = \hat{x}(t_0)$ and $\hat{p}^\dagger(t_0) = \hat{p}(t_0)$.

Exercise 3.12 *Verify that the two matrices defined in Eqs.3.42,3.43 with the help of Eqs.3.39,3.40, are formally hermitean. I am using the term "formally" here to indicate that, for the purposes of this exercise, you need not worry about potential subtleties that may arise because these matrices are infinite dimensional.*

Exercise 3.13 *Show that the hermitean conjugation of matrices reverses the order, i.e., that if A and B are linear maps, then $(AB)^\dagger = B^\dagger A^\dagger$. To this end, write out the matrices with indices and use that hermitean conjugating a matrix means transposing and complex conjugating it.*

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 see, therefore, why the imaginary unit i appeared 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 $\hat{p}\hat{x} - \hat{p}\hat{x} = k^*\mathbf{1}$, i.e., we obtain $k = -k^*$. Thus, k has to be imaginary. And of course it is: $k = i\hbar$.

[5] Still, it is true also that the use of the a and $a^\dagger$ will be particularly convenient when considering the special case of harmonic oscillators.

[6] I 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.

3.4.4 Example: Solving the equations of motion for a free particle with matrix-valued functions

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:

$$\frac{d}{dt}\hat{x}(t) = \frac{1}{m}\hat{p}(t) \tag{3.44}$$

$$\frac{d}{dt}\hat{p}(t) = 0 \tag{3.45}$$

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

$$\hat{x}(t) = \hat{x}(t_0) + \frac{(t-t_0)}{m}\hat{p}(t_0) \tag{3.46}$$

and

$$\hat{p}(t) = \hat{p}(t_0), \tag{3.47}$$

where $\hat{x}(t_0)$ and $\hat{p}(t_0)$ are the matrices of Eqs.3.42,3.43. Concretely, by substituting in the matrices a and $a^\dagger$, we have:

$$\hat{x}(t)_{n,m} = \begin{pmatrix} 0 & \sqrt{1}\left(L - \frac{i\hbar(t-t_0)}{2Lm}\right) & 0 & \\ \sqrt{1}\left(L + \frac{i\hbar(t-t_0)}{2Lm}\right) & 0 & \sqrt{2}\left(L - \frac{i\hbar(t-t_0)}{2Lm}\right) & \\ 0 & \sqrt{2}\left(L + \frac{i\hbar(t-t_0)}{2Lm}\right) & 0 & \\ & & & \ddots \end{pmatrix} \tag{3.48}$$

$$\hat{p}(t)_{n,m} = \begin{pmatrix} 0 & -\sqrt{1}\frac{i\hbar}{2L} & 0 & \\ \sqrt{1}\frac{i\hbar}{2L} & 0 & -\sqrt{2}\frac{i\hbar}{2L} & \\ 0 & \sqrt{2}\frac{i\hbar}{2L} & 0 & \\ & & & \ddots \end{pmatrix} \tag{3.49}$$

For simplicity, not all the many zeros in these matrices are shown. The only nonzero terms are immediately to the left and right of the diagonal.

Exercise 3.14 *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 constructed the representation in such a way that the commutation relation and the hermiticity condition hold 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 systems[7] 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: $d/dt\ \{\hat{x}, \hat{p}\} = \{\{\hat{x}, \hat{p}\}, \hat{H}\} = \{1, \hat{H}\} = 0$. We can also turn the logic around. Assume we know nothing about Hamiltonians and about the dynamics of quantum systems. Except, we may want to assume that, whatever the time evolution is, it must preserve the Poisson algebra structure, i.e., we require that the Poisson brackets be conserved in time. The structure of the Poisson algebra then demands (we don't show this explicitly here) that the time evolution must be generated through an equation of the type of the Hamilton equation, by some generator which we may call H, and which we may then as well call the Hamiltonian.

3.4.5 Example: Solving the equations of motion for a harmonic oscillator with matrix-valued functions

The vibrational degree of freedom of a diatomic molecule such as HF, CO or HCl can be described as a harmonic oscillator (as long as the oscillations are small). Now let x stand for the deviation from the equilibrium distance between the two nuclei. This distance oscillates harmonically and is described by this effective Hamiltonian of the form of a harmonic oscillator:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2 \tag{3.50}$$

The term "Effective Hamiltonian" expresses the fact that this Hamiltonian is not really the exact Hamiltonian but that it is a good approximation to the Hamiltonian in the regime of low energies (i.e., of small oscillations) that we are considering here. By the way, how do we know that the true Hamiltonian is not simply a harmonic oscillator? Easy: we know from experiments that diatomic molecules will, for example, split apart at sufficiently high temperatures, i.e., that they do not have infinite binding energy. A harmonic oscillator potential, however, just keeps going up faster with distance and therefore if you tried to pull apart the two particles in the diatomic molecule, they would just get pulled together more and more strongly. Diatomic molecules could never be split if they were truly harmonically bound.

So then if we know that the true potential is not harmonic, how do we know that a harmonic potential is a good approximation at low energies? That's because

[7] With the possible exception of systems that involve black hole horizons or other gravitational horizons or singularities.

any potential $V(x)$ is normally well described by a smooth function $V(x)$ with a minimum at, say x_0, around which we can Taylor expand the potential: $V(x) = V(x_0) + V'(x_0)(x - x_0) + \frac{1}{2}V"(x_0)(x - x_0)^2 + ...$ Here, $V'(x_0) = 0$ because x_0 is assumed to be the minimum and therefore in the regime of small oscillations, where x is close to x_0, any potential starts out quadratic, i.e., as a harmonic oscillator.

Now let us remember that the matrix-valued functions $\hat{x}(t)$ and $\hat{p}(t)$ that we want to solve the harmonic oscillator problem Eq.3.50 have to obey three conditions, as always in quantum theory. First, the matrix-valued functions $\hat{x}(t)$ and $\hat{p}(t)$ must obey the equations of motion of the harmonic oscillator given in Eq.3.50. Second, they must be hermitean, $\hat{x}(t) = \hat{x}^\dagger(t)$ and $\hat{p}(t) = \hat{p}^\dagger(t)$. Third, they must obey the canonical commutation relation $[\hat{x}(t), \hat{p}(t)] = i\hbar 1$ (where 1 is now to be represented by the identity matrix).

Now the first problem would seem easy to solve: we know that the equations of motion are solved any linear combination of $\sin(\omega t)$ and $\cos(\omega t)$. But these are just number-valued functions and cannot by themselves obey the commutation relations! On the other hand, we do have matrices a and $a^\dagger$ that are beautifully noncommutative. But they are constant in time and therefore cannot alone describe the dynamics given by the equations of motion. This suggests that we try to construct the solution by combining the a and $a^\dagger$ matrices with the sine and cosine solutions. Let's try this ansatz[8]

$$\hat{x}(t) = \xi(t)a + \xi^*(t)a^\dagger, \tag{3.51}$$

and let us choose $\xi(t)$ to be a linear combination of the sine and cosine solutions to the equations of motion:

$$\xi(t) := r\sin(\omega t) + s\cos(\omega t) \tag{3.52}$$

Here, for now, r and s can be any complex numbers. By the way, whenever one makes an ansatz of this kind (and one often does, especially in quantum field theory), then $\xi(t)$ is called a mode function. The task now is to find out if we can find coefficients r and s such that the matrix-valued function $\hat{x}(t)$ and the corresponding $\hat{p}(t) = m\dot{\hat{x}}(t)$ obey the equations of motion, are hermitean and obey the canonical commutation relation.

Exercise 3.15 *Show that for any arbitrary choice of complex numbers r, s, the matrix-valued functions $\hat{x}(t)$ and $\hat{p}(t)$ defined through Eqs.3.51,3.52 obey the equations of motion at all time.*

Exercise 3.16 *Show that, again for any arbitrary choice of complex numbers r, s, the matrix-valued functions $\hat{x}(t)$ and $\hat{p}(t)$ defined through Eqs.3.51,3.52 obey the hermiticity conditions at all time.*

[8] An ansatz is an educated guess.

Exercise 3.17 *Find the equation that the complex numbers r, s have to obey so that the matrix-valued functions $\hat{x}(t)$ and $\hat{p}(t)$ defined through Eqs.3.51,3.52 obey the canonical commutation relations at all time. This equation for r, s is called the Wronskian condition and it has many solutions. Give an example of a pair of complex numbers r, s that obey the Wronskian condition and write down $\hat{x}(t)$ explicitly with these values for r, s filled in.*

Since there are many pairs r, s that obey the Wronskian condition, our ansatz given by Eqs.3.51,3.52 allows us to generate many solutions! We knew that there are always many linear representations of $\hat{x}(t)$ and $\hat{p}(t)$ in quantum theory (and that they are all physically equivalent because they are all related by changes of bases). Here, with our ansatz we have found already infinitely many of these representations of $\hat{x}(t)$ and $\hat{p}(t)$ for the harmonic oscillator. Actually, among the representations that we just found, some representations are particularly convenient and most of the time one uses one of those. These choices of r, s turn out to be convenient because the matrix-representation of the Hamiltonian $\hat{H}(t)$ is much simpler for clever choices of r, s than for other choices.

Exercise 3.18 *Use Eqs.3.51,3.52 to express the Hamiltonian in terms of functions and the operators $a, a^\dagger$. There should be terms proportional to a^2, to $(a^\dagger)^2, aa^\dagger$ and $a^\dagger a$.*

Exercise 3.19 *It turns out that it is possible to choose the coefficients r and s so that the terms in the Hamiltonian which are proportional to a^2 and $(a^\dagger)^2$ drop out. Find the condition which the equation that r and s have to obey for this to happen. Choose a pair of complex numbers r, s such that the Hamiltonian simplifies this way, and of course such that the Wronskian condition is obeyed. Write down $\hat{H}(t)$ as an explicit matrix for this choice of r, s. It should be a diagonal matrix.*

Remark: Notice that this convenient choice of r, s depends on the parameters m and ω of the harmonic oscillator. This means that each harmonic oscillator has its own optimal choices of parameters r, s. Making such wise choices of the parameters r, s is particularly useful in quantum field theory where each wavevector (and therefore frequency ω) of a quantum field has its own harmonic oscillator degree of freedom, and should therefore best have its own convenient choice of r, s that diagonalizes its Hamiltonian.

3.4.6 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 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.3.48,3.49.

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 first 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 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

$$\sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}(t) \phi_m \tag{3.53}$$

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, which we will denote by $\bar{f}(t)$ for a measurement of $\hat{f}(t)$ at time t, is[9] of the form:

$$\bar{f}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}(t) \psi_m \tag{3.54}$$

Now that we are using the same vector left and right, 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.3.54 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 $\hat{f}^*_{s,r} = \hat{f}_{r,s}$. Indeed[10]:

$$\bar{f}(t)^* = \left(\sum_{r,s=1}^{\infty} \psi_r^* \hat{f}_{r,s}(t) \psi_s \right)^* = \sum_{r,s=1}^{\infty} \psi_r \hat{f}^*_{r,s}(t) \psi_s^* = \sum_{r,s=1}^{\infty} \psi_s^* \hat{f}_{s,r}(t) \psi_r = \bar{f}(t) \tag{3.55}$$

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

$$\bar{x}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{x}_{n,m}(t) \psi_m \tag{3.56}$$

$$\bar{p}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{p}_{n,m}(t) \psi_m \tag{3.57}$$

$$\bar{L}^{(i)}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{L}^{(i)}_{n,m}(t) \psi_m \tag{3.58}$$

$$\bar{H}(t) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{H}_{n,m}(t) \psi_m \tag{3.59}$$

[9]There 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 $\bar{f}(t) = \mathrm{Tr}(\rho \hat{f}(t)) = \sum_{i,j} \rho_{i,j} \hat{f}_{j,i}(t)$ 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 $\mathrm{Tr}(\rho) = 1$.

[10]This 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!

$$\psi_n = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{pmatrix} \tag{3.60}$$

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

$$\psi_n \;=\; \frac{1}{5} \begin{pmatrix} 4 \\ 3i \\ 0 \\ 0 \\ \vdots \end{pmatrix} \tag{3.61}$$

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

Obviously the choice of ψ determines the predictions that we make for **all** possible 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 last observation reveals what the choice of the vector ψ means physically: the choice of ψ is the choice of initial conditions!

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, i.e., at the time t_0, 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 *any arbitrary observable,* $\hat{f}$, right at the initial time t_0. Since there are infinitely many observables, it is plausible that we get to choose infinitely coefficients in ψ to fix their initial conditions. For example, some of these observables are:

$$\bar{x}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{x}_{n,m}(t_0) \psi_m \tag{3.62}$$

$$\bar{p}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{p}_{n,m}(t_0) \psi_m \tag{3.63}$$

$$\bar{L}^{(i)}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{L}^{(i)}_{n,m}(t_0)\psi_m \tag{3.64}$$

$$\bar{H}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{H}_{n,m}(t_0)\psi_m \tag{3.65}$$

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

$$\bar{f}(t_0) = \sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}(t_0)\psi_m \tag{3.66}$$

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 $\bar{f}(\hat{x}(t), \hat{p}(t))$? If that were true, then this should be true:

$$\bar{f}(\hat{x}(t), \hat{p}(t)) = f(\bar{x}(t), \bar{p}(t)) \tag{3.67}$$

Actually, this equation does generally not hold in quantum mechanics! Mathematically, it is because, for example, when A is a square matrix and v is a vector then generally $v^\dagger A^n v \neq (v^\dagger A v)^n$. Therefore, just because we have a prediction for the position and the momentum does not mean that we have a prediction for other measurements such as the energy or the angular momentum!

Exercise 3.20 *Give a counter example for Eq.3.67. To this end, write out Eq.3.67 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.3.67 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 $\bar{f}$. 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

$$(\Delta(Q))^2 := \overline{\left(Q - \bar{Q}\right)^2} = \overline{Q^2} - \bar{Q}^2 \tag{3.68}$$

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

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 $(\hat{f}^2)^\dagger = \hat{f}^2$. And 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:

$$(\Delta f(t))^2 = \overline{(f(t) - \bar{f}(t))^2} = \overline{f^2(t)} - \bar{f}^2 \tag{3.69}$$

For example, from Eq.3.69:

$$(\Delta x(t))^2 = \sum_{r,s,t=1}^{\infty} \psi_r^* \hat{x}_{r,s} \hat{x}_{s,t} \psi_t - \left(\sum_{u,v=1}^{\infty} (\psi_u^* \hat{x}_{u,v} \psi_v) \right)^2 \tag{3.70}$$

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

$$\bar{f}(\hat{x}(t), \hat{p}(t)) \neq f(\bar{x}(t), \bar{p}(t)), \tag{3.71}$$

the predictions should come with uncertainty. They should be statistical. Our predictions for observables $\bar{f}(t)$ such as $\bar{x}(t), \bar{p}(t), \bar{H}(t), \bar{L}_i(t)$, ... 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.3.48,3.49, we can calculate the mean 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.3.56,3.57. 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 Eq.3.69 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.3.54 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 $\hat{f}$:

$$\bar{x}(t_0) = a_1, \quad \bar{p}(t_0) = a_2, \quad \overline{x^2}(t_0) = a_3, \quad \overline{p^2}(t_0) = a_4, \quad \overline{x^3}(t_0) = a_5, \quad ... \tag{3.72}$$

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.3.72, which are infinitely many equations for the unknown vector components $\{\psi_i\}$ in terms of the given coefficients $\{a_j\}$:

$$\sum_{i,j} \psi_i^* \hat{x}_{i,j}(t_0)\psi_j = a_1, \qquad \sum_{i,j} \psi_i^* \hat{p}_{i,j}(t_0)\psi_j = a_2, \qquad \sum_{i,j,k} \psi_i^* \hat{x}_{i,k}\hat{x}_{k,j}(t_0)\psi_j = a_3, \quad ... \tag{3.73}$$

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 3.21 *Verify that ψ of Eq.3.61 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.

Exercise 3.22 *Spell out the step of the second equality in Eq.3.68.*

3.7 The Hilbert space of quantum mechanics, and Dirac's notation

In the previous sections, we solved the equations of motion for matrix-valued functions $\hat{x}(t)_{ij}$ and $\hat{p}(t)_{ij}$. And, of course, once we have the $\hat{x}(t)_{ij}$ and $\hat{p}(t)_{ij}$, we can easily calculate from them the matrix-valued function of time $\hat{f}(t)_{ij}$ for every observable $\hat{f}$, i.e., for every polynomial in the $\hat{x}$ and $\hat{p}$ that obeys $\hat{f} = \hat{f}^\dagger$.

We also found out how to specify initial conditions, namely 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 an abstract vector. Following Dirac, we will call that abstract vector a "ket" and denote by $|\psi\rangle$. By working with column vectors and matrices, we have implicitly chosen a basis in the vector space. Let us give these basis vectors the names $|b_n\rangle$, for $n = 1, 2, ...$

The fact that the numbers ψ_i are the coefficients of $|\psi\rangle$ in the basis $\{|b_n\rangle\}_{n=1}^{\infty}$ that we have implicitly chosen is expressed in the equation:

$$|\psi\rangle = \sum_n \psi_n |b_n\rangle \tag{3.74}$$

Therefore, the abstract vector $|\psi\rangle$ is an element of the abstract vector space of initial state vectors. Once we choose a basis, then the abstract vector $|\psi\rangle$ has a unique collection of coefficients ψ_n that expand $|\psi\rangle$ in that basis. The coefficients ψ_n are usually arranged in a column and that column of numbers is also being called a *column vector*. But keep in mind that this column is really just the collection of the numbers ψ_n. This *column vector* of coefficients is not the state vector $|\psi\rangle$. Instead, one says

that the column vector of the coefficients *represents* the abstract initial state vector $|\psi\rangle$ in the basis $\{|b_n\rangle\}_{n=1}^{\infty}$.

Similarly, the matrix elements $\hat{x}(t)_{ij}$ themselves are of course simply numbers. They are the coefficients of the abstract linear map $\hat{x}(t)$ from the vector space into itself, in the basis $\{|b_n\rangle\}_{n=1}^{\infty}$. We say that the matrices, such as $(\hat{x}(t)_{ij})_{i,j=1}^{\infty}$ provide a *representation* of the abstract operator $\hat{x}(t)$ that maps the abstract vector space of initial state vectors into itself. Whenever then we change basis in the vector space, as we will freely do later, then this implies a corresponding change of the coefficients of the matrices $\hat{x}(t)_{ij}, \hat{p}(t)_{ij}$, ..., $\hat{f}(t)_{ij}$ as well as the coefficients ψ_i of the vector ψ. The abstract initial state vector $|\psi\rangle$ and the abstract operators such as $\hat{x}(t), \hat{p}(t)$ and $\hat{f}(t)$ stay the same as we change the choice of basis. But their representation in terms of column vectors and matrices will change[11].

The representations that we have used so far have been so-called discrete representations, i.e., representations that use a basis of countably many, i.e., discretely-labeled basis vectors $\{|b_n\rangle\}_{n=1}^{\infty}$. One of the perks of working with infinite-dimensional vector spaces is that they also admit so-called continuous representations. One may loosely think of these representations as using a set of continuously-labeled basis $\{|b_\lambda\rangle\}_{\lambda\in K}$ where $K = \mathbb{R}$ or $K \subset \mathbb{R}$. It is also possible to have representations using bases that are in part discrete and in part continuous[12]. This means that the way that concrete quantum mechanical calculations look like can depend very much on which basis one chooses to work with, i.e., on the choice of representation.

Always, however, the number-valued predictions of expectation values of a quantum theory only depend on the abstract operators $\hat{x}(t), \hat{p}(t)$,... and on the abstract vectors ψ. The predictions, $\bar{x}(t), \bar{p}(t)$, ... $\bar{f}(t)$ are scalars, i.e., they are basis independent. Choices of basis, i.e., choices of representation are merely issues of convenience.

Let us now study what ultimately matters, namely the abstract structure of the vector space of initial state vectors and the structure of the abstract operators acting as maps on this abstract vector space.

3.7.1 Hilbert spaces

We begin by reviewing the definition of a complex vector space. In brief, a complex vector space 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

[11] Notice that we can always rotate any basis 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.

[12] For example, the eigenbasis of the Hamiltonian of the Hydrogen atom is of this form, containing a countable number of eigenstates that are bound states and an uncountable number of eigenstates that are scattering states.

a) it possesses an operation $\mathcal{H} \times \mathcal{H} \to \mathcal{H}$ 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 $\mathbb{C} \times \mathcal{H} \to \mathcal{H}$ obeying the following axioms for all $|v\rangle, |w\rangle \in \mathcal{H}$ and for all $\alpha, \beta \in \mathbb{C}$:

$$(\alpha + \beta)|v\rangle = \alpha|v\rangle + \beta|v\rangle \tag{3.75}$$
$$\alpha(|v\rangle + |w\rangle) = \alpha|v\rangle + \alpha|w\rangle \tag{3.76}$$
$$(\alpha\beta)|v\rangle = \alpha(\beta|v\rangle) \tag{3.77}$$
$$1|v\rangle = |v\rangle \tag{3.78}$$

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 continuous[13] linear maps $\hat{v} \to \mathbb{C}$. We call the 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:

$$\langle r| : \quad \mathcal{H} \to \mathbb{C} \tag{3.79}$$
$$\langle r| : \quad |v\rangle \to \langle r|v\rangle \tag{3.80}$$

That they are linear maps means:

$$\langle r| : \quad (\alpha|v\rangle + \beta|w\rangle) \to \alpha\langle r|v\rangle + \beta\langle r|w\rangle \tag{3.81}$$

Exercise 3.23 *Verify that $\mathcal{H}^*$ is a complex vector space.*

[13] Aren't all linear maps continuous? Well no, not necessarily in infinite-dimensional spaces! Consider for example the map $\phi : \mathcal{H} \to \mathbb{C}$ that maps vectors $|\psi\rangle$ into numbers through $|\psi\rangle \to \phi(|\psi\rangle) := \sum_n n!\psi_n$ in some basis. The map ϕ is clearly a map from the vector space into the complex numbers that is linear. But ϕ is not continuous. To see this, consider that arbitrarily small changes to the vector $|\psi\rangle$ (i.e., to the coefficients ψ_n) can change $\phi(|\psi\rangle)$ arbitrarily strongly.

Definition: A complex vector space is said to possess a *scalar product* or also called *inner product* if it possesses a map $\mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that obeys certain conditions that we'll discuss below. So an inner product takes two vectors in $\mathcal{H}$ and maps then into a number. Equivalently, a vector space with an inner product is a vector space that possesses a map $\mathcal{H} \to \mathcal{H}^*$ obeying certain conditions. How does this yield an inner product of two vectors in $\mathcal{H}$? Take the first vector, map it into a vector in $\mathcal{H}^*$ and then use that dual vector to linearly map the second vector in $\mathcal{H}$ to obtain a number, which is then the value of the inner product.

Concretely, such a map is usually denoted by $^\dagger : \mathcal{H} \to \mathcal{H}^*$ and it is called a hermitean conjugation map. We can use the hermitean conjugation map to map any element $|v\rangle \in \mathcal{H}$ into an element in $|v\rangle^\dagger \in \mathcal{H}^*$. One could give this image of the vector $|v\rangle$ any name but one traditionally chooses to call it by the same name as its pre-image: $\langle v| := |v\rangle^\dagger \in \mathcal{H}^*$. The hermitian conjugation map is required to be what is called anti-linear, i.e., linear but with a complex conjugation:

$$^\dagger : \quad \mathcal{H} \to \mathcal{H}^* \tag{3.82}$$

$$^\dagger : \quad r|v\rangle + s|w\rangle \to r^*\langle v| + s^*\langle w| \qquad \forall |v\rangle, |w\rangle \in \mathcal{H}, \forall r, s \in \mathbb{C} \tag{3.83}$$

What then is the list of requirements that defines an inner product, or scalar product? 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 these conditions:

$$\langle u| \left(\alpha|v\rangle + \beta|w\rangle\right) = \alpha\langle u|v\rangle + \beta\langle u|w\rangle \tag{3.84}$$

$$\langle v|w\rangle^* = \langle w|v\rangle \tag{3.85}$$

$$\langle v|v\rangle \geq 0 \tag{3.86}$$

$$\langle v|v\rangle = 0 \quad \text{only if} \quad |v\rangle = 0 \tag{3.87}$$

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.3.74 we need to define what it means to sum up an infinite number of basis vectors.

Definition: We say that a sequence of vectors $|v_n\rangle$ converges (namely to some vector $|v\rangle$), and we write $\lim_{n\to\infty} |v_n\rangle = |v\rangle$, iff (i.e., if and only if) it is true that

$\lim_{n\to\infty} d(|v_n\rangle, |v\rangle) = 0$.

Definition: A sequence $\{|v_n\rangle\}$ is called fundamental (or Cauchy), if and only if $\lim_{n,m\to\infty} d(|v_n\rangle, |v_m\rangle) = 0$. By this is meant that $\forall\epsilon > 0$, $\exists R$ so that $\forall n, m > R$: $d(|\hat{v}_n\rangle, |v_m\rangle) < \epsilon$.

Definition: A pre-Hilbert space is called a Hilbert space if all its fundamental sequences converge[14].

Remark: Since the distance to a vector of infinite length would be infinite, no sequence could converge to a vector of infinite length. 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 $\hat{f}_{r,s}$?

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:

$$|\psi\rangle = \sum_n \psi_n |b_n\rangle \tag{3.88}$$

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

Are they big enough for quantum mechanics? Yes, actually we know already that in quantum mechanics we can work with a separable Hilbert because we have succeeded

[14]This means that in order to obtain a Hilbert space from a given pre-Hilbert space it may be necessary to add new elements to it - for the Cauchy sequences to converge to - until every Cauchy sequence has an element to converge to. This is similar to how the set of real numbers $\mathbb{R}$ is obtained from the set of rational numbers, $\mathbb{Q}$: There are Cauchy sequences in $\mathbb{Q}$ which have no rational number to converge to. Consider, for example, the sequence of rational numbers s_n where $s_n = trunk_n(\sqrt{2})$, namely, s_n is given by the first n digits of $\sqrt{2}$. Clearly, the sequence (s_n) is a Cauchy sequence but it does not converge to any number in $\mathbb{Q}$. That's because $\sqrt{2} \notin \mathbb{Q}$. So we add $\sqrt{2}$ and all other such irrational numbers to finally obtain $\mathbb{R}$. Now you may say yes but coming from $\mathbb{Q}$ we don't know what, for example, $\sqrt{2}$ is. How can we add it to $\mathbb{Q}$ then? Good that you ask! Here is roughly how you can do it. Consider all Cauchy sequences that seem to want to converge to the same point, say $\sqrt{2}$. How do we know that two Cauchy sequences want to converge to the same point? When their difference converges to zero. All these Cauchy sequences that want to converge to the same point in that sense form an equivalence class of sequences. We can call that equivalence class $\sqrt{2}$. In this way, any real number is really a set of equivalence classes of Cauchy sequences, where the elements within an equivalence class all converge to the same point (in the sense above). The completion of a pre-Hilbert space to obtain a Hilbert space is analogous.

above in representing the canonical commutation relations in a representation of matrices and column vectors. In fact, in quantum mechanics, it always suffices to consider a separable Hilbert space. Separable Hilbert spaces are the smallest Hilbert spaces after finite-dimensional Hilbert spaces. 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 single 1-dimensional harmonic oscillator has the same dimension (namely countable infinity) as does the Hilbert space of a space ship[15].

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.3.88, 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:

$$\psi_n = \langle b_n|\psi\rangle \tag{3.89}$$

Eq.3.89 is easily verified by applying $\langle b_m|$ from the left in Eq.3.88:

$$\langle b_m|\psi\rangle \;\; = \;\; \langle b_m| \sum_n \psi_n |b_n\rangle$$

[15]Following up on previous comments: In 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.

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

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 $|x\rangle$.

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:

$$\mathbf{1} = \sum_{\mathrm{n}} |\mathrm{b_n}\rangle\langle \mathrm{b_n}| \tag{3.90}$$

This is called the resolution of the identity in terms of the basis vectors $|b_n\rangle$. For example, using Eq.3.89:

$$\mathbf{1}|\psi\rangle = \sum_{\mathrm{n}=1}^{\infty} |\mathrm{b_n}\rangle\langle \mathrm{b_n}|\psi\rangle = \sum_{\mathrm{n}} \psi_{\mathrm{n}} |\mathrm{b_n}\rangle \tag{3.91}$$

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:

$$\begin{aligned} \langle\psi|\psi\rangle &= \langle\psi|1|\psi\rangle && (3.92)\\ &= \sum_n \langle\psi|b_n\rangle\langle b_n|\psi\rangle && (3.93)\\ &= \sum_n \psi_n^* \psi_n && (3.94) \end{aligned}$$

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

$$\sum_n \psi_n^* \psi_n < \infty \tag{3.95}$$

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

$$\hat{f} = \sum_{r,s=1}^{\infty} |b_r\rangle \; f_{r,s} \; \langle b_s| \tag{3.96}$$

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

$$\hat{f}_{r,s} = \langle b_r|\hat{f}|b_s\rangle \tag{3.97}$$

This equation follows from Eq.3.96 by applying to it $\langle b_n|$ 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 $\hat{f}|\psi\rangle \in \mathcal{H}$. This is 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 this matrix:

$$(f)_{n,m} := \begin{pmatrix} 1^2 & 0 & 0 & 0 & \\ 0 & 2^2 & 0 & 0 & \\ 0 & 0 & 3^2 & 0 & \\ 0 & 0 & 0 & 4^2 & \\ & & & & \ddots \end{pmatrix}_{n,m} \tag{3.98}$$

The column vector[16]

$$(\psi)_n = \begin{pmatrix} 1/1 \\ 1/2 \\ 1/3 \\ 1/4 \\ \vdots \end{pmatrix}_n \tag{3.99}$$

is square summable[17]. Therefore, the coefficients ψ_n define a vector $|\psi\rangle \in \mathcal{H}$. Now the action of the matrix $(f)_{n,m}$ on this column vector would yield a column vector with the components $\phi_n = \sum_{m=1}^{\infty} f_{nm}\psi_m = n$. But the length of this column vector ϕ would be infinite because $\sum_{n=1}^{\infty} n^2 = \infty$. Therefore, the numbers ϕ_n cannot be the coefficients of any vector in the Hilbert space relative to any basis. Therefore, the abstract operator $\hat{f}$ defined by the matrix above cannot be allowed to act on the abstract Hilbert space vector $|\psi\rangle$ defined by the column vector above (namely because it would map it outside the Hilbert space).

Definition: We call the set $D_f \subseteq \mathcal{H}$ of vectors on which $\hat{f}$ is allowed to act the *domain* of $\hat{f}$. In general, we have some freedom in choosing the domain of an operator. But

[16]Keep in mind that this so-called 'column vector' is really just the collection of the coefficients of an abstract Hilbert space vector relative to a choice of basis.

[17]Namely, $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$. In the year 1735, Leonard Euler, who was 28 at the time, was the first to solve this tough old problem and that started his fame.

we have to keep in mind that domains must always be chosen such that the image of every vector in the domain is also inside the Hilbert space:

$$\forall|v\rangle \in D_f : \hat{f}|v\rangle \in \mathcal{H} \tag{3.100}$$

Note that we can allow the matrix $(\hat{f})_{n,m}$ of Eq.3.98 to act, for example, 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.3.98. But of course the domain of this $\hat{f}$ can be made a little larger by allowing it to act on vectors with infinitely many nonzero coefficients that, however, have to decay fast enough.

Exercise 3.24 *Let us consider how large we can make the domain $D_f \subset \mathcal{H}$ of the operator $\hat{f}$ defined in Eq.3.98. Concretely, how fast do the coefficients ψ_n of $|\psi\rangle$ have to decay, as $n \to \infty$. For this exercise, assume that the coefficients $|\psi_n|$ scale as n^s for $n \to \infty$. What then are the allowed values for s? Hint: The so-called Dirichlet series representation of the Riemann zeta function $\zeta(r) := \sum_{n=1}^{\infty} 1/n^r$ diverges for $r \leq 1$ but converges for $r > 1$.*

3.7.5 Changes of basis

Using the Dirac bra-ket notation, we can now rewrite equations such as Eq.3.54 in a much more compact form:

$$\bar{f}(t) = \langle\psi|\hat{f}(t)|\psi\rangle \tag{3.101}$$

We can easily recover Eq.3.54 from Eq.3.101, simply by twice inserting the resolution of the identity:

$$\begin{aligned}
\bar{f}(t) &= \langle\psi|\hat{f}(t)|\psi\rangle && (3.102)\\
&= \langle\psi|1\hat{f}(t)1|\psi\rangle && (3.103)\\
&= \sum_{n,m=1}^{\infty} \langle\psi|b_n\rangle\langle b_n|\hat{f}|b_m\rangle\langle b_m|\psi\rangle && (3.104)\\
&= \sum_{n,m=1}^{\infty} \psi_n^* \hat{f}_{n,m}\psi_m && (3.105)
\end{aligned}$$

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, say $\{|b_n\rangle\}_{n=1}^{\infty}$ then we can simply suitably insert resolutions of the identity in terms of the basis vectors. If we then wish to change basis then we can do this by inserting resolutions of the identity for the new basis, say $\{|c_n\rangle\}_{n=1}^{\infty}$:

$$\bar{f}(t) = \langle\psi|\hat{f}(t)|\psi\rangle \tag{3.106}$$

$$= \sum_{n,m} \langle\psi|b_n\rangle\langle b_n|\hat{f}|b_m\rangle\langle b_m|\psi\rangle \tag{3.107}$$

$$= \sum_{n,m} \langle\psi|1|b_n\rangle\langle b_n|1\hat{f}1|b_m\rangle\langle b_m|1|\psi\rangle \tag{3.108}$$

$$= \sum_{n,m,r,s,u,v} \langle\psi|c_r\rangle\langle c_r|b_n\rangle\langle b_n|c_s\rangle\langle c_s|\hat{f}|c_u\rangle\langle c_u|b_m\rangle\langle b_m|c_v\rangle\langle c_v|\psi\rangle \tag{3.109}$$

Let us denote the linear map that maps the basis vectors of one basis into the basis vectors of the other basis by $\hat{U}$:

$$\hat{U}: \ |b_n\rangle \rightarrow \hat{U}|b_n\rangle = |c_n\rangle \tag{3.110}$$

Since $\hat{U}$ is a linear map, it can be written in the form:

$$\hat{U} = \sum_{r,s} |b_r\rangle \ U_{rs} \ \langle b_s| \tag{3.111}$$

Applying $\langle b_m|$ to the right of $\hat{U}|b_n\rangle = |c_n\rangle$ we can now identify the change of basis coefficents of Eq.3.109 as the matrix elements of $\hat{U}$:

$$\langle b_m|c_n\rangle = U_{mn} \tag{3.112}$$

Exercise 3.25 *By comparing Eqs.3.108 and Eq.3.109, find out how to express the wave function of $|\psi\rangle$ in the b basis in terms of its wave function in the c basis, using the matrix of $\hat{U}$. Also express the representation of $\hat{f}$ as a matrix in the b basis in terms of its representation as a matrix in the c basis, using the matrix representation of $\hat{U}$. Finally write Eq.3.109 in terms of a row vector, some matrices and a column vector.*

Exercise 3.26 *Assume that the basis vectors $|b_n\rangle$ are orthonormal. Show that if the new vectors $|c_n\rangle$ are to be an orthonormal basis too then $\hat{U}$ has to obey $\hat{U}^\dagger = \hat{U}^{-1}$. This means that $\hat{U}$ is what is called a unitary operator.*

Definition: An operator that obeys $\hat{U}^\dagger = \hat{U}^{-1}$ is called a unitary operator.

Remark: *In real vector spaces, matrices R that obey $R^t = R^{-1}$ are called rotation matrices because they rotate orthonormal basis vectors into orthonormal basis vectors. Here, we are working with Hilbert spaces, i.e., with complex vector spaces. Here, as Eq.3.110 shows, the unitaries play the role of rotations of orthonormal basis vectors into orthonormal basis vectors. Notice that R^t is defined through $R^t_{nm} = R_{mn}$, i.e., it is the transpose of R. In contrast, $U^\dagger$ is the transposed and also complex conjugated of U, i.e., it obeys $U^\dagger_{nm} = (U_{mn})^*$.*

In Dirac's notation it is clear 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 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 when we represented $\hat{x}$ and $\hat{p}$ using the matrices a and $a^\dagger$, in Eqs.3.42,3.43 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 3.27 *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 normalized*[18] *vector, which we denote by $|0\rangle$, which obeys $b|0\rangle = 0$. 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.*

[18]Notice 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.

Chapter 4

Eigenbases of Observables and the Spectral Theorem

We saw 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 4.1 *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 an orthonormal basis of eigenvectors of $\hat{Q}$ in the Hilbert space. The situation is more subtle in infinite dimensions, as the full spectral theorem for self-adjoint operators on general Hilbert spaces shows. Physicists speak of self-adjoint operators that possess non-normalizable eigenvectors which are not in the Hilbert space. What is the underlying mathematics?

4.1 Self-adjointness

What we have to look out for when defining self-adjointness in the infinite-dimensional case is that an operator can only act on those Hilbert space vectors that it maps into vectors that are in the Hilbert space, i.e., that are again normalizable. Since this requirement is nontrivial in infinite dimensional Hilbert spaces, the domain $D_{\hat{Q}}$ of an operator $\hat{Q}$, i.e., the set of vectors on which it is allowed to act can be smaller than the entire the Hilbert space.

Actually, for an operator on an infinite-dimensional Hilbert space it happens easily that its domain is smaller than the Hilbert space. For example, recall the operator

$\hat{Q}$ whose matrix representation in a Hilbert basis is diagonal with the diagonal entries being $1, 2^5, 3^5, 4^5, ...$, and consider the vector with the coefficients $(1, 1/2, 1/3, 1/4, ...)$ in the same basis. The vector is is in the Hilbert space because its coefficients are square-summable. But this vector is not in the domain of the operator $\hat{Q}$. That's because the image of our vector would be the vector with the coefficients $(1, 2^4, 3^4, 4^4, ...)$ 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:

$$D_{\hat{Q}^\dagger} = \Big\{ |\phi\rangle \in \mathcal{H} \mid \exists\ |\varphi\rangle \in \mathcal{H} \ \text{ so that } \ \langle\varphi|\psi\rangle = \langle\phi|\hat{Q}|\psi\rangle \ \ \forall |\psi\rangle \in D_{\hat{Q}} \Big\} \tag{4.1}$$

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

$$\hat{Q}^\dagger |\phi\rangle = |\varphi\rangle \tag{4.2}$$

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: $D_{\hat{Q}} = D_{\hat{Q}^\dagger}$.

4.2 The spectrum of an operator

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

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

$$\hat{Q}|\psi\rangle = \lambda|\psi\rangle \tag{4.3}$$

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.4.3 in this form:

$$(\hat{Q} - \lambda 1)|\psi\rangle = 0 \tag{4.4}$$

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 if λ is an eigenvalue of $\hat{Q}$, then the operator $(\hat{Q} - \lambda 1)$ is not invertible[1].

This is a very important observation because we can now refer to eigenvalues without referring to eigenvectors! That will come in handy when we want to deal with what would be eigenvectors if they weren't non-normalizable. 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 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 - in can happen that there are values λ for which $(\hat{Q} - \lambda 1)$ does have an inverse $(\hat{Q} - \lambda 1)^{-1}$, but this inverse 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 $\mathcal{H}$ and (c) there are Hilbert space vectors orthogonal to the domain of $(\hat{Q} - \lambda 1)^{-1}$ (in this case, we say that the domain of $(\hat{Q} - \lambda 1)^{-1}$ is not dense in $\mathcal{H}$). The set of these λ forms 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.

[1]Technically, the operator is not injective. Let's 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:

$$\forall \epsilon > 0 \quad \exists |\psi\rangle \in \mathcal{H} \quad \text{with} \;\; || \; |\psi\rangle \; || = 1 \;\; \text{so that} \;\; ||(\hat{Q} - \lambda 1)|\psi\rangle|| < \epsilon \tag{4.5}$$

Therefore, there will always be Hilbert space vectors that approximate what would be eigenvectors arbitrarily closely, namely in the sense of Eq.4.5. It is in this sense that one speaks of an approximate eigenvalue having an approximate eigenvector or an improper eigenvector or a non-normalizable eigenvector.

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:

$$\int_a^b f(x) \; \delta(x) \; dx = \lim_{\epsilon \to 0} \int_a^b f(x) \; \frac{\epsilon}{\pi} \frac{1}{x^2 + \epsilon^2} \; dx \tag{4.6}$$

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.

4.3 The spectral theorem for self-adjoint operators

Theorem (Spectral theorem for self-adjoint operators): Assume that $\hat{f}$ is a self-adjoint operator: $\hat{f} = \hat{f}^\dagger$. Then, the spectrum of $\hat{f}$ is real and consists of a point spectrum or a continuous spectrum, or both. The operator $\hat{f}$ possesses an eigenbasis consisting of a discrete set of eigenvectors corresponding to the eigenvalues from the point spectrum and a continuous set of approximate eigenvectors corresponding to the continuous spectrum.

Exercise 4.2 *Assume that $|\psi\rangle$ is an eigenvector of a self-adjoint operator $\hat{f}$ with eigenvalue λ. Show that λ is real.*

Definition: Instead of the terms eigenvectors and approximate eigenvectors, one sometimes also speaks of proper and improper eigenvectors respectively.

In the case of self-adjoint operators on finite-dimensional Hilbert spaces, the spectrum consists purely of a point spectrum because in finite dimensional Hilbert spaces all vectors can be normalized. 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 $q_1, ..., q_N$. In this case, it is easy to write down a resolution of the identity by summing over the values of the spectrum:

$$1 \;=\; \sum_{n=1}^{N} |q_n\rangle\langle q_n| \tag{4.7}$$

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

$$\hat{Q} \;=\; \sum_{n=1}^{N} q_n \; |q_n\rangle\langle q_n| \tag{4.8}$$

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 which we will come back to in Sec.6.2.

4.3.1 Case 1: The self-adjoint operator $\hat{f}$ possesses only a point spectrum

We covered this case before. Let's do a quick review and close with an important new definition at the end of the section. So there is then an eigenbasis $\{|f_n\rangle\}$ of vectors obeying:

$$\hat{f}\ |f_n\rangle = f_n\ |f_n\rangle \quad \text{and} \quad \langle f_n|f_m\rangle = \delta_{nm} \tag{4.9}$$

We can sum them up to obtain a resolution of the identity:

$$\mathbf{1} = \sum_n |f_n\rangle\langle f_n| \tag{4.10}$$

It allows us to quickly expand vectors in the eigenbasis of $\hat{f}$:

$$|\psi\rangle = \mathbf{1}|\psi\rangle = \sum_n |f_n\rangle\langle f_n|\psi\rangle = \sum_n \psi_n|f_n\rangle \tag{4.11}$$

In the last step we used this definition:

Definition: The collection of numbers $\psi_n = \langle f_n|\psi\rangle$ is called the wave function of the state $|\psi\rangle$ in the eigenbasis of $\hat{f}$.

Notice that it is a function of n which takes only discrete values because the spectrum is here discrete. Later, in the case of the continuous spectrum, the wave function will be defined over the continuous spectrum.

The resolution of the identity also allows us to quickly expand arbitrary operators, $\hat{g}$, in the eigenbasis of $\hat{f}$:

$$\hat{g} = \mathbf{1}\hat{g}\mathbf{1} = \sum_{n,m} |f_n\rangle\langle f_n|\hat{g}|f_m\rangle\langle f_m| = \sum_{n,m} g_{nm}|f_n\rangle\langle f_m| \tag{4.12}$$

In the last step we used this definition:

Definition: We call $g_{nm} := \langle f_n|\hat{g}|f_m\rangle$ the matrix elements of the operator $\hat{g}$ in the eigenbasis of $\hat{f}$.

For example, if we choose for $\hat{g}$ to be $\hat{g} = \hat{f}$ then we obtain

$$f_{nm} = \langle f_n|\hat{f}|f_m\rangle = f_n\ \langle f_n|f_n\rangle = f_n\ \delta_{nm} \tag{4.13}$$

i.e., $\hat{f}$ is diagonal in its own eigenbasis, as expected. For illustration, let us also express an equation, say

$$|\phi\rangle = \hat{g}|\psi\rangle \tag{4.14}$$

in the $\hat{f}$ eigenbasis. We begin by multiplying the equation with $\langle f_r|$ from the left:

$$\langle f_r|\phi\rangle = \langle f_r|\hat{g}|\psi\rangle \tag{4.15}$$

Therefore:

$$\langle f_r|\phi\rangle = \langle f_r|\hat{g}\mathbf{1}|\psi\rangle = \sum_n \langle f_r|\hat{g}|f_n\rangle\langle f_n|\psi\rangle \tag{4.16}$$

We obtain the following equation for explicit column vectors (the wave functions) and a matrix:

$$\phi_r = \sum_n g_{r,n}\psi_n \tag{4.17}$$

We obtained these expansions of vectors and operators and their equations in eigenbases by using the resolution of the identity. It it will be very useful later also to consider partial resolutions of the identity, namely by not summing all the way:

Definition: We define a partial resolution of the identity:

$$\hat{E}_N := \sum_n^N |f_n\rangle\langle f_n| \tag{4.18}$$

Clearly, in the limit, we obtain the resolution of the identity:

$$\lim_{N\to\infty} \hat{E}_N = \mathbf{1} \tag{4.19}$$

It will be important later that each $\hat{E}_N$ is an operator, defined on the entire Hilbert space, which is a projector, namely, that it obeys $\hat{E}_N = \hat{E}_N^\dagger$ and $\hat{E}_N^2 = \hat{E}_N$.

Definition: An operator $\hat{P}$ is called a projector if it is self-adjoint, and obeys: $\hat{P}^2 = \hat{P}$.

Exercise 4.3 *Show that $\hat{E}_N$ is a projector.*

We now come to case 2 of the spectral theorem:

4.3.2 Case 2: The spectrum of $\hat{f}$ is purely continuous

In this case, let us give the spectrum a name:

$$J := \mathrm{spec}(\hat{f}) \subseteq \mathbb{R} \tag{4.20}$$

Notice that J can be a subset of the real line but it can also be the full real line. J can also consist of several disconnected pieces[2]. To start with, let us use the notation that

[2]For example, the Hamiltonian of electrons in a periodic potential, e.g., in a crystal, tends to possess a continuous spectrum that consists of several disjoint stretches, called bands. Insulators, semiconductors and conductors are distinguished by the degree to which the states in the highest energetically accessible band are filled with electrons.

is common in physics. In physics notation, $\hat{f}$ has a basis of approximate eigenvectors $\{|f\rangle\}_{f\in J}$ that are continuum normalized

$$\langle f|f'\rangle = \delta(f - f') \tag{4.21}$$

and yield a resolution of the identity:

$$\mathbf{1} = \int_J |f\rangle\langle f| \; df \tag{4.22}$$

Similarly to case 1 where $\hat{f}$ possesses only a purely point spectrum, here too we obtain wave functions for states and matrix elements for operators:

$$|\psi\rangle = \mathbf{1}|\psi\rangle = \int_J |f\rangle\langle f|\psi\rangle \; df = \int_J \psi(f) \; |f\rangle \tag{4.23}$$

In the last step we used this definition:

Definition: We call $\psi(f) := \langle f|\psi\rangle$ the wave function of the state $|\psi\rangle$ in the eigenbasis of $\hat{f}$.

Further, we obtain for operators:

$$\hat{g} = \mathbf{1}\hat{g}\mathbf{1} = \int_J \int_J |f\rangle\langle f|\hat{g}|f'\rangle\langle f'| \; df \; df' = \int_J \int_J g(f, f')|f\rangle\langle f'| \; df \; df' \tag{4.24}$$

In the last step we used this definition:

Definition: We call $g(f, f') := \langle f|\hat{g}|f'\rangle$ the matrix elements of $\hat{g}$ in the eigenbasis of $\hat{f}$. We may also sometimes use the notation $g_{f,f'}$ for $g(f, f')$. The function $g(f, f')$ is also called the integral kernel of $\hat{g}$ in the continuous eigenbasis of $\hat{f}$.

Notice that $g_{f,f'}$ behaves like a matrix with continuous rather than discrete indices. In fact, everything is the same as for a matrix, except that summations over indices are now integrations.

For example, let us express the equation $|\phi\rangle = \hat{g}|\psi\rangle$ in the $\hat{f}$ eigenbasis, as in the discussion of case 1 above. Now, however, we pick an operator $\hat{f}$ whose spectrum is continuous. We begin by multiplying the equation with $\langle\hat{f}|$ from the left:

$$\langle f|\phi\rangle = \langle f|\hat{g}|\psi\rangle \tag{4.25}$$

Therefore:

$$\langle f|\phi\rangle = \langle f|\hat{g}\mathbf{1}|\psi\rangle = \int_J \langle f|\hat{g}|f'\rangle\langle f'|\psi\rangle \; df' \tag{4.26}$$

We obtain the following equation for explicit column vectors (the wave functions) and a matrix, though now all indices are continuous:

$$\phi(f) = \int_J g(f, f')\psi(f') \; df' \tag{4.27}$$

Let us also quickly consider the matrix elements, i.e., the integral kernel, of $\hat{f}$ in its eigenbasis, in the case where the spectrum is continuous:

$$f(f, f') = \langle f|\hat{f}|f'\rangle = f'\langle f|f'\rangle = f'\ \delta(f - f') \tag{4.28}$$

Therefore:

$$\hat{f} = \int_J \int_J f'\ \delta(f - f')|f\rangle\langle f'|\ df\ df' = \int_J f|f\rangle\langle f|\ df \tag{4.29}$$

4.3.3 Case 3: The self-adjoint operator $\hat{f}$ has a point spectrum and a continuous spectrum

In this case, we have:

$$\mathbf{1} = \sum_n |f_n\rangle\langle f_n| + \int_J |f\rangle\langle f|\ df \tag{4.30}$$

$$\hat{f} = \sum_n f_n\ |f_n\rangle\langle f_n| + \int_J f\ |f\rangle\langle f|\ df \tag{4.31}$$

Exercise 4.4 *Derive the expression for the abstract equation $|\phi\rangle = \hat{g}|\psi\rangle$ in this basis.*

Exercise 4.5 *Look up and state the complete spectrum of the Hamiltonian of the Hydrogen atom (without spin or higher order corrections of any form). Give the exact source in a textbook.*

Hint: Essentially every textbook on quantum mechanics treats the Hydrogen atom. What do they do, from our perspective? They start with the Hamiltonian of the Hydrogen atom, expressed in the position basis. Then they calculate the eigenbasis of the Hamiltonian, i.e., they calculate the energy eigenstate's wave functions in the position basis. They are the wave functions of the Hydrogen orbitals. In principle, it's just like we will calculate in Ch.**??** the eigenstates of the position operator in the eigenbasis of a harmonic oscillator and vice versa. But never mind the details of this. It's not what the question is about.

The question is merely about the spectrum of the Hamiltonian of the Hydrogen atom. What are the proper eigenvalues and what are the approximate eigenvalues? I.e., the task is for you to learn to read the literature with the new mathematical and physical insights in the background. The question is to thereby find out the point spectrum and also the continuous spectrum of the Hamiltonian of the Hydrogen atom. Yes, the Hamiltonian of the Hydrogen atom has both types of spectrum. If your textbook doesn't mention the continuous spectrum (whose eigenfunctions describes electrons that are not bound to the proton but merely scatter off it), check another textbook. Regarding the point spectrum, do list also how often each eigenvalue occurs.

Remember that eigenvalues can be degenerate, i.e., the eigenspace to an eigenvalue can be multi-dimensional. I.e., list the dimension of each eigenvalue's eigenspace. No need to list the wave functions of the Hamiltonian's eigenvectors.

Comment: We can see here why a smart phone battery demands about 5V from the USB charger and why nuclear fuel can release about a million times more energy per kilo fuel than a chemical fuel such as oil or coal does:

You'll find that the spacing of the first energy levels of the Hydrogen atom is usually given in a unit of energy that is called an electronvolt, [eV]. By definition, 1 eV is the amount of energy that an electron gains when you let it travel between two electrodes that have a voltage bias of 1 Volt between them. The larger spacings between the discrete energy levels (belonging to bound states) of the Hydrogen atom, and of molecules in general, is in the range of a few eV. This is the reason why batteries and photovoltaic cells operate at the voltage levels of a few eV. All chemical and biochemical reactions operate in the range of just a few eV too.

The formation of a molecules of Vitamin D, for example, happens to require an amount of energy, namely about 4.2 eV, that is in the upper range in the field of biochemistry and the other molecules of our bodies just don't have that kind of large energy gap available! That's why we need to consume vitamin D or make it using ultraviolet-B sunlight. UV-B photons that carry the required 4.2eV of energy needed to make Vitamin D molecules have a wavelength of about 295nm. We can't see sunlight of such short wavelength. Ultraviolet light is bluer than blue or even violet, which is why it is called ultraviolet. Sunlight scatters of air molecules through Rayleigh scattering and that effect is the stronger the smaller the wavelength. That's why during the day the sky is full of scattered light - which is blue. UV-B light makes it all the way down to the surface essentially only when the sun stands steeper in the sky than about 45 degrees. Else, it's just scattered around and away. So that's why we get enough vitamin D from sunlight only is the summer. In the winter, it's good to have vitamin D supplements. You know, the tablets that taste like dust. Yummy!

Now atomic nuclei also have a Hamiltonian. These nuclei can be excited so that they vibrate (point spectrum) or so that they split apart into non-bound fragments (continuous spectrum). The typical energy gaps in the spectra of nuclei are 1 MeV. That's why you can get roughly on the order of a million times more energy out of a kilo of plutonium than out of a litre of oil. Still, I am not advocating for nuclear energy. Why? The tremendous amount of energy in nuclear fuel also means that unlike a chemical fire that runs out of energy in a few days, radioactive contamination can easily spew radioactive radiation energy for millions of days (depending on the element of course). I flew over Fukushima. It's terrible.

4.4 The spectral theorem for unitary operators

The spectral theorem for self-adjoint operators f translates directly into a spectral theorem for unitary operators, i.e., for operators obeying $U^\dagger = U^{-1}$. This is because there is a bijective map, called the Cayley transform, that maps every self-adjoint operator into a unitary operator and vice versa:

$$\hat{U} := (\hat{f} - i\mathbf{1})(\hat{f} + i\mathbf{1})^{-1} \tag{4.32}$$

Exercise 4.6 *Show that $\hat{U}$ is unitary.*

The Cayley transform is invertible:

$$\hat{f} = -i(\hat{U} - 1)^{-1}(\hat{U} + 1) \tag{4.33}$$

Clearly, in the eigenbasis of $\hat{f}$, where $\hat{f}$ is diagonal, also $\hat{U}$ is diagonal. This means that $\hat{f}$ and $\hat{U}$ possess the same eigenvectors and approximate eigenvectors, with the same multiplicities (i.e. with the same dimension of the eigenspaces). Further, since both operators are simultaneously diagonal, if $r \in \mathbb{R}$ is in the spectrum of $\hat{f}$, then $\alpha = (r - i)(r + i)^{-1}$ (which is a Moebius transform of r), is in the spectrum of $\hat{U}$.

Exercise 4.7 *Show that each α is a complex number on the unit circle of the complex plane, i.e., that its modulus squared is $\alpha\alpha^* = 1$.*

Therefore, the spectral theorem for self-adjoint operators implies a corresponding spectral theorem for unitary operators. They too only possess a point and or a continuous spectrum, with corresponding eigenvectors and approximate eigenvectors.

Remark: You may wonder why we here did we not use exponentiation to obtain a unitary operator for any self-adjoint operator, through $\tilde{U} := \exp(i\hat{f})$. The reason is that, while we could use exponentiation, it has a drawback. The drawback is that the exponentiation map is not invertible because of the periodicity of $\exp(ir)$. Multiple eigenvalues of $\hat{f}$ could map into one eigenvalue for $\tilde{U}$. The Cayley transform is invertible, i.e., it maps each unitary operator uniquely to a self-adjoint operator and vice versa.

The most general spectral theorem is for normal operators: The spectral theorem naturally generalizes beyond self-adjoint and unitary operators, namely to all operators that are called normal. An operator $\hat{Q}$ is called normal if it commutes with its adjoint, i.e., if $[\hat{Q}, \hat{Q}^\dagger] = 0$. For example, the operators a and $a^\dagger$ that obey $[a, a^\dagger] = 1$ are clearly not normal. Indeed, the eigenstates of a that you calculated in Exercise 3.27, the so-called coherent states, can be shown to be not orthogonal to another.

Exercise 4.8 *Assume that $\hat{f}$ is a self-adjoint operator with a purely point spectrum $\{f_n\}$. Further assume that $\hat{U}$ is the Cayley transform of f. Determine whether or not the operator $\hat{Q} := \hat{f} + \hat{U}$ is a normal operator. If not, why not? If yes, what is the spectrum of $\hat{Q}$?*

Chapter 5

The position, momentum and energy representations

5.1 The eigenbasis of the position operator

We had found a representation of the position operator (parametrized by a parameter L that specifies the unit of length) in Eqs.3.42,3.48:

$$\hat{x}_{n,m} = \begin{pmatrix} 0 & \sqrt{1}L & 0 & 0 & \\ \sqrt{1}L & 0 & \sqrt{2}L & 0 & \\ 0 & \sqrt{2}L & 0 & \sqrt{3}L & \\ 0 & 0 & \sqrt{3}L & 0 & \\ & & & & \ddots \end{pmatrix}_{n,m} \tag{5.1}$$

Here, the subscript n, m indicates the matrix element in the $(n+1)$'st row and $(m+1)$'st column, because we will count from 0. We are here obviously dealing with a discrete, i.e., countable, basis. Let us call these basis vectors $|E_n\rangle$ for $n = 0, 1, 2, 3, ...$ Then we have

$$\langle E_n|\hat{x}|E_m\rangle = \hat{x}_{n,m} \tag{5.2}$$

and

$$\hat{x} = \sum_{n,m=0}^{\infty} |E_n\rangle \hat{x}_{n,m} \langle E_m| \tag{5.3}$$

Clearly, $\hat{x}$ is not diagonal in the $\{|E_n\rangle\}$ basis. Now to find a basis in which $\hat{x}$ is diagonal, i.e., to find an eigenbasis of $\hat{x}$, we need to find the solutions to the eigenvalue equation:

$$\hat{x}|x\rangle = x|x\rangle \tag{5.4}$$

To this end, we first act with $\langle E_n|$ from the left to obtain:

$$\langle E_n|\hat{x}|x\rangle = x\langle E_n|x\rangle \tag{5.5}$$

Inserting a resolution of the identity we obtain a concrete eigenvalue equation in a basis:

$$\sum_{m=0}^{\infty} \langle E_n|\hat{x}|E_m\rangle\langle E_m|x\rangle = x\langle E_n|x\rangle \tag{5.6}$$

We know the matrix elements $\langle E_n|\hat{x}|E_m\rangle$ of the position operator, namely

$$\langle E_n|\hat{x}|E_m\rangle = L \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \\ \sqrt{1} & 0 & \sqrt{2} & 0 & 0 & \\ 0 & \sqrt{2} & 0 & \sqrt{3} & 0 & \\ 0 & 0 & \sqrt{3} & 0 & \sqrt{4} & \\ 0 & 0 & 0 & \sqrt{4} & 0 & \\ & & & & & \ddots \end{pmatrix}_{n,m} \tag{5.7}$$

and our task is to determine for which real eigenvalues x there exists a column vector that is an eigenvector. Its components are then the wave function $\langle E_n|x\rangle$ for a position eigenstate $|\hat{x}\rangle$ in the $\{|E_n\rangle\}$ basis. Once we have these wave functions, $\langle E_n|x\rangle$, of position eigenstates, we can write down the position eigenvectors $|x\rangle$ themselves in the basis $\{|E_n\rangle\}$:

$$|x\rangle = \sum_n |E_n\rangle\langle E_n|x\rangle \tag{5.8}$$

Now Eq.5.6 yields an equation for each $n \in \{0, 1, 2, ...$ These equations are

$$L\sqrt{1}\langle E_1|x\rangle = x\langle E_0|x\rangle \tag{5.9}$$

$$L\sqrt{1}\langle E_0|x\rangle + L\sqrt{2}\langle E_2|x\rangle = x\langle E_1|x\rangle \tag{5.10}$$

$$L\sqrt{2}\langle E_1|x\rangle + L\sqrt{3}\langle E_3|x\rangle = x\langle E_2|x\rangle \tag{5.11}$$

and so on. We have arrived at a recursion for the coefficients $\langle E_n|x\rangle$ which gives us a solution, i.e., a full set of coefficients $\langle E_n|x\rangle$ for any choice of x! This is because once we pick an arbitrary value for the first coefficient, $\langle E_0|x\rangle$, we can then calculate the second coefficient, namely $\langle E_1|x\rangle$. From these two we can calculate the third, $\langle E_2|x\rangle$. In this way, iteratively, all $\langle E_n|x\rangle$ are determined and can be calculated. An important conclusion we can draw from this is that no matter what value for x we try, we always find an eigenvector for it! This means that the spectrum of $\hat{x}$ is the real line: $\text{spec}(\hat{x}) = \mathbb{R}$.

The recursion problem can be solved in closed form. Without proof (which is straightforward but tedious) we only state the result:

$$\langle E_n|x\rangle = \frac{e^{-\frac{x^2}{4L^2}}}{\sqrt{\pi^{1/2}2^n(n)!}}\ H_n\left(\frac{x}{\sqrt{2}L}\right) \tag{5.12}$$

Here, H_n is the n'th Hermite polynomial, defined through:

$$H_n(z) \;:=\; (-1)^n\, e^{z^2}\, \frac{d^n}{dz^n}\, e^{-z^2} \tag{5.13}$$

But then, didn't we have for each eigenvalue x a choice in what value we wanted to give to the first coefficient, $\langle E_0|x\rangle$? Yes we did. In the equation Eq.5.12, can we make this choice such that the eigenvectors $|x\rangle$ come out normalizable? Here, let us remember that we just established that the spectrum is the real line and is, therefore, purely continuous. From the spectral theorem, we conclude that the eigenvectors $|\hat{x}\rangle$ are not normalizable. Instead, therefore, the choice of the first coefficient is made for each x such that the resulting eigenvectors come out to be what is called *continuum normalized.* Namely, one can show that with Eq.5.12, we have:

$$\langle x|x'\rangle = \sum_{n=0}^{\infty} \langle x|n\rangle\langle n|x'\rangle = \delta(x-x') \tag{5.14}$$

Therefore, we have now found the wave function $\langle E_n|x\rangle$ of each position eigenstate $|x\rangle$ in the basis $\{|E_n\rangle\}_n$. In fact, we have inadvertently also obtained the position wave functions of the eigenstates of harmonic oscillators! To see this, let us briefly review aspects of the harmonic oscillator.

5.2 The energy eigenbasis of a harmonic oscillator

Following up on Exercise (3.19), we express the harmonic oscillator's Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2 \tag{5.15}$$

in terms of operators $a, a^\dagger$ which obey $[a, a^\dagger] = 1$. In the $\{|E_n\rangle\}$ basis, these operators have the matrix elements given in Eqs.3.39,3.40. We make again the ansatz:

$$\hat{x} = L(a^\dagger + a), \qquad \hat{p} = iK(a^\dagger - a) \quad \text{with} \quad L, K \in \mathbb{R} \tag{5.16}$$

So far, the parameters L and K are still arbitrary real numbers. The ansatz needs to solve the hermiticity conditions $\hat{x} = \hat{x}^\dagger$ and $\hat{p} = \hat{p}^\dagger$ and it obviously does. A little calculation shows that this ansatz also succeeds in solving the canonical commutation relations $[\hat{x}, \hat{p}] = i1$, namely if we choose $K = (2L)^{-1}$ which we will henceforth do. Now, only L is still left as a free parameter in our ansatz. (For now, let us not worry about the time evolution and consider only one arbitrary point in time.) Therefore, we have found a valid way of writing $\hat{x}$ and $\hat{p}$.

At this point we are still free to choose any real value for L and, in principle, any value of L is as good as any other. Let us now express the Hamiltonian in terms of the $a, a^\dagger$ operators:

$$\hat{H} = \frac{-1}{2m\ 4L^2}\left((a^\dagger)^2 - a^\dagger a - aa^\dagger + a^2\right) + \frac{m\omega^2 L^2}{2}\left((a^\dagger)^2 + a^\dagger a + aa^\dagger + a^2\right) \tag{5.17}$$

We notice that we can make the $(a^\dagger)^2$ and a^2 terms cancel by choosing L so that the prefactors match, i.e., so that

$$\frac{1}{8mL^2} = \frac{m\omega^2 L^2}{2} \tag{5.18}$$

which is solved by:

$$L = (2m\omega)^{-1/2} \tag{5.19}$$

We conclude that for every harmonic oscillator, when expressing the $\hat{x}$ and $\hat{p}$ operators in terms of a and $a^\dagger$, there is an especially convenient choice of L :

$$\hat{x} = \frac{1}{\sqrt{2m\omega}}(a^\dagger + a), \qquad \hat{p} = i\sqrt{\frac{m\omega}{2}}(a^\dagger - a) \tag{5.20}$$

This is because with this choice of L, the form of the oscillator's Hamiltonian, after a short calculation, simplifies to:

$$\hat{H} = \hbar\omega(a^\dagger a + 1/2) \tag{5.21}$$

How does this help us to find the spectrum of $\hat{H}$? Recall that we know the matrix elements $a_{n,m} = \langle E_n|a|E_m\rangle$ and $a^\dagger_{n,m} = \langle E_n|a^\dagger|E_m\rangle$ of the operators a and $a^\dagger$ in the basis $\{|E_0\rangle, |E_1\rangle, |E_2\rangle, |E_3\rangle, ...\}$, namely:

$$a_{n,m} = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & 0 & \\ 0 & 0 & \sqrt{2} & 0 & 0 & \\ 0 & 0 & 0 & \sqrt{3} & 0 & \\ 0 & 0 & 0 & 0 & \sqrt{4} & \\ 0 & 0 & 0 & 0 & 0 & \\ & & & & & \ddots \end{pmatrix}_{n,m} \tag{5.22}$$

The hermitean conjugate is:

$$a^\dagger_{n,m} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \\ \sqrt{1} & 0 & 0 & 0 & 0 & \\ 0 & \sqrt{2} & 0 & 0 & 0 & \\ 0 & 0 & \sqrt{3} & 0 & 0 & \\ 0 & 0 & 0 & \sqrt{4} & 0 & \\ & & & & & \ddots \end{pmatrix}_{n,m} \tag{5.23}$$

Clearly then, the matrix elements of the so-called number operator $\hat{N} := a^\dagger a$ in the basis $\{|E_0\rangle, |E_1\rangle, ...\}$ are

$$\hat{N}_{n,m} = (a^\dagger a)_{n,m} = n\ \delta_{n,m} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \\ 0 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 2 & 0 & 0 & \\ 0 & 0 & 0 & 3 & 0 & \\ 0 & 0 & 0 & 0 & 4 & \\ & & & & & \ddots \end{pmatrix}_{n,m} \tag{5.24}$$

Clearly, $\hat{N}$ is a self-adjoint operator, $\hat{N}^\dagger = \hat{N}$, since it is diagonal in the basis $\{|E_0\rangle, |E_1\rangle, ...\}$ and is, therefore, diagonalizable. Since

$$\hat{H} = \hbar\omega(\hat{N} + 1/2) \tag{5.25}$$

we have that the Hamiltonian is diagonal in the basis $\{|E_0\rangle, |E_1\rangle, ...\}$, i.e., that $\hat{H}|E_n\rangle = E_n|E_n\rangle$, and that its eigenvalues are $E_n = \hbar\omega(n + 1/2)$ for $n \in \{0, 1, 2, ...\}$. The spectrum of the Hamiltonian of a harmonic oscillator is, therefore, purely a point spectrum.

Therefore, this Hamiltonian reads in its eigenbasis:

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2}{2}\hat{x}^2 = \sum_{n=0}^{\infty} E_n |E_n\rangle\langle E_n| \tag{5.26}$$

The state $|E_n\rangle$ can therefore also be interpreted the n'th excited state of a harmonic oscillator with frequency ω. Now what is the position wave function of the n'th eigenstate of this harmonic oscillator? By definition it is given by $\langle x|E_n\rangle$. But we have $\langle x|E_n\rangle = (\langle E_n|x\rangle)^* = \langle E_n|x\rangle$ because $\langle E_n|x\rangle$ is real. This means that we have inadvertently also obtained the position wave functions of the eigenstates of the harmonic oscillator!

Exercise 5.1 *Consider the harmonic oscillator of above, choose* $L = (2m\omega)^{-1/2}$ *and assume that the oscillator system is in this state:*

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|E_0\rangle - |E_1\rangle) \tag{5.27}$$

Calculate $\bar{H}$ *and* $\bar{x}$ *and the position wave function* $\psi(x) := \langle x|\psi\rangle$.

5.3 The position representation

We established that the spectrum of $\hat{x}$ is $\mathbb{R}$, with each of these approximate eigenvalues coming with a corresponding approximate (i.e., non-normalizable) eigenvector $\{|x\rangle\}_{x\in\mathbb{R}}$. This means that we have a new resolution of the identity[1]:

$$\mathbf{1} = \int_{\mathbb{R}} |x\rangle\langle x| \, dx \tag{5.28}$$

It allows us to express arbitrary states in the eigenbasis of $\hat{x}$, i.e., we can now calculate:

$$|\psi\rangle = \int_{\mathbb{R}} |x\rangle\langle x|\psi\rangle \, dx = \int_{\mathbb{R}} \psi(x) \, |x\rangle \, dx \tag{5.29}$$

[1] We are here using physicists' notation. The functional analytic details of operators with continuous spectra will be in an Appendix.

Here, $\psi(x) = \langle x|\psi\rangle$ is the position wave function of the state $|\psi\rangle$. Further, we can now calculate the scalar product of states in the position basis:

$$\langle\psi|\phi\rangle = \langle\psi|\mathbf{1}|\phi\rangle = \int_{\mathbb{R}} \langle\psi|x\rangle\langle x|\phi\rangle \; dx = \int_{\mathbb{R}} \psi(x)^* \phi(x) \; dx \tag{5.30}$$

For example, the normalization condition of a state $|\psi\rangle$ takes the form:

$$1 = \langle\psi|\psi\rangle = \int_{\mathbb{R}} \psi(x)^* \psi(x) \; dx \tag{5.31}$$

Operators can be expanded in the position eigenbasis by inserting resolutions of the identity:

$$\hat{f} = \mathbf{1}\hat{f}\mathbf{1} = \int_{\mathbb{R}}\int_{\mathbb{R}} |x\rangle\langle x|\hat{f}|x'\rangle\langle x'| \; dx \; dx' = \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,x')|x\rangle\langle x'| \; dx \; dx' \tag{5.32}$$

Here, the $f(x,x') = \langle x|\hat{f}|x'\rangle$ are called the matrix elements of $\hat{f}$ in the position representation. Of course, since x is a continuous rather than a discrete index for the matrix elements, we are not really dealing with a conventional matrix here. While the term matrix element is commonly used in the physics literature, the alternative terminology in proper mathematical language is to say that $f(x,x')$ is the *integral kernel* of $\hat{f}$ in the position eigenbasis.

Definition: If $\hat{f}$ is an operator, and $\{|\lambda\rangle\}$ is a continuous basis (such as the position eigenbasis) then we call $f(\lambda,\lambda') := \langle\lambda|\hat{f}|\lambda'\rangle$ the matrix elements of $\hat{f}$ in the $\{|\lambda\rangle\}$ basis. Of course, $f(\lambda,\lambda')$, which could also be written as $f_{\lambda,\lambda'}$, is a matrix whose two indices take continuous values. When such a 'matrix' acts on a vector then the common index needs to be integrated instead of summed: $\langle\lambda|\hat{f}|\psi\rangle = \int f(\lambda,\lambda')\psi(\lambda')d\lambda'$. For this reason, $f(\lambda,\lambda')$ is also called the integral kernel of $\hat{f}$ in the $\{|\lambda\rangle\}$ basis.

For example, for the operator $\hat{f} = \hat{x}$ we find that its integral kernel,

$$\hat{x}(x,x') = \langle x|\hat{x}|x'\rangle = x\langle x|x'\rangle = x\delta(x-x'), \tag{5.33}$$

is diagonal in the sense that the kernel vanishes off the diagonal, i.e., for $x \neq x'$. Let us also express the position operator in the position basis:

$$\hat{x} = \hat{x}1 = \int_{\mathbb{R}} \hat{x} \; |x\rangle\langle x| \; dx = \int_{\mathbb{R}} x \; |x\rangle\langle x| \; dx \tag{5.34}$$

Let us now calculate the momentum operator's matrix elements in the position representation:

$$\hat{p}(x,x') = \langle x|\hat{p}|x'\rangle = \sum_{n,m} \langle x|E_n\rangle\langle E_n|\hat{p}|E_m\rangle\langle E_m|x'\rangle = i\hbar\frac{d}{dx'}\delta(x-x') \tag{5.35}$$

Here, we inserted resolutions of the identity in terms of the eigenbasis $\{|E_n\rangle\}$ of a harmonic oscillator. This is useful because we already know the matrix elements $\langle n|\hat{p}|m\rangle$ of $\hat{p}$ explicitly from Eqs.5.20,5.22,5.23 and we know the coefficients $\langle E_n|x\rangle$ from Eq.5.12. The last step, i.e., to work out the sums over n and m is quite tedious and we skip the details here.

But what is the derivative of the Dirac delta? The Dirac delta itself is defined to obey:

$$\int_{\mathbb{R}} f(x)\delta(x-x')\ dx = f(x') \tag{5.36}$$

Now for the derivative of the Dirac delta we obtain via integration by parts:

$$\int_{\mathbb{R}} f(x)\frac{d}{dx}\delta(x-x')\ dx = -\int_{\mathbb{R}}\left(\frac{d}{dx}f(x)\right)\ \delta(x-x')\ dx = -\frac{df}{dx}(x') \tag{5.37}$$

We now obtain the operator $\hat{p}$ in the position representation:

$$\hat{p} = \int_{\mathbb{R}}\int_{\mathbb{R}} i\hbar\left(\frac{d}{dx'}\delta(x-x')\right)|x\rangle\langle x'|\ dx\ dx' \tag{5.38}$$

Remember that already in Eqs.3.48,3.49, we had found a representation of the operators $\hat{x}$ and $\hat{p}$ as infinite-by-infinite matrices. Now, with Eqs.5.33,5.35, we have found a representation of $\hat{x}$ and $\hat{p}$ in a continuous basis, namely in the eigenbasis of the position operator.

5.4 Shorthand notation for operators acting on wave functions

Let us consider an operator Q that maps a vector $|\psi\rangle$ into a vector $|\phi\rangle = Q|\psi\rangle$. Since $|\psi\rangle$ and $|\phi\rangle$ both possess position wave functions, namely $\psi(x)$ and $\phi(x)$, we have that Q also induces a map of wave functions into wave functions, namely $\hat{Q} : \psi(x) \to \phi(x)$. This is conveniently described by an important shorthand notation:

Definition: We denote the induced action of an operator on a wave function by a dot. For example, an operator $\hat{Q}$ acts on a state $|\psi\rangle$ by mapping it into a state $|\phi\rangle$, i.e., we have $|\phi\rangle = \hat{Q}|\psi\rangle$. If we choose some basis $\{|\lambda\rangle\}$, this also means that $\hat{Q}$ maps the wave function $\psi(\lambda)$ into the wave function $\phi(\lambda)$. We write this as:

$$\phi(\lambda) = \hat{Q}.\psi(\lambda) \tag{5.39}$$

For example, in the case of the position operator and the position eigenbasis:

$$\phi(x) = \langle x|\phi\rangle = \langle x|\hat{x}|\psi\rangle = x\langle x|\psi\rangle = x\psi(x) \tag{5.40}$$

In the shorthand notation, we can therefore write that the action of the position operator on position wave functions is:

$$\hat{x}.\psi(x) = x\psi(x) \tag{5.41}$$

What then is the action of $\hat{p}$ on position wave functions? Abstractly, we have $|\varphi\rangle = \hat{p}|\psi\rangle$. In the position eigenbasis, this is:

$$\begin{aligned}
\phi(x) &= \langle x|\phi\rangle = \langle x|\hat{p}|\psi\rangle && (5.42)\\
&= \int_{\mathbb{R}} \langle x|\hat{p}|x'\rangle\langle x'|\psi\rangle \; dx' && (5.43)\\
&= \int_{\mathbb{R}} i\hbar \left(\frac{d}{dx'}\delta(x-x')\right)\psi(x') \; dx' && (5.44)\\
&= -i\hbar \int_{\mathbb{R}} \delta(x-x') \; \frac{d}{dx'}\psi(x') \; dx' + i\hbar \left[\delta(x-x')\psi(x')\right]_{x'=-\infty}^{x'=\infty} && (5.45)\\
&= -i\hbar \frac{d}{dx}\psi(x) && (5.46)
\end{aligned}$$

This means that in the shorthand notation we can write:

$$\hat{p}.\psi(x) = -i\hbar \frac{d}{dx}\psi(x) \tag{5.47}$$

Exercise 5.2 *Verify the canonical commutation relation in the position representation, i.e., verify that, for all (differentiable) wave functions $\psi(x)$:*

$$(\hat{x}\hat{p} - \hat{p}\hat{x} - i\hbar 1).\psi(x) = 0 \tag{5.48}$$

5.5 The momentum representation

Similar to how we diagonalized the position operator, we could now calculate also the momentum operator's eigenbasis. Namely, we could diagonalize the momentum operator's matrix representation Eq.3.49 in the $\{|E_n\rangle\}_{n=0}^{\infty}$ basis of a harmonic oscillator. Instead, it will be more convenient to calculate the momentum operator's spectrum and eigenbasis by diagonalizing it in the position representation. Concretely, let us solve the eigenvector equation

$$\hat{p}|p\rangle = p|p\rangle \tag{5.49}$$

in the position basis, i.e., with the momentum eigenvector $|p\rangle$ written in the form:

$$|p\rangle = \mathbf{1}|p\rangle = \int_{\mathbb{R}} |x\rangle\langle x|p\rangle \; dx \tag{5.50}$$

Here, $\langle x|p\rangle$ is the wave function of the momentum eigenvector $|p\rangle$ to eigenvalue p in the position basis. If we can calculate $\langle x|p\rangle$, then we have the momentum eigenvector $|p\rangle$ through Eq.5.50. From Eq.5.49 we obtain:

$$\langle x|\hat{p}\mathbf{1}|p\rangle = p\ \langle x|p\rangle \tag{5.51}$$

$$\int_{\mathbb{R}} \langle x|\hat{p}|x'\rangle\langle x'|p\rangle\ dx' = p\ \langle x|p\rangle \tag{5.52}$$

$$\int_{\mathbb{R}} i\hbar \left(\frac{d}{dx'}\delta(x-x')\right) \langle x'|p\rangle\ dx' = p\ \langle x|p\rangle \tag{5.53}$$

$$\int_{\mathbb{R}} -i\hbar\delta(x-x')\frac{d}{dx'}\langle x'|p\rangle\ dx' = p\ \langle x|p\rangle \tag{5.54}$$

Therefore, the function $\langle x|p\rangle$ must obey:

$$-i\hbar\ \frac{d}{dx}\ \langle x|p\rangle = p\ \langle x|p\rangle \tag{5.55}$$

Notice that we could have found immediately that Eq.5.49 translates into Eq.5.55 in the position basis by using the shorthand notation Eq.5.47. Clearly, the solution to Eq.5.55 is:

$$\langle x|p\rangle = N\ e^{\frac{ixp}{\hbar}} \tag{5.56}$$

Since, therefore, we are able to find an eigenvector $|p\rangle$ for each value $p \in \mathbb{R}$, we can conclude that $\mathrm{spec}(\hat{p}) = \mathbb{R}$. The spectrum is, therefore, continuous and the approximate eigenfunctions $\langle x|p\rangle$ therefore require continuum normalization:

$$\begin{aligned}
\delta(p-p') &= \langle p|p'\rangle && (5.57)\\
&= \int_{\mathbb{R}} \langle p|x\rangle\langle x|p'\rangle\ dx && (5.58)\\
&= \int_{\mathbb{R}} N^*\ e^{\frac{-ixp}{\hbar}}\ N\ e^{\frac{ixp'}{\hbar}}\ dx && (5.59)\\
&= N^*N \int_{\mathbb{R}} e^{\frac{ix}{\hbar}(p-p')}\ dx && (5.60)
\end{aligned}$$

At this point we use the fact that

$$\int_{\mathbb{R}} e^{iab}\ da = 2\pi\ \delta(b) \tag{5.61}$$

to conclude that $N = (2\pi\hbar)^{-1/2}$. We obtain the continuum normalized position eigenfunctions of the momentum eigenvectors:

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}\ e^{\frac{ixp}{\hbar}} \tag{5.62}$$

Notice that this also yields the momentum wave functions of the position eigenvectors:

$$\langle p|x\rangle = \langle x|p\rangle^* = \frac{1}{\sqrt{2\pi\hbar}}\, e^{\frac{-ixp}{\hbar}} \tag{5.63}$$

We obtain, analogously to the calculation for $\hat{x}$ before, a resolution of the identity and a spectral resolution of the momentum operator:

$$\mathbf{1} = \int_{\mathbb{R}} |p\rangle\langle p|\ dp, \qquad \text{and} \qquad \hat{p} = \int_{\mathbb{R}} p\ |p\rangle\langle p|\ dp \tag{5.64}$$

The momentum wave function of a state $|\psi\rangle$ is then obtained from:

$$|\psi\rangle = \mathbf{1}|\psi\rangle = \int_{\mathbb{R}} |p\rangle\langle p|\psi\rangle\ dp = \int_{\mathbb{R}} \tilde{\psi}(p)\ |p\rangle\ dp \tag{5.65}$$

Here, $\tilde{\psi}(p) := \langle p|\psi\rangle$ is a common notation for the momentum wave function. The scalar product of two states is now calculated through

$$\langle \phi|\psi\rangle = \int_{\mathbb{R}} \langle \phi|p\rangle\langle p|\psi\rangle\ dp = \int_{\mathbb{R}} \tilde{\phi}(p)^*\tilde{\psi}(p)\ dp \tag{5.66}$$

We can now derive how the momentum wave function $\tilde{\psi}(p)$ can be expressed in terms of the position wave function $\psi(x)$:

$$\begin{aligned} \tilde{\psi}(p) &= \langle p|\psi\rangle && (5.67)\\ &= \int_{-\infty}^{\infty} \langle p|x\rangle\langle x|\psi\rangle dx && (5.68)\\ &= \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x)\ e^{\frac{-ixp}{\hbar}}\ dx && (5.69) \end{aligned}$$

We conclude that the position and momentum representations (i.e., bases) are related by Fourier transformation. Similarly, as we saw earlier, the position eigenbasis and the eigenbasis of the Hamiltonian of a harmonic oscillator are related by a Hermite transformation. Analogously, any two choices of (appropriately orthonormalized) bases are related by a transformation that implements the unitary change of basis.

Exercise 5.3 *Derive the action of $\hat{x}$ and $\hat{p}$ on momentum wave functions, i.e., derive the short hand notation $\hat{x}.\tilde{\psi}(p) = ?\tilde{\psi}(p)$ and $\hat{p}.\tilde{\psi}(p) = ?\tilde{\psi}(p)$ analogously to how we derived the short hand notation for the position representation.*

5.6 Energy representations: bound states and scattering states

Consider an arbitrary quantum system with Hamiltonian $\hat{H}$. Since $\hat{H}$ is normally self-adjoint, it possesses an eigenbasis, the so-called energy eigenbasis. Let us denote

the normalizable and nonnormalizable eigenvectors of $\hat{H}$ by $|E_\mu\rangle$ where μ is an index that runs through a continuous, a discrete or a mixed index set, as the case may be. Remember that, the Hamiltonians of some systems possess a purely discrete spectrum, as is the case for the harmonic oscillator. The Hamiltonians of other systems are purely continuous, as for example, that of the free particle. The spectrum can also consist of a discrete and a continuous spectrum, as, for example, the spectrum of the Hamiltonian of the hydrogen atom. What does this mathematical property of the spectrum being discrete or continuous make correspond to, physically?

To see the physical difference, let us recall that eigenvalues E_n of $\hat{H}$ that are in the point spectrum $E_n \in \text{spec}_{discr.}(\hat{H})$ come with eigenvectors $|E_n\rangle$ that are normalizable and that, in contrast, the approximate eigenvalues $E \in \text{spec}_{cont.}(\hat{H})$ of the continuous spectrum, the so-called approximate eigenvalues, come with non-normalizable approximate eigenvectors $|E\rangle$. Let us consider these normalizable and non-normalizable vectors in the position eigenbasis, i.e., $\psi_{E_n}(x) = \langle x|E_n\rangle$ and $\psi_E(x) = \langle x|E\rangle$ respectively.

Since each state $|E_n\rangle$ is normalizable, $\langle E_n|E_n\rangle = 1$, we have from Eq.5.31 that its position wave function is square integrable:

$$\int_{-\infty}^{\infty} \psi_{E_n}(x)^* \psi_{E_n}(x)\, dx = 1 \tag{5.70}$$

This implies that the wave function must decay quickly towards $\pm\infty$. For this reason, particles described by such a wave function are fairly localized and we say that they are in a bound state. This could be, for example, an electron bound to a proton, described by one of Hydrogen's orbital wave functions. In contrast, the wave functions $\psi_E(x)$ belonging to states of the continuous spectrum are not normalizable and describe particles that are entirely delocalized, much like plane waves. These eigenstates of the Hamiltonian describe particles that come in from infinity, interact with our system, perhaps change direction a bit and then go away out to infinity. The states of the continuous spectrum of the Hamiltonian are, therefore, also called scattering states. The Hamiltonians of electrons that orbit the nuclei of an atom or a molecule always possess both a discrete energy spectrum and above it also a continuous energy spectrum that reaches up to infinite energies. This is because the electrons can be bound to the nuclei at discrete energy levels, until at sufficiently high energies the nuclei can no longer hold on to the particle and the particle merely scatters off the nuclei.

There is a classical analog in the solar system. Comets are in bound states around the sun. In contrast, the extra-solar rock called Oumuamua, for example, had much more kinetic energy than a comet and it only scattered off our solar system when it visited our solar system in 2017. Similarly, the hydrogen atom's well known discrete energy levels constitute only the point spectrum. The Hamiltonian of an electron interacting with a proton also possesses a continuous spectrum which ranges from 0 to infinity.

Chapter 6

Continuous versus discrete in quantum mechanics

6.1 It takes only a countable infinite amount of information to specify a quantum state

Let us review how it can be that a state $|\psi\rangle$ can be specified by both, an uncountable amount or also by a countable amount of coefficients. Namely, $|\psi\rangle$ can be specified through uncountably many numbers $\psi(x) = \langle x|\psi\rangle$, where x runs through all the uncountably infinitely many real numbers, or also through the countably infinitely many numbers $\psi_n = \langle n|\psi\rangle$ where n runs through a countably infinite number of integers.

The answer is subtle. The countably many coefficients ψ_n really do suffice to specify $|\psi\rangle$ because $|\psi\rangle = \sum_{n=1}^{\infty} \psi_n |n\rangle$. Therefore, when we instead describe $|\psi\rangle$ through the uncountably infinitely many data $\psi(x) \ \ \forall x \in \mathbb{R}$ this description must possess a large redundancy. Indeed, there are many wave functions that describe the same state!

Consider two wave functions $\psi(x)$ and $\phi(x)$ which are exactly the same for all $x \in \mathbb{R}$, except at a finite number of points $x_1, x_2, ..., x_N$, where they differ, for example, $\phi(x_2) = 77$ but $\psi(x_2) = -3.2$. Any such pairs of wave functions $\psi(x)$ and $\phi(x)$ are actually physically equivalent in the sense that the predictions come out the same. The reason is that when we calculate an expectation value with either $\psi(x)$ or $\phi(x)$ the prediction always comes out to be the same. Consider, for example, the position expectation value: $\bar{x} = \int \psi^*(x) x \psi(x) = \int \phi^*(x) x \phi(x)$. The reason is that the value of an integral is insensitive to changing the value of its integrand function at a finite number of points.

Why? It's because when you integrate, i.e., when you calculate the area under a curve, then if you make the integrand function have a spike at one isolated point, this does not have any effect on the area under the curve. A spike at a single point doesn't affect the area under the curve because such a spike is just a line and lines don't have area. If you are interested in the details of the integration theory behind this, look up

the notion of Lebesgue integration.

So in conclusion, since we can add isolated spikes everywhere on the x axis, each position wave function has uncountably infinitely many brother and sister position wave functions that differ only at individual points and that are indistinguishable under the integral and are, therefore, also indistinguishable in the calculation of any expectation value. So they are physically equivalent.

Each individual state $|\psi\rangle$, possesses, therefore, an equivalence class of uncountably infinitely many position wave functions. What is specified by $|\psi\rangle$ is merely the equivalence class, but not any particular wave function in it. This can pose a problem when wanting to differentiate the wave function. That's because wherever the wave function is not differentiable because of one of those spikes, we can only differentiate before and after the spike but there is no derivative at the spike. That's not a problem under the integrals because it happens only at isolated points, but it's tedious. Therefore, in practice, we always choose the most differentiable member of the equivalence class. I know, all these things are often not spelled out in textbooks, as many authors just follow the herd. But in this course we are not afraid to dig somewhat deeper into what's really going on.

So the redundancy in the form of wave functions grouping into large equivalence classes for each state is why it is that a(n) (equivalence class of) position wave function $\psi(x)$ can be specified by just giving countably many data, such as the data $\{\psi_n\}$ for $n = 1, 2, ...$ Namely, by specifying, for example, the ψ_n we obtain $|\psi\rangle = \sum_{n=1}^{\infty} \psi_n |n\rangle$ and from it we obtain a $\psi(x) = \langle x|\psi\rangle$. Let us remember that it is not the unique $\psi(x)$ that we obtain in this way. It's just the differentiable member of a whole equivalence class of wave functions, such as $\phi(x)$ that would describe the exact same physics.

6.2 Stieltjes integration explains the Dirac delta

Definition (Stieltjes integral): The Stieltjes integral

$$\int_a^b f(x)\ dm(x) \tag{6.1}$$

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

$$\int_a^b f(x)\ dm(x) := \lim_{\epsilon \to 0} \sum_i f(\tilde{x}_i)\ (m(x_{i+1}) - m(x_i)) \tag{6.2}$$

Here, $f(\tilde{x}_i)$ is f evaluated at an arbitrary point in the i'th interval: $\tilde{x}_i \in [x_i, x_{i+1}]$ and the limit is the limit in which the maximum spacing, ϵ occurring in a sequence

of partitionings goes to zero. Notice that if the integrator function is constant, then the integral vanishes. Also, if we add any constant to the integrator function $m(x)$, this makes no difference to the integral because only differences of its values occur in Eq.6.2.

Now how does the Stieltjes integral relate to ordinary integration?

Proposition: Assume that the integrator function is differentiable and that the derivative $m'(x)$ of $m(x)$ is continuous. Then the Stieltjes integral can be expressed in terms of the ordinary integral:

$$\int_a^b f(x)\ dm(x) = \int_a^b f(x)\ \frac{dm(x)}{dx}\ dx \tag{6.3}$$

So in particular, if we choose the integrator function to be $m(x) = x$ then we have $dm/dx = 1$ and we recover the ordinary integral over f.

But what then is the advantage of the Stieltjes integration method? The key advantage is that we can do more types of integrals now because the integrator function m need not have a continuous derivative! For example, consider the case of m being a Heaviside step function:

$$m(x) = \theta(x) = \begin{cases} 0 \ \text{ if } \ x < 0 \\ 1/2 \ \text{ if } \ x = 0 \\ 1 \ \text{ if } \ x > 0 \end{cases} \tag{6.4}$$

Assuming that $a < 0 < b$ and applying the definition Eq.6.2 shows that the Stieltjes integral in this case yields:

$$\int_a^b f(x)\ dm(x) = f(0) \tag{6.5}$$

For comparison, the Dirac delta is defined to be the mathematical object that obeys:

$$\int_a^b f(x)\ \delta(x)dx = f(0) \tag{6.6}$$

Comparing with Eq.6.3, we now see that the Dirac delta behaves like the derivative of the Heaviside function

$$\frac{d}{dx}\ \theta(x) = \delta(x) \tag{6.7}$$

This equation is commonly stated but let us keep in mind that, strictly speaking, this equation holds when functions are being integrated over using the Stieltjes integration method.

Exercise 6.1 *Explain Eq.6.5 using a sketch of the plot of a function and a partitioning of the integration interval.*

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.

$$\int_a^b f(x)\ dm(x) = \sum_i f(x_i) \tag{6.8}$$

More generally, by using an integrator function that is in some parts of the real line smooth and in some parts of the real line it has steps of various sizes, one single Stieltjes integral can express arbitrarily complicated combinations of sums and Riemann integrals.

Exercise 6.2 *Sketch the graph of 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 $x = 6$.*

Let us recall why we here introduce the Stieltjes integral. It is because it allows us to avoid using the Dirac delta. Of course, we could bring back the Dirac delta. For example, using the Dirac Delta, we could rewrite Eq.6.8:

$$\int_a^b f(x)\ dm(x) = \int_a^b f(x)\ \frac{dm(x)}{dx}\ dx = \int_a^b f(x)\ \sum_i \delta(x - x_i)\ dx = \sum_i f(x_i) \tag{6.9}$$

In practice, working with the Dirac delta is indeed convenient, even though it is not a function but is instead what is called a distribution[1]. The reason why we need a way to work without using the Dirac delta is that when it comes to integrating not over functions but integrating over operators, then we need to be able to work with operator-valued functions, such as $|x\rangle\langle x|$ rather than working with operator-valued distributions such as $\delta(x)$. The problem with "operator" such as $|x\rangle\langle x|$ is that they are not really operators because, for example, we cannot multiply them onto themselves: $|x\rangle\langle x|x\rangle\langle x| = \delta(0)|x\rangle\langle x|$ is undefined.

The Dirac delta is a distributional generalization of a number-valued function. But we would here need distributional generalizations also of vector- and operator-valued functions too, because this would come up in resolutions of the identity and in spectral representations of observables. Even if in practice one works with operator-valued distributions as if there were no subtlety, it is good to know, at least in principle, how things can be put on a solid footing by using the Stieltjes integral.

[1] A distribution such as the Dirac delta can be viewed as a continuous linear map from a specified space of functions to numbers. For example, the Dirac delta in $\int f(x)\ \delta(x - x_0)\ dx$ maps the function f linearly into the number $f(x_0)$. For details, see, e.g., the appendix on distributions in Messiah's book on quantum mechanics. Distributions can also be viewed as elements of a type of Hilbert space called a Sobolev space, in which the inner product involves derivatives.

6.3 The Stieltjes integral explains operators such as $|x\rangle\langle x|$.

In order to make sense of expressions such as $|x\rangle\langle x|$, let us first define the notion of a partial resolution of the identity:

Definition: Assume that $\hat{f}$ is a self-adjoint operator whose spectrum is purely continuous and consists of the entire real line. Then we define the partial resolution of the identity in the eigenbasis of $\hat{f}$ in this way:

$$\hat{E}(f) := \int_{-\infty}^{f} |\tilde{f}\rangle\langle\tilde{f}| \; d\tilde{f} \tag{6.10}$$

We recover the full resolution of the identity as:

$$\mathbf{1} = \hat{E}(\infty) = \lim_{f\to\infty} \int_{-\infty}^{f} |\tilde{f}\rangle\langle\tilde{f}| \; d\tilde{f} \tag{6.11}$$

The partial resolution of the identity in Eq.6.10 has the advantage that we can differentiate it with respect to the integration boundary:

$$\frac{d\hat{E}(f)}{df} = |f\rangle\langle f| \tag{6.12}$$

So now we have found an expression for the object that we are interested in: $|x\rangle\langle x|$. How does this help? Let us rewrite the resolution of the identity:

$$\mathbf{1} = \int_{\mathbb{R}} |f\rangle\langle f| = \int_{\mathbb{R}} \frac{d\hat{E}(f)}{df} \; df \tag{6.13}$$

From this, we now obtain, in the Stieltjes sense:

$$\mathbf{1} = \int_{\mathbb{R}} d\hat{E}(f) \tag{6.14}$$

And similarly:

$$\hat{f} = \int_{\mathbb{R}} f \; d\hat{E}(f) \tag{6.15}$$

Why are Eqs.6.14,6.15 useful? Remember that, in keeping with physics notation, we used expressions such as $|f\rangle\langle f|$ which are actually ill defined if $|f\rangle$ is an improper or 'approximate' eigenvector. For example, its square is infinite: $|f\rangle\langle f|f\rangle\langle f| = \delta(0)\,|f\rangle\langle f|$.

It would seem that so far not much is gained. Doesn't Eq.6.14 still make use of $|f\rangle\langle f|$? It does not! Let us look at the definition of the Stieltjes integral in Eq.6.2 again. The role of $m(x)$ there is played here by $\hat{E}(f)$. Notice that in Eq.6.2 only evaluations of $m(x)$ occur and that there is no need to evaluate its derivative $dm(x)/dx$ ever. Similarly

here, in Eqs.6.14,6.15, only $\hat{E}(f)$ occurs but not anymore its derivative (which would be $|f\rangle\langle f|$).

And $\hat{E}(f)$ is a perfectly well-behaved operator! One can show that for every value of f it is a projector. For this reason, Eqs.6.14,6.15 are at the heart of the rigorous mathematical formulations and applications of the spectral theorem.

In fact, while in the current section we only considered the case where the spectrum of $\hat{f}$ is the real line, the spectrum of a self-adjoint operator can of course contain arbitrary continuous and discrete parts. The Stieltjes integral takes care of this automatically, i.e., we no longer have to use a mix of sums and integrals. All we need is the Stieltjes integrals Eqs.6.14,6.15 over the full real line. In them, the summations of the point spectrum, see, e.g., Eqs.4.30,4.31, arise, as usual for Stieltjes integrals, through discontinuities in the integrator function $\hat{E}(f)$. The continuous spectrum is where the integrator function changes continuously and on stretches of $\mathbb{R}$ where there is no continuous or point spectrum, the integrator function $\hat{E}(f)$ is constant.

Chapter 7

Uncertainty principles

7.1 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 $\bar{f}^2 \neq \overline{f^2}$. 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 $(49^2 + 50^2 + 51^2)/3\ m^2 = (2500 + 2/3)m^2$.

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 $\bar{f}^2 \neq \overline{f^2}$. 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 $\bar{f}^2 = \overline{f^2}$.

The extent to which the equation $\bar{f}^2 = \overline{f^2}$ 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:

$$\begin{aligned}
(\Delta f(t))^2 &= \overline{(\hat{f}(t) - \overline{f}(t))^2} && (7.1)\\
&= \langle\psi| \left(\hat{f}(t) - \langle\psi|\hat{f}(t)|\psi\rangle\right)^2 |\psi\rangle && (7.2)\\
&= \langle\psi|\hat{f}^2(t)|\psi\rangle - \langle\psi|\hat{f}|\psi\rangle^2 && (7.3)\\
&= \overline{f(t)^2} - \overline{f(t)}^2 && (7.4)
\end{aligned}$$

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:

$$\Delta f(t)\ \Delta g(t)\ \geq\ \frac{1}{2}|\langle\psi|[\hat{f}(t), \hat{g}(t)]|\psi\rangle| \tag{7.5}$$

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

$$\Delta x(t)\ \Delta p(t)\ \geq\ \frac{1}{2}|\langle\psi|i\hbar|\psi\rangle| = \frac{\hbar}{2} \tag{7.6}$$

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:

$$\Delta x_i^{(r)}(t)\ \Delta p_j^{(s)}(t)\ \geq\ \frac{\hbar}{2}\ \delta_{i,j}\delta_{r,s} \tag{7.7}$$

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

$$|\phi\rangle = \left(\hat{f} - \bar{f}\mathbf{1} + \mathrm{i}\alpha(\hat{\mathrm{g}} - \bar{\mathrm{g}}\mathbf{1})\right)|\psi\rangle \tag{7.8}$$

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

$$\langle\psi|\left((\hat{f} - \bar{f}\mathbf{1}) - \mathrm{i}\alpha(\hat{\mathrm{g}} - \bar{\mathrm{g}}\mathbf{1})\right)\left((\hat{f} - \bar{f}\mathbf{1}) + \mathrm{i}\alpha(\hat{\mathrm{g}} - \bar{\mathrm{g}}\mathbf{1})\right)|\psi\rangle \geq 0 \tag{7.9}$$

Thus:

$$\langle\psi|(\hat{f} - \bar{f}\mathbf{1})^2|\psi\rangle + \alpha^2\langle\psi|(\hat{\mathrm{g}} - \bar{\mathrm{g}}\mathbf{1})^2|\psi\rangle + \alpha\langle\psi|\mathrm{i}(\hat{\mathrm{f}}\hat{\mathrm{g}} - \hat{\mathrm{g}}\hat{\mathrm{f}})|\psi\rangle \geq 0 \tag{7.10}$$

Therefore:

$$(\Delta f)^2 + \alpha^2(\Delta g)^2 + \alpha\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle \geq 0 \tag{7.11}$$

Thus, completing the squares for α:

$$(\Delta f)^2 + (\Delta g)^2\left(\alpha + \frac{\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle}{2(\Delta g)^2}\right)^2 - \frac{\left(\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle\right)^2}{\left(2(\Delta g)^2\right)^2}(\Delta g)^2 \geq 0 \tag{7.12}$$

In the last step, we assumed[1] that $\Delta g \neq 0$. 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:

$$(\Delta f)^2(\Delta g)^2 - \frac{\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle^2}{4} \geq 0 \tag{7.13}$$

and therefore, finally:

$$\Delta f(t)\ \Delta g(t)\ \geq\ \frac{1}{2}|\langle\psi|[\hat{f}(t), \hat{g}(t)]|\psi\rangle| \tag{7.14}$$

[1] If $\Delta g = 0$ then it depends on whether or not also $\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle$ is zero. First, we observe that $\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle$ is a real number because $\left(i[\hat{f}, \hat{g}]\right)^\dagger = i[\hat{f}, \hat{g}]$. Now if $\Delta g = 0$ and $\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle = 0$, then Eq.7.11 implies no condition on Δf. If, however, $\Delta g = 0$ and $\langle\psi|i[\hat{f}, \hat{g}]|\psi\rangle \neq 0$, then $\Delta f = +\infty$ because it must then be larger than any positive real number (since α can be chosen arbitrarily).

Exercise 7.1 *There are indications from studies of quantum gravity*[2]*, that the uncertainty relation between positions and momenta acquire corrections due to gravity effects and should be of the form:* $\Delta x \Delta p \geq \frac{\hbar}{2}(1+\beta(\Delta p)^2+...)$*, where* β *is expected to be a small positive number. (a) Show that this type of uncertainty relation arises if the canonical commutation relation is modified to read* $[\hat{x}, \hat{p}] = i\hbar(1+\beta\hat{p}^2)$*. (b) Sketch the modified uncertainty relation* $\Delta x \Delta p \geq \frac{\hbar}{2}(1+\beta(\Delta p)^2)$ *in the* Δp *versus* Δx *plane. (c) 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.

7.2 The time and energy uncertainty relation

We have seen that in quantum mechanics the position coordinates are observables whose uncertainties can be calculated through $\Delta x = \langle\psi|(\hat{x} - \langle\psi|\hat{x}|\psi\rangle)^2|\psi\rangle^{1/2}$. In contrast, quantum mechanics treats the time coordinate t as a mere parameter. Therefore,

[2]Gravity 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. A lot of follow-up papers (about two a week) are calculating the impact on atoms, black holes and cosmology.

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, $\bar{f}(t)$, and its uncertainty $\Delta f(t)$ generally change over time.

Now consider that we prepare an ensemble (i.e., a collection) of identical experiments, all starting with the same initial condition. Let's say that in a sub-ensemble we measure the observable $\hat{f}$ at a time t_0 and in the remaining cases of the ensemble we measure the observable at a time $t_0 + \Delta t$. The individual outcomes of the measurements of $\hat{f}(t_0)$ will generally differ from the outcomes of the individual measurements of $\hat{f}(t_0+\Delta t)$. There are two reasons for this. One reason is that the system has evolved from time t_0 to time $t_0 + \Delta t$ and so the expectation values $\hat{f}(t_0)$ and $\hat{f}(t_0 + \Delta t)$ are generally different and so of course the typical measurement outcomes are different at the two times. The other reason why the measurements generally differ at the times t_0 and $t_0 + \Delta t$ is that there are quantum fluctuations, namely as predicted by $\Delta f(t_0)$ and $\Delta f(t_0 + \Delta t)$.

This leads to the question which of the two reasons is the dominant one. And that depends! If we choose a short Δt then the system will not have had much time to evolve and therefore $\bar{f}(t_0)$ and $\bar{f}(t_0 + \Delta t)$ are still approximately the same. This means that the differing outcomes of the measurements of $\hat{f}$ at t_0 and at $t_0 + \Delta t$ are almost entirely due to quantum fluctuations. If one the other hand, we make Δt very large then the system had time to evolve significantly and the expectation values $\bar{f}(t_0)$ and $\bar{f}(t_0 + \Delta)$ can therefore differ significantly. In this case the difference between the individual outcomes of the individual measurements of $\hat{f}$ at t_0 or $t_0 + \Delta t$ are almost entirely due to the evolution of the system and not due to the quantum fluctuations.

There must be a Δt that is the threshold between the two cases! Namely, starting from some time t_0, how much time, Δt do we have to wait until the expectation value of $\hat{f}$ has changed appreciably, i.e., until it exceeds the quantum uncertainty? That time is crucial: it is the time we have to wait to be able to see the evolution of a quantum system in spite of the presence of quantum noise!

Concretely, we need to determine the amount of time, Δt, that it takes for the expectation value $\bar{f}(t)$ to change by the average standard deviation $\overline{\Delta f}$. Here, we define the average standard deviation as the time-average over the standard deviation from t_0 to $t_0 + \Delta t$:

$$\overline{\Delta f} := \frac{\int_{t_0}^{t_0+\Delta t} \Delta f(t)\ dt}{\Delta t} \tag{7.15}$$

We want to determine that amount of time Δt that it takes for the change of the expectation value $\bar{f}$ to start exceeding $\overline{\Delta f}$, because that's the amount of time it takes for change due to the dynamics of the system to dominate over the quantum noise. So we are looking for that amount of time Δt for which:

$$|\bar{f}(t_0 + \Delta t) - \bar{f}(t_0)| = \overline{\Delta f} \tag{7.16}$$

Combining Eqs.7.15,7.16, we obtain for the critical Δt:

$$\Delta t = \frac{\int_{t_0}^{t_0+\Delta t} \Delta f(t)\ dt}{|\bar{f}(t_0+\Delta t) - \bar{f}(t)|} \tag{7.17}$$

Interestingly, quantum mechanics implies a fundamental relation between the amount of time, Δt, that it takes for an expectation value $\bar{f}$ to change by more than the average quantum fluctuation and the uncertainty in the energy ΔH. Namely, assuming that the Hamiltonian does not depend on time[3], we obtain:

$$\Delta f(t)\Delta H \geq \frac{1}{2}\left|\langle\psi|[\hat{f}(t),\hat{H}]|\psi\rangle\right| \tag{7.18}$$

and

$$\frac{d\hat{f}(t)}{dt} = \frac{1}{i\hbar}\ [\hat{f}(t),\hat{H}] \tag{7.19}$$

we obtain:

$$\Delta f(t)\Delta H \geq \frac{\hbar}{2}\left|\langle\psi|\frac{d\hat{f}(t)}{dt}|\psi\rangle\right| = \frac{\hbar}{2}\left|\frac{d}{dt}\langle\psi|\hat{f}(t)|\psi\rangle\right| \tag{7.20}$$

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

$$\Delta f(t)\Delta H \geq \frac{\hbar}{2}\left|\frac{d\bar{f}(t)}{dt}\right| \tag{7.21}$$

Integrating over time, we obtain:

$$\begin{aligned}\Delta H\int_{t_0}^{t_0+\Delta t}\Delta f(t)\ dt &\geq \frac{\hbar}{2}\int_{t_0}^{t_0+\Delta t}\left|\frac{d\bar{f}(t)}{dt}\right|\ dt \\ &\geq \frac{\hbar}{2}\left|\int_{t_0}^{t_0+\Delta t}\frac{d\bar{f}(t)}{dt}\ dt\right| = \frac{\hbar}{2}|\bar{f}(t_0+\Delta t) - \bar{f}(t_0)|\end{aligned} \tag{7.22}$$

Now inserting Eq.7.17 we finally obtain the time-energy uncertainty relation:

$$\Delta t\ \Delta H\ \geq\ \frac{\hbar}{2} \tag{7.23}$$

It is quite remarkable that this uncertainty relation looks so similar to that between position and momentum! And, notice that all dependence on the choice of $\hat{f}$ has dropped out. What this means is that if the energy of a system has uncertainty ΔH then for *every* observable $\hat{f}$ it takes at least an amount of time Δt that obeys Eq.7.23 for the mean value $\bar{f}$ of the observable to change appreciably, i.e., for it to change by at

[3]Of course, if we allow ourselves to change the Hamiltonian arbitrarily then we can make observables change arbitrarily quickly too.

least the average uncertainty $\overline{\Delta f}$. In the extreme case where 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! For example, there is definitely no observable rotation of the electron about the proton in any way - because *all* observables' expectation values are constant. Conversely, if a system is to possess any observables $\hat{f}$ which do change appreciably on a short time scale Δt, for example, if we want to run 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.7.23.

Exercise 7.2 *Ultimately, every clock is a quantum system, with the clock's pointer or display consisting of one or more observables of the system. Even small quantum systems such as a nucleus, an electron, atom or molecule have been made to serve as clocks. Assume now that you want to use a small system, such as a molecule, as a clock by observing how one of its observables changes over time. Assume that your quantum clock possess a discrete and bounded*[4] *energy spectrum* $E_1 \leq E_2 \leq E_3 \leq ... \leq E_{max}$ *with* $E_{max} - E_1 = 1eV$ *(1eV=1 electronvolt) which is a typical energy scale in atomic physics. (a) Calculate the maximum uncertainty in energy,* ΔE *that your quantum clock can possess. (b) Calculate the maximally achievable accuracy for such a clock. I.e., what is the shortest time interval (in units of seconds) within which any observable property of the clock could change its expectation value by a standard deviation?*

[4]The energy could be bounded for several reasons. For example, your quantum clock could be a molecule that is such that if you were to supply any more energy that E_{max} to your quantum clock, it would break up, i.e., you would reach the continuous spectrum.