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