Lecture 3

Eslam Khalaf

September 13, 2023

1 Measurement

In the previous lecture, we have started building a dictionary between the abstract formalism of Hilbert spaces and the different physical objects in the quantum theory. We have identified the state of a quantum system with rays in the Hilbert space. We have also identified physical observables with hermitian operators. The real eigenvalues of a Hermitian operators correspond to the possible measurement outcomes. For example, $S_x = \pm\hbar/2$ in the Stern-Gerlach experiment. The eigenkets of the Hermitian operator correspond to the possible states associated with the different measurement outcomes. The fact that the eigenkets of a Hermitian operator form a complete basis is necessary for this interpretation.

However, we have not answered the very important question: what determines which measurement outcome is realized in a given experiment? Let us for concreteness consider again a cascaded Stern-Gerlach setup of SG$z+$, which yields a beam polarized in the $+z$ direction, followed by SGx. As we discussed in Lecture 1, the intensity of the beam is split equally between the two possible outcomes representing $+x$ and $-x$ polarization. Now we ask the question: what would happen if the input $+z$ polarized beam has only one atom? Since quantum mechanics is a fundamental theory of microscopic objects (rather than an emergent theory where concepts only make sense at a certain scale), we should be able to write a ket to describe the state of a single atom. However, a single atom cannot split into two with $\pm x$ polarization. Instead, the atom 'picks' one of the two outcomes with equal probability of 0.5 each. This means that are two very unusual aspects of the quantum theory. First, although the input state was a definite state without built-in uncertainty, the output of the experiment can only be determined probabilistically in the quantum theory. Second, the measurement device seems to play an active rule, forcing the atom that was living happily in a superposition of $|+x\rangle$ and $|-x\rangle$ to make up its mind and choose of of the two alternatives.

If we accept these two pecularities of the quantum world, we can incorporate them in the formalism by introducing the notion of a measurement. To every measurement device, we associate an observable which is a Hermitian operator and we define the measurement as a process which forces the input state into one of the eigenkets of the corresponding operator. The probability of obtaining a certain measurement outcome associated with the eigenvalue λ_a and an eigenket $|a\rangle$ starting from a general normalized ket $|u\rangle$ is given by $|\langle a|u\rangle|^2$. This is known as the Born rule. Let us first make sure that such rule makes sense i.e. the numbers $p_a(u) = |\langle a|u\rangle|^2$ can be interpreted as probabilities. For a set of numbers p_a to represent probabilities, they need to be non-negative and sum to 1. This can seen using the resolution of the unity

$$\langle u|u\rangle = 1 = \sum_a \langle u|a\rangle\langle a|u\rangle = \sum_a |\langle u|a\rangle|^2 = \sum_a p_a(u) \tag{1}$$

An important class of measurement is called selective measurements or filtrations. These are measurements which selects only one of the eigenkets of the measurement device. In the Stern-Gerlach setup, these were the devices SG$\hat{n}\pm$ which only let one of the beams through. A selective measurement which selects the outcome a is described by the operator $\Lambda_a = |a\rangle\langle a|$. This operator satisfies the important property that it is equal to its square

$$\Lambda_a^2 = |a\rangle\langle a|a\rangle\langle a| = |a\rangle\langle a| = \Lambda_a \tag{2}$$

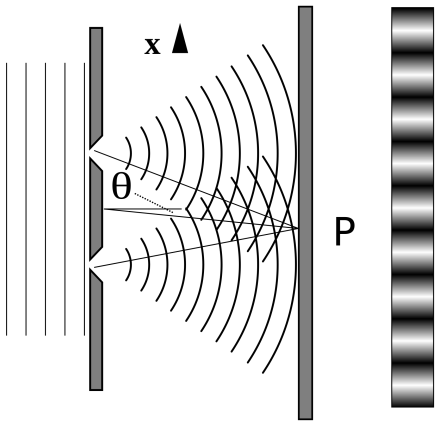

Figure 1: Schematic illustration of the double slit experiment

Physically, this reflects the fact that applying a filtration device Λ_a to the output of an identical filtration device Λ_a does nothing. Operators satisfying (2) are called projection operators [1].

Although the final measurement outcome is obtained through a probability distribution, there is a very important distinction between the quantum theory and standard probability theory. Let us illustrate this with an example. Imagine we have a cascade of three filteration devices (which we can think of as some SG devices with some relative angles), so that the first device $A = \Lambda_a$ selects some outcome $|a\rangle$, the second $B = \Lambda_b$ selects some outcome b and the third $C = \Lambda_c$ selects some outcome c. Assuming the beam coming out of a is normalized, the probability of getting an output in B (the intensity of the B beam) is $|\langle a|b\rangle|^2$. The probability of getting an output in C is $|\langle c|b\rangle|^2|\langle b|a\rangle|^2$ since probabilities are multiplicative. If we repeat the experiment each time filtering a different output of b, the total probability will be

$$\sum_b |\langle c|b\rangle|^2|\langle b|a\rangle|^2 = \sum_b \langle c|b\rangle\langle b|a\rangle\langle a|b\rangle\langle b|c\rangle \tag{3}$$

Remarkably, this turns out to be different from the probability we get if we remove the B device altogether which is given by

$$|\langle c|a\rangle|^2 = \Big|\sum_b \langle c|b\rangle\langle b|a\rangle\Big|^2 = \sum_{b,b'} \langle c|b\rangle\langle b|a\rangle\langle a|b'\rangle\langle b'|c\rangle \tag{4}$$

This means that although we have allowed for all possible intermediate values of B, the fact that we have measured it (and we can record it) makes this different from removing the B device altogether. The reason for the distinction is that in quantum mechanics, if we do not perform a measurement, the intermediate state is a ket that can be in a superposition. Another way to see this is that, unlike probability theory where we add the probabilities of mutually exclusive intermediate steps, in the quantum theory, we add the complex coefficients of basis expansion of a certain ket $\langle a|u\rangle$ whose absolute value squares is the probability (these complex coefficients are called probability amplitudes).

1.1 The double-slit experiment

A very dramatic illustration of this effect is realized in the double-slit experiment, shown in Fig. 1. The double slit experiment is a very old experiment to illustrate the wave nature of light. Its setup involves letting a monochromatic light through a screen with two narrow slits in it. A screen is then placed behind the slits to observe the light. The main observation in this experiment is that the light exhibits an interference pattern of dark and light spots. This pattern is not a simple sum of the output of each individual slit (see

[1] Note that a projection operator is more general since any operator of the form $P = \sum_a |a\rangle\langle a|$ satisfying $P^2 = P$ where the sum goes over any subset of eigenstates

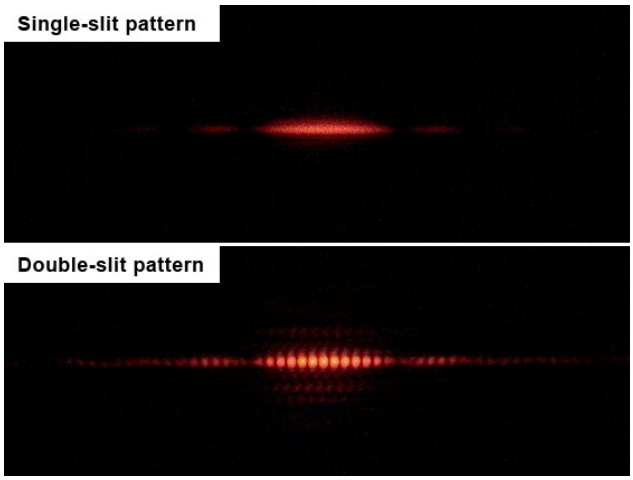

Figure 2: Comparison of the output of a single slit and double slit with the latter displaying clear interference pattern

Fig. 2). Instead it features interference where waves which are in phase add and those which are out of phase subtract. The phase difference arises because the light travels a different distance between the two slits to the screen.

When performing the double slit experiment with electrons instead of light, the same interference pattern is observed. Although the electrons are registered on the screen as individual spots, their distribution follows the same interference pattern as the light waves. The interference pattern is compatible with the rules we have discussed where the probability amplitudes for the two paths are first added, then we take the absolute value squared of the result, leading to an interference pattern. The setup we discussed with the filtration device would correspond to blocking one of the slits and then the others and adding the outputs. Clearly, the resulting output will not show any interference pattern. In fact, even if we let both slits open, but we have some way to measure which slit the electron goes through, for example by shining light on the slits, this will disturb the electrons enough to destroy the interference pattern. This illustrates a fundamental aspects of quantum mechanics where the measurement is not a passive operation *even if we allow for all intermediate measurement outcomes*.

1.2 Expectation value and variance

The born rule means that a given state, $|u\rangle$ defines a probability distribution over the space of possible measurement outcomes with the probability $p_a(u) = |\langle a|u\rangle|^2$ associated with the measurement outcome λ_a. Two important quantities for any probability distribution are the mean or expectation value and the variance. The expectation value of an operator A is defined as $\langle u|A|u\rangle$. It can be written in the familiar form of the mean value for a probability distribution using the resolution of unity:

$$\langle u|A|u\rangle = \sum_{a,b}\langle u|a\rangle\langle a|A|b\rangle\langle b|u\rangle = \sum_a \lambda_a|\langle u|a\rangle|^2 = \sum_a \lambda_a p_a(u) \tag{5}$$

It is important to note that although the allowed values of an observable are discrete, e.g. $S_z = \pm\hbar/2$, the expectation value $\langle S_z\rangle$ can take any value between $-\hbar/2$ and $+\hbar/2$. This is similar to saying that the average number of students per class at Harvard is let's say 12.3 although the allowed numbers can only be positive integers. Another important quantity for a given probability distribution is the variance

$$\langle(\Delta A)^2\rangle = \langle(A - \langle A\rangle)^2\rangle = \langle A^2\rangle - \langle A\rangle^2 = \sum_a(\lambda_a - \langle A\rangle)^2 p_a(u) \tag{6}$$

Clearly, the variance is a non-negative real number $\langle(\Delta A)^2\rangle \geq 0$. The variance of an operator A in a state $|u\rangle$ is zero if and only if $|u\rangle$ is an eigenket of A. To see this, note that the last expression in (6) is a sum

of non-negative numbers so for the sum to vanish, each term has to vanish which implies $\lambda_a = \langle A \rangle$ or $p_a(u) = 0$ for each a. To satisfy normalization, at least one p_a has to be non-zero. If the eigenvalues λ_a are all distinct, then there can only be one $\lambda_a = \langle A \rangle$. Thus, $p_a(u) = \delta_{a,b}$ for some b i.e. $|u\rangle$ is an eigenket of A. Even if the eigenvalues of A has some degeneracy, we can still deduce from the above argument that all non-zero values for p_a correspond to the same eigenvalue $\lambda_a = \langle A \rangle$ which means that $|u\rangle$ is in one of the eigenspaces of A.

1.3 Compatible and incompatible observables

As we have seen in the Stern-Gerlach device, there is a big distinction between cascading devices which have the same set of eigenkets, for example SGz, SG$z+$ and SG$z-$, and cascaded devices with different eigenkets e.g. SGz and SGx. This distinction is formalized through the notion of compatible and incompatible observables. Compatible observables are described by operators A and B which commute with each other

$$[A, B] = AB - BA = 0 \tag{7}$$

As a result, they have a common set of eigenkets (but not eigenvalues). In linear algebra, this corresponds to the statement that commuting matrices can be diagonalized simultaneously. Let us see how this is realized in the language of the quantum theory. Let us first consider the non-degenerate case where all the eigenvalues of A are distinct. Then we can write

$$0 = \langle a|[A, B]|b\rangle = \langle a|AB|b\rangle - \langle a|BA|b\rangle = (\lambda_a - \lambda_b)\langle a|B|b\rangle \tag{8}$$

For $a \neq b$, this implies $\langle a|B|b\rangle = 0$. This means that the off-diagonal matrix elements of B vanish in the eigenket basis of A. In other words, B is diagonal in that basis and has the same eigenkets.

Let us now consider the case where the eigenvalues are degenerate. For example, there are two linearly independent eigenkets $|a, 1\rangle$ and $|a, 2\rangle$ with the same eigenvalues λ_a. Clearly, any linear combination of these two kets is also an eigenket. This means that the eigenkets corresponding to λ_a form a two dimensional subspace. Eq. 8 implies that the matrix elements connecting any state in this space with a state with different eigenvalues λ_b vanish but it does not restrict the form of B in this subspace which will be some 2×2 Hermitian matrix. We can diagonalize this 2×2 Hermitian matrix and label the common eigenstates of A and B in this subspace by $|a, b\rangle$. This example reveals the general matrix structure of commuting operators. In the A eigenbasis, B decomposes into a Block-diagonal form with each Block corresponding to a degenerate subspace of A eigenkets with the same eigenvalue. We can choose special linear combinations of these eigenkets to be eigenkets of B. If some degeneracies are still let, we can have a third commuting operator C and repeat the procedure.

Operators which do not satisfy (7) are called incompatible. Incompatible observables do not have a complete set of common eigenkets since if $A|a, b\rangle = \lambda_a|a, b\rangle$ and $B|a, b\rangle = \lambda_b|a, b\rangle$ for all $|a, b\rangle$ then

$$AB|a, b\rangle = \lambda_b A|a, b\rangle = \lambda_a \lambda b|a, b\rangle = B(\lambda_a|a, b\rangle) = BA|a, b\rangle \tag{9}$$

which implies $AB = BA$. Is there a way to quantify how incompatible two observables are? clearly, some pairs of observables can be almost compatible, in the sense that they deviate from the relation (7) by a small amount (think of an SGz device followed by one rotated by a small angle θ relative to the z axis) while some can be maximally incompatible.

Operator incompatibility has important physical consequences. It is the reason why a beam polarized in $|+z\rangle$ retains this information when it goes through the device SGz but completely forgets it when it goes through SGx. This can be quantified through uncertainty relations which places constraints on the variance of different observables in the same state. Recall, the variance of an operator in a given state measures how close this state is to being an eigenstate of this operator. For incompatible observables, reducing the variance of one observables would increase the variance of the other due to the absence of a common eigenstate. To derive the uncertainty principle, we use the Cauchy-Schwarz inequality

$$|\langle u|v\rangle|^2 \leq \|u\|^2\|v\|^2 \tag{10}$$

The intuitive meaning of the Cauchy-Schwarz inequality is that the projection of a vector onto another cannot be longer than the vector itself. We now take $|u\rangle = \Delta A|w\rangle$ and $|v\rangle = \Delta B|w\rangle$ leading to

$$\|u\|^2\|v\|^2 = \langle(\Delta A)^2\rangle\langle(\Delta B)^2\rangle \geq |\langle\Delta A\Delta B\rangle|^2 \tag{11}$$

We now use the relation $AB = \frac{1}{2}([A,B] + \{A,B\})$ where $\{A,B\}$ is the anticommutator defined as $\{A,B\} = AB + BA$. Due to the relation $(AB)^\dagger = B^\dagger A^\dagger$, the commutator of two hermitian operators is anti-hermitian $[A,B]^\dagger = -[A^\dagger, B^\dagger] = -[A,B]$ whereas the anticommutator is hermitian $\{A,B\}^\dagger = \{A^\dagger, B^\dagger\} = \{A,B\}$. Since the expectation value of a hermitian operator is real and that of an anti-hermitian operator is imaginary, we have

$$|\langle\Delta A\Delta B\rangle|^2 = \frac{1}{4}|\langle[\Delta A, \Delta B]\rangle|^2 + \frac{1}{4}|\langle\{\Delta A, \Delta B\}\rangle|^2 \tag{12}$$

This implies

$$\langle(\Delta A)^2\rangle\langle(\Delta B)^2\rangle \geq \frac{1}{4}|\langle[\Delta A, \Delta B]\rangle|^2 = \frac{1}{4}|\langle[A,B]\rangle|^2 \tag{13}$$

In the last inequality we used the fact that the commutator of a constant with any operator vanishes. This means we can drop out the constant $\langle A\rangle$ in ΔA. Eq. 13 is the uncertainty relation between general observables A and B. When taking A and B to represent position and momentum, this reduces to the famous Heisenberg uncertainty which will be discussed in the next lectures.

1.4 Basis transformations

So far we have been discussing hermitian operators which correspond to physical observables in the quantum theory. There is another class of important operators of the quantum theory; those corresponding to basis transformations. These can be understood as analogs of rotating the $x-y$ coordinates by some angle in a real vector space. Equivalently, we can also think of active transformations where we rotate all vectors while keeping the $x-y$ axes fixed. In both cases, such transformation keeps the lengths of vectors as well as the angles between them fixed. This means that the transformation $|u\rangle \mapsto U|u\rangle$ satisfies $\langle u|U^\dagger U|v\rangle = \langle u|v\rangle$ for every u and v. This is only possible if $U^\dagger U$ is the identity operator:

$$U^\dagger U = \mathbb{1} \tag{14}$$

which means that U is unitary. Any pair of orthonormal complete bases are related by a unitary transformation. To see this, let us call the two bases vectors $|a_l\rangle$ and $|b_l\rangle$ where $l = 1, \ldots, N$ (note that here a and b do not label eigenvalues). Then the matrix $U = \sum_l |b_l\rangle\langle a_l|$ provides a basis transformation since $U|a_l\rangle = |b_l\rangle$ and $U^\dagger U = \sum_{l,m} |a_l\rangle\langle b_l|b_m\rangle\langle a_m| = \sum_l |a_l\rangle\langle a_l| = \mathbb{1}$. Two operators related via $B = UAU^\dagger$ are called unitary equivalent. Unitary equivalent operators have the same eigenvalues since for

$$A|u_a\rangle = \lambda_a|u_a\rangle \tag{15}$$

we can construct $|\tilde{u}_a\rangle = U|u_a\rangle$ with

$$B|\tilde{u}_a\rangle = UAU^\dagger U|u_a\rangle = UA|u_a\rangle = \lambda_a U|u_a\rangle = \lambda_a|\tilde{u}_a\rangle \tag{16}$$

Useful unitary invariants are the determinant and the trace of the operator which are expressed purely in terms of eigenvalues. For the trace, this follows from the cyclic property $\mathrm{tr}XY = \mathrm{tr}YX$ which implies $\mathrm{tr}U^\dagger AU = \mathrm{tr}A$. A hermitian matrix can always be diagonalized through a unitary basis transformation.