As an example, consider a qubit with eigenvalues p and $1-p$. Then $\text{tr}(\rho^\alpha) = p^\alpha + (1-p)^\alpha$ so that Eq. (2.147) becomes

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \ln \Big\{ p^\alpha + (1-p)^\alpha \Big\}. \tag{2.152}$$

This result is plotted in Fig. 2.6 for several values of α. As can be seen, except for $\alpha \to 0$, which is kind of silly, the behavior of all curves is qualitatively similar.

Integral representations of $\ln(\rho)$

When dealing with more advanced calculations, sometimes dealing with $\ln(\rho)$ in terms of eigenvalues can be hard. An alternative is to write the logarithm of operators as an integral representation. I know two of them. If you know more, tell me and I can add them here. A simple one is

$$\ln(\rho) = (\rho - 1) \int_0^1 \frac{\mathrm{d}x}{1 + x(\rho - 1)}. \tag{2.153}$$

Here whenever ρ appears in the denominator, what is meant is the matrix inverse. Another formula is[4]

$$\ln(\rho) = (\rho - 1) \int_0^\infty \frac{\mathrm{d}x}{(1+x)(\rho + x)} \tag{2.154}$$

$$= \int_0^\infty \mathrm{d}x \left(\frac{1}{1+x} - \frac{1}{\rho + x} \right). \tag{2.155}$$

This last formula in particular can now be used as the starting point for a series expansion, based on the matrix identity

$$\frac{1}{A+B} = \frac{1}{A} - \frac{1}{A} B \frac{1}{A+B}. \tag{2.156}$$

For instance, after one iteration of Eq. (2.155) we get

$$\ln(\rho) = \int_0^\infty \mathrm{d}x \left(\frac{1}{1+x} - \frac{1}{x} + \frac{\rho}{x^2} - \frac{\rho^2}{x^2} \frac{1}{\rho + x} \right). \tag{2.157}$$

2.10 Generalized measurements and POVMs

So far our discussion of measurements has been rather shallow. What I have done so far is simply postulate the idea of a projective measurement, without

[4]See M. Suzuki, *Prog. Theo. Phys*, **100**, 475 (1998)

discussing the physics behind it. I know that this may be a bit frustrating, but measurements in quantum mechanics are indeed a hard subject and experience shows that it is better to postulate things first, without much discussion, and then later on study models which help justify these postulates.

Starting from next chapter, we will begin to work out several models of measurements, so I promise things will get better. What I would like to do now is discuss measurements from a mathematical point of view and try to answer the question "what is the most general structure a measurement may have?" We will further divide this in two questions. First, what is a measurement? Well, it is *an assignment from states to probabilities*. That is, given an arbitrary state ρ, we should ask what is the most general way of assigning probabilities to it? This will introduce us to the idea of **POVMs** (Positive Operator-Valued Measures). The next question is, what should be the state of the system after the measurement? That will lead us to the idea of **Kraus operators**.

Ok. So let's start. We have a system prepared in a state ρ. Then we use as our starting point the postulate that a measurement is a probabilistic event for which different outcomes can be obtained with different probabilities. Let us label the outcomes as $i = 1, 2, 3, \ldots$. At this point the number of possible outcomes has no relation in principle to the dimension of the Hilbert space or anything of that sort. All we want to do is find an operation which, given ρ, spits out a set of probabilities $\{p_i\}$. Well, from density matrices, information is always obtained by taking the expectation value of an operator, so this association must have the form

$$\boxed{p_i = \text{tr}(E_i \rho),} \tag{2.158}$$

where E_i are certain operators, the properties of which are determined once we impose that the p_i should behave like probabilities. First, the p_i must be non-negative for any ρ, which can only occur if the operators E_i are positive semi-definite. Second, $\sum_i p_i = 1$ so that $\sum_i E_i = 1$. We therefore conclude that if the rule to associate probabilities with quantum states has the structure of Eq. (2.158), the set of operators $\{E_i\}$ must satisfy

$$\boxed{E_i \geq 0, \qquad \sum_i E_i = 1.} \tag{2.159}$$

A set of operators satisfying this is called a **POVM**: Positive-Operator Valued Measure, a name which comes from probability theory. The set of POVMs also contain projective measurements as a special case: the projection operators P_i are positive semi-definite and add up to the identity.

POVMs for a qubit

Here is an example of a POVM we can construct by hand for a qubit:

$$E_1 = \lambda|0\rangle\langle 0|$$
$$E_2 = (1-\lambda)|0\rangle\langle 0| + |1\rangle\langle 1|. \tag{2.160}$$

These guys form a POVM provided $\lambda \in [0,1]$: they are positive semi-definite and add up to the identity. However, this is in general not a projective measurement, unless $\lambda = 1$. The logic here is that outcome E_1 represents the system being found in $|0\rangle$, but outcome E_2 means it can be in either $|0\rangle$ or $|1\rangle$ with different probabilities. For a general qubit density matrix like (2.18), we get

$$p_1 = \text{tr}(E_1\rho) = \lambda p, \tag{2.161}$$

$$p_2 = \text{tr}(E_2\rho) = 1 - \lambda p. \tag{2.162}$$

So even if $p = 1$ (the system is for sure in $|0\rangle$), then we can still obtain the outcome E_2 with a certain probability. From such a silly example, you are probably wondering "can this kind of thing be implemented in the lab?" The answer is yes and the way to do it will turn out to be much simpler than you can imagine. So hang in there!

What is cool about POVMs is that we can choose measurement schemes with more than two outcomes, even though our qubit space is two-dimensional. For instance, here is an example of a POVM with 3 outcomes (taken from Nielsen and Chuang):

$$E_1 = q|0\rangle\langle 0| \tag{2.163}$$

$$E_2 = q|+\rangle\langle +| \tag{2.164}$$

$$E_3 = 1 - E_1 - E_2. \tag{2.165}$$

To illustrate what you can do with this, suppose you are walking down the street and someone gives you a state, telling you that *for sure* this state is either $|1\rangle$ or $|-\rangle$, but he/she doesn't know which one. Then if you measure the system and happen to find outcome E_1, you know for certain that the state you were given could not be $|1\rangle$, since $\langle 0|1\rangle = 0$. Hence, it must have been $|-\rangle$. A similar reasoning holds if you happen to measure E_2, since $|+\rangle$ and $|-\rangle$ are orthogonal. But if you happen to measure E_3, then you don't really know anything. So this is a POVM where the observer never makes a mistake about what it is measuring, but that comes at the cost that sometimes he/she simply doesn't learn anything.

Generalized measurements

We now come to the much harder question of what is the state of the system after the measurement. Unlike projective measurements, for which the state always collapses, for general measurements many other things can happen. So we should instead ask what is the most general mathematical structure that a state can have after a measurement. To do that, I will postulate something, which we will only prove on later chapters, but which I will try to give a reasonable justification below. You can take this next postulate as the ultimate measurement postulate: that is, it is a structure worth remembering because every measurement can be cast in this form. The postulate is as follows.

Measurement postulate: any quantum measurement is fully specified by a set of operators $\{M_i\}$, called **Kraus operators**, satisfying

$$\boxed{\sum_i M_i^\dagger M_i = 1.} \tag{2.166}$$

The probability of obtaining measurement outcome i is

$$\boxed{p_i = \mathrm{tr}(M_i \rho M_i^\dagger),} \tag{2.167}$$

and, if the outcome of the measurement is i, then the state after the measurement will be

$$\boxed{\rho \to \frac{M_i \rho M_i^\dagger}{p_i}.} \tag{2.168}$$

Ok. Now breath! Let us analyze this in detail. First, for projective measurements $M_i = P_i$. Since $P_i^\dagger = P_i$ and $P_i^2 = P_i$ we then get $P_i^\dagger P_i = P_i$ so that Eqs. (2.166)-(2.168) reduce to

$$\sum_i P_i = 1, \qquad p_i = \mathrm{tr}(P_i \rho), \qquad \rho \to \frac{P_i \rho P_i}{p_i}, \tag{2.169}$$

which are the usual projective measurement/collapse scenario. It also does not matter if the state is mixed or pure. In particular, for the latter $\rho = |\psi\rangle\langle\psi|$ so the state after the measurement becomes (up to a constant) $P_i|\psi\rangle$. That is, we have projected onto the subspace where we found the system in.

Next, let us analyze the connection with POVMs. Define

$$E_i = M_i^\dagger M_i. \tag{2.170}$$

Then Eqs. (2.166) and (2.167) become precisely Eqs. (2.158) and (2.159). You may therefore be wondering why define POVMs separately from these generalized measurements. The reason is actually simple: different sets of measurement operators $\{M_i\}$ can give rise to the same POVM $\{E_i\}$. Hence, if one is only interested in obtaining the probabilities of outcomes, then it doesn't matter which set $\{M_i\}$ is used, all that matters is the POVM $\{E_i\}$. However, states having the same POVM can lead to completely different post-measurement states.

Examples for a qubit

Consider the following measurement operators

$$M_1 = \begin{pmatrix} \sqrt{\lambda} & 0 \\ 0 & 0 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} \sqrt{1-\lambda} & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.171}$$

These operators satisfy (2.166). Moreover, $E_1 = M_1^\dagger M_1$ and $E_2 = M_2^\dagger M_2$ give exactly the same POVM as in Eq. (2.160). Suppose now that the system is initially in the pure state $|+\rangle = \frac{1}{\sqrt{2}}(1, 1)$. Then the outcome probabilities and the states after the measurements will be

$$\begin{aligned} p_1 &= \frac{\lambda}{2} & |+\rangle &\to |0\rangle \\ p_2 &= 1 - \frac{\lambda}{2} & |+\rangle &\to \frac{\sqrt{1-\lambda}|0\rangle + |1\rangle}{\sqrt{2-\lambda}} \end{aligned} \tag{2.172}$$

Thus, unless $\lambda = 1$, the state after the measurement will not be a perfect collapse.

Next consider the measurement operators defined by

$$M_1' = \begin{pmatrix} 0 & 0 \\ \sqrt{\lambda} & 0 \end{pmatrix}, \qquad M_2' = M_2 = \begin{pmatrix} \sqrt{1-\lambda} & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.173}$$

Compared to Eq. (2.171), we have only changed M_1. But note that $M_1'^\dagger M_1' = M_1^\dagger M_1$. Hence this gives the same POVM (2.160) as the set $\{M_i\}$. However, the final state after the measurement is completely different: if outcome 1 is obtained, then instead of (2.172), the state will now collapse to

$$|+\rangle \to |1\rangle. \tag{2.174}$$

To give a physical interpretation of what is going on here, consider an atom and suppose that $|0\rangle = |e\rangle$ is the excited state and $|1\rangle = |g\rangle$ is the ground-state. The system is then initially in the state $|+\rangle$, which is a superposition of the two. But if you measure and find the atom in the excited state, then that means it must have emitted a photon and therefore decayed to the ground-state. The quantity λ in Eq. (2.173) therefore represents the probability of emitting a photon *during the time-span* of the observation. If it emits, then the state is $|1\rangle = |g\rangle$ because it must have decayed to the ground-state. If it doesn't emit, then it continues in a superposition, but this superposition is now updated to $\sim \sqrt{1-\lambda}|0\rangle + |1\rangle$. This is really interesting because it highlights the fact that *if nothing happens, we still update our information about the atom.* In particular, if λ is very large, for instance $\lambda = 0.99$, then the state after the measurement will be very close to $|1\rangle$. This means that if the atom did not emit, there is a huge chance that it was actually in the ground-state $|1\rangle$ to begin with.

Origin of generalized measurements

Now I want to show you one mechanism through which generalized measurements appear very naturally: *a generalized measurement is implemented by making a projective measurement on an ancilla that is entangled with the system.* That is, instead of measuring A, we first entangle it with an auxiliary system B (which we call *ancilla*) and then measure B using projective measurements. Then, from the point of view of A, this will be translated into a generalized measurement.

To illustrate the idea, suppose we have a system in a state $|\psi\rangle_A$ and an ancilla prepared in a state $|0\rangle_B$. Then, to entangle the two, we first evolve them with a joint unitary U_{AB}. The joint state of AB, which was initially product, will then evolve to a generally entangled state

$$|\phi\rangle = U_{AB}\Big[|\psi\rangle_A \otimes |0\rangle_B\Big]. \tag{2.175}$$

We now perform a projective measurement on B, characterized by a set of projection operators $P_i^B = 1_A \otimes |i\rangle_B\langle i|$. Then outcome i is obtained with probability

$$p_i = \langle\phi|P_i^B|\phi\rangle, \tag{2.176}$$

and the state after the measurement, if this outcome was obtained, collapses to $P_i^B|\phi\rangle$.

Now let's see how all this looks from the perspective of A. The next calculations are a bit abstract, so I recommend some care. Have a first read all the way to the end and then come back and try to understand it in more detail. The point is that here the $\otimes$ can be a curse. It is better to get rid of it and write, for instance, $P_i^B = |i\rangle_B\langle i|$ where the fact that this is an operator acting only on Hilbert space B is implicit in the subscript. Similarly we write $|\psi\rangle_A \otimes |0\rangle_B = |\psi\rangle_A|0\rangle_B$. With this we then get, for instance,

$$p_i = {}_A\langle\psi|\,{}_B\langle 0|U_{AB}^\dagger|i\rangle_B\langle i|U_{AB}|\psi\rangle_A|0\rangle_B. \tag{2.177}$$

This quantity is a scalar, so we are contracting over everything. But what we could do is leave the contraction $\langle\psi|(\ldots)|\psi\rangle$ for last. Then the $(\ldots)$ will be an operator acting only on the Hilbert space of A. If we define the operators

$$\boxed{M_i = {}_B\langle i|U_{AB}|0\rangle_B = \Big(1\otimes\langle i|\Big)U_{AB}\Big(1\otimes|0\rangle\Big).} \tag{2.178}$$

acting only Hilbert space A, then we get

$$p_i = {}_A\langle\psi|M_i^\dagger M_i|\psi\rangle_A, \tag{2.179}$$

which is precisely Eq. (2.167) for the probabilities of a generalized measurement. Moreover, we can also check that the $\{M_i\}$ satisfy the normalization condition (2.166):

$$\sum_i M_i^\dagger M_i = \sum_i {}_B\langle 0|U_{AB}^\dagger|i\rangle_B\langle i|U_{AB}|0\rangle_B$$

$$= {}_B\langle 0|U_{AB}^\dagger U_{AB}|0\rangle_B$$

$$= {}_B\langle 0|0\rangle_B$$

$$= 1_A,$$

so they indeed form a set of measurement operators.

We now ask what is the reduced density matrix ρ_A^i of system A, given that the outcome of the measurement on B was i. Well, this is simply obtained by taking the partial trace over B of the new state $P_i^B|\phi\rangle$:

$$\rho_A^i = \text{tr}_B\left\{P_i^B|\phi\rangle\langle\phi|P_i^B\right\}$$

$$= {}_B\langle i|\phi\rangle\langle\phi|i\rangle_B$$

$$= {}_B\langle i|U_{AB}|\psi\rangle_A|0\rangle_B\;{}_A\langle\psi|\;{}_B\langle 0|U_{AB}^\dagger|i\rangle_B$$

Using Eq. (2.178) this may then be written as

$$\rho_A^i = M_i\Big(|\psi\rangle\langle\psi|\Big)M_i^\dagger, \tag{2.180}$$

which is exactly the post-measurement state (2.168). Thus, as we set out to prove, if we do a projective measurement on a ancila B which is entangled with A, from the point of view of A we are doing a generalized measurement.

The above calculations are rather abstract, I know. It is a good exercise to do them using $\otimes$ to compare. That can be done decomposition $U_{AB} = \sum_\alpha A_\alpha \otimes B_\alpha$. Eq. (2.177), for instance, then becomes:

$$p_i = \sum_{\alpha,\beta}\Big(\langle\psi|\otimes\langle 0|\Big)\Big(A_\alpha^\dagger\otimes B_\alpha^\dagger\Big)\Big(1\otimes|0\rangle\langle 0|\Big)\Big(A_\beta\otimes B_\beta\Big)\Big(|\psi\rangle\otimes|0\rangle\Big).$$

I will leave for you as an exercise to check that this indeed gives (2.179). Also, try to check that the same idea leads to Eq. (2.180).