CS191 – Fall 2014
Lecture 10: Generalized measurement, partial trace, and distance in state space

Mohan Sarovar
(Dated: October 1, 2014)

I. REVIEW OF PROJECTIVE (VON NEUMANN) MEASUREMENTS

You have seen that a conventional measurement in quantum mechanics (sometimes referred to as a *von Neumann* measurement) is defined as a collection of projectors. That is, given an observable O, since it is a Hermitian operator, we know it has a spectral decomposition as:

$$O = \sum_i \lambda_i P_i, \tag{1}$$

where the P_i are projectors ($P_i P_j = P_i \delta_{ij}$) and the λ_i are unique, meaning that the P_i are not necessarily rank 1. If the P_i is not rank 1, then it can be expressed as a sum of rank 1 projectors: *i.e.*, if the rank of P_i is k, $P_i = \left|\phi_1^i\right\rangle\left\langle\phi_1^i\right| + \left|\phi_2^i\right\rangle\left\langle\phi_2^i\right| ... + \left|\phi_k^i\right\rangle\left\langle\phi_k^i\right|$. Measuring an observable O on a state $|\psi\rangle$ yields a result that we label with the λ_i, and we think of the measurement as projecting onto one of the P_i. The probability of getting λ_i is

$$\Pr\{i\} = \langle\psi| P_i |\psi\rangle . \tag{2}$$

If instead of a pure state, we have a mixed state description of the system, ρ, then this probability is:

$$\Pr\{i\} = \mathrm{tr}\,(P_i \rho). \tag{3}$$

Exercise: Confirm that in the special case that ρ is a pure state, *i.e.*, $\rho = |\psi\rangle\langle\psi|$, Eq. (3) becomes the same as Eq. (2).

The state of the system after performing the measurement (the *post-measurement* state) is prescribed by the update formulas

$$\text{Pure states}: \quad |\psi\rangle \;\rightarrow\; \frac{P_i|\psi\rangle}{\sqrt{\langle\psi|P_i|\psi\rangle}} = \frac{P_i|\psi\rangle}{\sqrt{\Pr\{i\}}}, \tag{4}$$

$$\text{Mixed states}: \quad \rho \;\rightarrow\; \frac{P_i \rho P_i}{\mathrm{tr}\,(P_i\rho)} = \frac{P_i \rho P_i}{\Pr\{i\}}, \tag{5}$$

given that we obtained result i from the measurement. Note that if the state before measurement is pure, then the post-measurement state is also pure.

Example 1 Consider a general one qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$. Suppose we measure the observable σ_z. It has a spectral decomposition:

$$\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$$

The projectors, P_i, are rank 1 in this case. There are two possible results: $+1$ or -1, with probabilities:

$$\Pr\{+1\} = |\langle\psi|\;\;|0\rangle\langle 0|\;\;|\psi\rangle|^2 = |\alpha|^2$$
$$\Pr\{-1\} = |\langle\psi|\;\;|1\rangle\langle 1|\;\;|\psi\rangle|^2 = |\beta|^2$$

If we obtain $+1$ as the measurement results, the post-measurement state is

$$\frac{|0\rangle\langle 0|\;\;|\psi\rangle\langle\psi|\;\;|0\rangle\langle 0|}{|\alpha|^2} = |0\rangle\langle 0|$$

Example 2 Consider a system of two qubits, and a measurement of the σ_z observable of the first qubit. The observable in this case is $\sigma_z \otimes I$, which has spectral decomposition:

$$\sigma_z \otimes I = (|00\rangle\langle 00| + |01\rangle\langle 01|) - (|10\rangle\langle 10| + |11\rangle\langle 11|) \tag{6}$$

This is an example of an observable with a spectral decomposition with projectors that are not rank 1. What this means physically is that measurements of this observable are not able to distinguish between the states $|00\rangle$ and $|01\rangle$, and between the state $|10\rangle$ and $|11\rangle$. This makes sense since we're only doing a measurement of qubit 1's properties.

We often say we measure the expectation value of an observable O when the system is in some state $|\psi\rangle$. The value of this expectation value, denoted $\langle O\rangle_\psi$, is $\langle\psi| O |\psi\rangle$ (or $\mathrm{tr}\,(O\rho)$ in the more general case of a mixed state). However, the expectation value is the average value of the observable under the given state, and like all averages it cannot be determined by a single measurement. Instead, it should be kept in mind that how one experimentally determines the $\langle O\rangle_\psi$ is:

1. Prepare state $|\psi\rangle$.
2. Perform the measurement of O, which results in projection onto one of the P_i in the spectral decomposition of O, Eq. (1). The measurement result is m_1 (which is one of the λ_i in Eq. (1)), record this value.
3. Repeat the above two steps N times to generate $m_2, m_3, ...m_N$.
4. The expectation value is then the statistical average of all the measurement results: $\frac{1}{N}\sum_{j=1}^{N} m_j$. This empirical average converges to the observable's expectation value as $N \to \infty$.

II. GENERALIZED MEASUREMENTS

The above description of projective measurements turns out to be too restrictive. There are measurements that can be performed on a system that cannot be described within this formalism. These are called non-projective measurements, generalized measurements, or positive operator-valued measures (POVMs) [1].

A generalized measurement in quantum mechanics is described a collection of positive operators $E_i \geq 0$ that satisfy $\sum_i E_i = I$. We denote such a measurement as $M = \{E_i\}$. Each E_i is associated with an outcome of the measurement and since $E_i \geq 0$, it has the decomposition $E_i = M_i^\dagger M_i$. For a state ρ (pure or mixed), the probability of obtaining the result associated with E_i is

$$\Pr\{i\} = \mathrm{tr}\,(E_i\rho). \tag{7}$$

Similarly, the post-measurement state after obtaining result i is

$$\rho \to \frac{M_i\rho M_i^\dagger}{\mathrm{tr}\,(E_i\rho)} = \frac{M_i\rho M_i^\dagger}{\Pr\{i\}} \tag{8}$$

We see from these definitions that the projective measurements described above are a special case where $M_i = M_i^\dagger = P_i$, and in this case $E_i = P_i$ as well since $P_i^\dagger P_i = P_i$. The main difference between projective measurements and POVM elements is that the POVM elements do not have to be orthogonal. That is, while $P_iP_j = P_i\delta_{ij}$, the same orthogonality relation does not exist for the M_i or E_i.

Exercise: What is the definition of a positive operator? Is a positive operator also Hermitian?

The simplest example of the need for a POVM description of measurements is given by the following scenario: consider a measurement apparatus that measures a single qubit in the computational basis. However, the apparatus can fail with probability p and when it does, it does not interact with the qubit and therefore returns no measurement result. In this case, there is no projective description of the measurement, instead it is described by POVM elements:

$$\begin{aligned} E_1 &= pI \\ E_2 &= (1-p)\,|0\rangle\langle 0| \\ E_3 &= (1-p)\,|1\rangle\langle 1| \end{aligned}$$

Example 3 Another example of a generalized measurement on a single qubit is given by the POVM elements [from Nielsen & Chuang, p.92]:

$$\begin{aligned} E_1 &= \frac{\sqrt{2}}{1+\sqrt{2}}\,|1\rangle\langle 1| \\ E_2 &= \frac{\sqrt{2}}{1+\sqrt{2}}(|0\rangle - |1\rangle)(\langle 0| - \langle 1|) \\ E_3 &= I - E_1 - E_2 \end{aligned} \tag{9}$$

---

[1] There is a reason for this last name, but going into it would take us too far afield. Therefore, we will simply use the terminology "POVM" without explanation. See the references in the further reading section if you really want to know the reason for this terminology.

Note that $E_iE_j \neq \delta_{ij}$ for any i, j, and hence this is clearly not a projective measurement. This POVM allows one to discriminate between the state $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. These two states are not orthogonal and therefore there is no measurement that completely distinguishes them. For example, consider measuring in the computational basis. If we get $|1\rangle$ as the measurement result then the state must have been $|\psi_2\rangle$, but if we get $|0\rangle$ as the result then the state could have been either one of the two. However, note that E_1 is orthogonal to $|\psi_1\rangle$ and E_2 is orthogonal to $|\psi_2\rangle$. Therefore if we get a measurement result corresponding to $E_1(E_2)$, then we know the state must have been $|\psi_2\rangle\,(|\psi_1\rangle)$. The cost for this discrimination is that if we get a result that corresponds to E_3, then the measurement tells us nothing.

Exercise: Verify that $\sum_i E_i = I$ for the the POVM elements given in example 3.

III. IMPLEMENTING POVMS AND NEUMARK'S THEOREM

How do we actually implement a generalized measurement? This is usually done through the coupling of the system to be measured to an *ancilla* system and then doing projective measurement of the ancilla. To formally see what this looks like, consider a system to be measured in some pure state $|\psi\rangle$. If this system is coupled to an ancilla (maybe a set of qubits), and both evolve under some unitary, U_{SA}, the resulting composite state is: $U_{SA}\,|\psi\rangle_S \otimes |\phi\rangle_A$, where $|\phi\rangle$ is the initial state of the ancilla , and the subscripts S, A denote the system and ancilla. Then assume we do a conventional projective measurement on the ancilla in a basis $|m_i\rangle_A \langle m_i|$ for $i = 1, 2, ...$ The probability of getting result i is (see Eq. (2))

$$p_i = (\langle\psi|_S \otimes \langle\phi|_A)U_{SA}^\dagger \left[I_S \otimes |m_i\rangle_A \langle m_i|\right] U_{SA}(|\psi\rangle_S \otimes |\phi\rangle_A). \tag{10}$$

The composite state after this projective measurement with result i is (see Eq. (4))

$$|\Psi_{SA}\rangle = \frac{\left(M_i\,|\psi\rangle_S\right) \otimes |m_i\rangle}{\sqrt{p_i}}, \tag{11}$$

where M_i is an operator *acting on the system only* that takes the form

$$M_i\,|\psi\rangle_S \equiv \langle m_i|\,U_{SA}(|\psi\rangle_S \otimes |\phi\rangle_A). \tag{12}$$

This operator depends on $|\phi\rangle_A, |m_i\rangle$ and U_{SA}, and it defines a generalized measurement on the system. To see this, note that

$$\langle\psi|_S \sum_i M_i^\dagger M_i\,|\psi\rangle_S = \sum_i (\langle\psi|_S \otimes \langle\phi|_A)U_{SA}^\dagger \left[I_S \otimes |m_i\rangle_A \langle m_i|\right] U_{SA}(|\psi\rangle_S \otimes |\phi\rangle_A) = 1, \tag{13}$$

and since this is true for any state $|\psi\rangle$, it must be that $\sum_i M_i^\dagger M_i = I$. Hence $E_i \equiv M_i^\dagger M_i$ satisfy the two properties that define POVM elements for a generalized measurement, namely $E_i \geq 0$ and $\sum_i E_i = I$.

Exercise: Why is Eq. (13) true?

Note that this generalized measurement can be tuned by choosing three things: (i) the initial state of the ancilla, $|\phi\rangle_A$, (ii) the unitary that couples the system and ancilla, U_{SA}, and (iii) the basis that the ancilla is measured in $\{|m_i\rangle\}$. In fact, one can show that it is possible to engineer *any* POVM using this approach of coupling to an ancilla of suitable size and performing projective measurement on it. This is the content of *Neumark's theorem.* We won't say more about this theorem here, but you can find more details on it on page 285 of *"Quantum theory: concepts and methods"* by Asher Peres (Kluwer Academic Publishers), 1995.

Exercise: What happens if the unitary U_{SA} does not couple the system and ancilla, *i.e.*, it factors as $U_{SA} = U_S \otimes U_A$?

IV. PARTIAL TRACE AND REDUCED STATES

You have seen that the way to describe composite quantum systems is to use the tensor product operation. That is, if I have two qubits and their states are $|\psi_1\rangle$ and $|\psi_2\rangle$, then the state of the whole system composed of both qubits, is $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$. Importantly, in quantum mechanics there exists states of composite systems that have extremely non-classical properties. These are *entangled* states and you will hear more about them in coming lectures (and a little later).

But here I want to consider the opposite of composition of individual systems. Namely, how do we describe the state of subsystems of a composite system. For example, let the state of a two qubit composite system be $|\phi\rangle$ (represented as a vector, this is a 4×1 normalized vector with complex entries). Now, what is the state of qubit 1 in this system? The way to find the state of such a subsystem is to take the *partial trace*. More generally, given the state of a composite system ρ^{AB} (the subsystems are labeled A and B), the state of subsystem A is

$$\rho^A = \mathrm{tr}_B(\rho^{AB}), \tag{14}$$

where $\mathrm{tr}_B(\cdot)$ is the partial trace over subsystem B. ρ^A is referred to as the *reduced density operator* of system A. Operationally, the way to perform a partial trace is to contract the composite density matrix with a complete set of states in the Hilbert space of being traced over. That is, suppose system B is k dimensional, and a basis for this k dimensional space is $\left|\phi_1^B\right\rangle, \left|\phi_2^B\right\rangle, ... \left|\phi_k^B\right\rangle$ (here the superscript B just reminds us that these states live in the Hilbert space of the second subsystem). Then the partial trace is

$$\rho^A = \mathrm{tr}_B(\rho^{AB}) = \sum_{j=1}^{k} \left\langle \phi_j^B \right| \rho_{AB} \left| \phi_j^B \right\rangle \tag{15}$$

Let us look at a couple of examples.

Example 4 Let the composite state of two qubits be

$$\left|\psi^{AB}\right\rangle = \left|0^A\right\rangle \otimes \frac{1}{\sqrt{2}}(\left|0^B\right\rangle + \left|1^B\right\rangle)$$

The composite density matrix is

$$\rho^{AB} = \left(\left|0^A\right\rangle \otimes \frac{1}{\sqrt{2}}(\left|0^B\right\rangle + \left|1^B\right\rangle)\right)\left(\left\langle 0^A\right| \otimes \frac{1}{\sqrt{2}}(\left\langle 0^B\right| + \left\langle 1^B\right|)\right)$$

The reduced density operator of A is

$$\begin{aligned} \rho^A &= \left\langle 0^B\right| \rho^{AB} \left|0^B\right\rangle + \left\langle 1^B\right| \rho^{AB} \left|1^B\right\rangle \\ &= \frac{1}{2}\left|0^A\right\rangle\left\langle 0^A\right| + \frac{1}{2}\left|0^A\right\rangle\left\langle 0^A\right| = \left|0^A\right\rangle\left\langle 0^A\right| \end{aligned}$$

In this case the reduced state is pure, and we can equivalently write it as $\left|\psi^A\right\rangle = \left|0^A\right\rangle$.

Example 5 Let the composite state of two qubits be

$$\left|\psi^{AB}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0^A\right\rangle \otimes \left|0^B\right\rangle + \left|1^A\right\rangle \otimes \left|1^B\right\rangle\right)$$

The composite density matrix is

$$\rho^{AB} = \frac{\left|0^A\right\rangle\left|0^B\right\rangle\left\langle 0^A\right|\left\langle 0^B\right| + \left|1^A\right\rangle\left|1^B\right\rangle\left\langle 0^A\right|\left\langle 0^B\right| + \left|0^A\right\rangle\left|0^B\right\rangle\left\langle 1^A\right|\left\langle 1^B\right| + \left|1^A\right\rangle\left|1^B\right\rangle\left\langle 1^A\right|\left\langle 1^B\right|}{2},$$

where we dispense with writing the $\otimes$ and the superscripts explicitly for conciseness – *i.e.*, $\left|0^A\right\rangle\left|0^B\right\rangle \equiv \left|0^A\right\rangle \otimes \left|0^B\right\rangle$ and so on. The reduced density operator of A is

$$\begin{aligned} \rho^A &= \left\langle 0^B\right| \rho^{AB} \left|0^B\right\rangle + \left\langle 1^B\right| \rho^{AB} \left|1^B\right\rangle \\ &= \frac{\left|0^A\right\rangle\left\langle 0^A\right|}{2} + \frac{\left|1^A\right\rangle\left\langle 1^A\right|}{2} \end{aligned}$$

In this case the reduced state is mixed, in fact it is the completely mixed state – *i.e.*, $\rho^A = I/2$.

The above examples illustrate some important properties of the partial trace and reduced density matrices:

1. Just like the trace operation, the partial trace is linear.
2. If the composite system is in a product state, $\rho^{AB} = \rho^A \otimes \rho^B$, then the partial trace just results in the component state – *e.g.*, $\mathrm{tr}_B(\rho^A \otimes \rho^B) = \rho^A$.
3. Even if the composite system is in a pure state (a state vector), the reduced state for a subsystem can be mixed (a density matrix). For this reason it is always best to work out the partial trace in density matrix form, even if the initial state is pure (as we did in example 4). This is actually a deep result which gets at one of the primary differences between quantum and classical physics. Pure states of composite systems that lead to mixed states of a subsystem (as a result of partial tracing) possess entanglement.

A final important property of reduced density matrices is that they contain all the information one needs to make predictions about the subsystem not being traced over. That is, consider again a composite system with state ρ^{AB}. Then any observable only on subsystem A has the form $O^A \otimes I^B$, where O^A is the observable and I^B is the identity operator on the Hilbert space of system B. Then the expectation value of this observable is:

$$\langle O^A \rangle = \mathrm{tr}_{AB}((O^A \otimes I^B)\rho^{AB}) = \mathrm{tr}_A(O^A \rho^A), \tag{16}$$

where $\rho_A = \mathrm{tr}_B(\rho^{AB})$, and we have explicitly noted the Hilbert spaces over which each trace is being taken.

Exercise: One way to get the last equality in Eq. (16) is to use the fact that $\mathrm{tr}_{AB}(\cdot) = \mathrm{tr}_A \mathrm{tr}_B(\cdot) = \mathrm{tr}_B \mathrm{tr}_A(\cdot)$; *i.e.*, one can take the full trace by taking partial traces over subsystems in any order. Convince yourself of this fact.

V. MEASURES OF DISTANCE BETWEEN QUANTUM STATES

Often we want to compare two quantum states and ask how similar they are. For example, if we want to check how well a quantum device that we designed is performing, one way to do this would be to calculate how close the state that the device outputs is to the ideal state it should output when given an input state. There are several ways to quantify the similarity of quantum states, and we will examine one of the most commonly used ones, the *state fidelity*.

Given two states of the same system (*i.e.*, from the same Hilbert space), ρ and σ, the *fidelity* between the states is:

$$F(\rho, \sigma) = \mathrm{tr}\,(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}). \tag{17}$$

This definition involves the square root of positive operators, which is defined through the spectrum of the operator. That is, if $A \geq 0$, then it possess a spectral decomposition: $A = UDU^\dagger$, where U is a unitary matrix whose columns are the eigenvectors of A and D is a diagonal matrix containing the (positive) eigenvalues of A. Then $\sqrt{A} = U\sqrt{D}U^\dagger$, where the square root of a diagonal matrix is given by simply taking the square root of its elements.

Exercise: Show that if $\rho = |\psi_\rho\rangle\langle\psi_\rho|$ and $\sigma = |\psi_\sigma\rangle\langle\psi_\sigma|$ are both pure states, then the fidelity simplifies to:

$$F(\rho, \sigma) = |\langle\psi_\sigma|\psi_\rho\rangle| \tag{18}$$

This expression for fidelity is a little hard to manipulate and calculate (mainly because of all the square roots), but there is a strong operational motivation for defining it this way. Consider how we would actual go about determining how similar ρ and σ are if we were given many copies of them. We would perform some measurement on both and compare the measurement results. Let us call the measurement we perform a general POVM with elements $\{E_i\}$. Then it can be shown that [see Nielsen & Chuang, p.412]

$$F(\rho, \sigma) = \min_{\{E_i\}} F_{\mathrm{cl}}(\{p_i\}, \{q_i\}), \tag{19}$$

where $F_{\mathrm{cl}}(\{p_i\}, \{q_i\})$ is the classical fidelity between two classical probability distributions (defined below), and $\{p_i = \mathrm{tr}\,(E_i\rho)\}$ and $\{q_i = \mathrm{tr}\,(E_i\sigma)\}$ are the probability distributions for the POVM elements under the states ρ and σ.

This says that the fidelity is actually the distance between the probability distributions produced by the two states, when you do an general measurement on the system. In fact, it's the minimum distance, when you optimize over all

possible measurements you could do. In this sense the fidelity has a very strong operational meaning, which is why it is commonly used in quantum information.

The classical fidelity used above is a measure of distance between discrete probability distributions, and is defined as:

$$F_{\mathrm{cl}}(\{p_i\},\{q_i\}) = \sum_i \sqrt{p_i q_i} \tag{20}$$

Some properties of the fidelity [most of these are proved in Nielsen & Chuang, section 9.2.2]:

1. The fidelity is symmetric in its arguments – *i.e.*, $F(\rho,\sigma) = F(\sigma,\rho)$.
2. When one of the states being compared is pure, the fidelity reduces to $F(|\psi\rangle\langle\psi|,\sigma) = \sqrt{\langle\psi|\sigma|\psi\rangle}$. This is easier to calculate than the general case, and it very applicable in a lot of quantum computing applications where one wants to compare to an ideal pure state.
3. The fidelity is invariant under unitary transformations – *i.e.*, $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho,\sigma)$.
4. The fidelity is also non-decreasing under more general transformations on states, but we will get to this after we cover these transformations.

VI. REFERENCES AND FURTHER READING

1. Nielsen & Chuang. Sections 2.2.3 – 2.2.6, 2.4.3, 9.2.2.
2. *"Quantum theory: concepts and methods"* by Asher Peres (Kluwer Academic Publishers), 1995. This book has a great treatment of the basic quantum formalism, POVMs and Neumark's theorem.