Quantum Computation **(CMU 18-859BB, Fall 2015)**

Lecture 20: Pretty Good Measurement

11/16/15

Lecturer: John Wright *Scribe: Lachlan Lancaster*

1 Introduction: Quantum Hypothesis Testing

So we start with a promise problem: given a set of density matrices $\{\sigma_i\}$ and a quantum state ρ we are promised that ρ is in state σ_i with probability p_i. In the general case we have $i \in [m]$ and of course $\sum_{i=1}^{m} p_i = 1$. Our Goal is then to succesfully identify which of the σ_i that our state ρ is actually in, this is known as *Quantum Hypothesis Testing*. The "real life" analog of this is what you can imagine Bob trying to do for the case of the transported information from alice in terms of the Holevo Bound for information transportation.

Getting at the problem we want to at least maximize our probability of getting the state right. This maximization with respect to both the probabilities on each state as well as with respect to any randomness that our approach employs. We then must choose a Quantum POVM (Positive-Operator Valued Measure) $\{E_i\}$ that carries put a measurement and maximizes our probability of getting the state right.

So say we pick a POVM then we know from our previous lectures that:

$$\Pr[\text{observe } \sigma_i] = \mathrm{Tr}\left(\sigma_i E_i\right) \tag{1}$$

So that our overall probability of success, (as in telling that we have the right state), is then:

$$\Pr[\text{success}] = \sum_i p_i \mathrm{Tr}\left(\sigma_i E_i\right) \tag{2}$$

So this is the quantity that we will seek to maximize in this lecture, we can see that in the case that $m = 2$, this is quite easy.

1.1 Case $m = 2$

In this case we only have two possible outcomes of density matrices. That is, our state ρ can only be in state σ_1 (with probability p_1) or in state σ_2 (with probability p_2). It is easy to see that we must have $p_2 = 1 - p_1$, and in fact that is the main advantage of this case which results in it being an easy problem. In light of this, we will relabel p_1 as simply p. In this case, using equation (2) we have:

$$\Pr[\text{success}] = p\text{Tr}(\sigma_1 E_1) + (1-p)\text{Tr}(\sigma_2 E_2)$$

Where E_1 and E_2 come from the POVM that we have chosen, which is still arbitrary as of now. We then notice that, due to the nature of the POVM, we are abel to make a similar observation that $E_2 = I - E_1$ where I is the identity matrix. We can then again relabel E_1 as simply E and substituting this into the above we have:

$$\begin{aligned}\Pr[\text{success}] &= p\text{Tr}(\sigma_1 E) + (1-p)\text{Tr}(\sigma_2(I-E)) \\ &= p\text{Tr}(\sigma_1 E) + (1-p)\left(\text{Tr}(\sigma_2) - \text{Tr}(\sigma_2 E)\right) \\ &= (1-p) + p\text{Tr}(\sigma_1 E) - (1-p)\text{Tr}(\sigma_2 E) \\ &= (1-p) + \text{Tr}\left(E(p(\sigma_1+\sigma_2)-\sigma_2)\right)\end{aligned}$$

Where we note that above we used the linearity of the trace operation as well as the fact that density matrices are required to have trace one. From the result we see that in order to maximize our probability of success we need only maximize the trace of the matrix in the second part of the above equation $E(p(\sigma_1+\sigma_2)-\sigma_2)$. The way to maximize this trace is then to simply choose E so that is projects onto the positive eigenspace of the matrix it is multiplying, that is the matrix $p(\sigma_1+\sigma_2)-\sigma_2$. Thus we have a direct strategy of picking the best POVM for any set of σ_i, so the problem is solved!

We will see that the interesting case is when $m \geq 3$, this problem is **unsolved** in modern quantum information theory. But, as you might have been able to tell from the title of the lecture, we have some pretty ideas of how to do pretty good with this problem.

2 Some Tractable Cases

We wil now begin to tackle the problem of qunatum hypothesis testing in the case that $m \geq 3$ and see some of the problems that arise along the way in not allowing us to do perfectly well. In order to do this well we will first define some notation/terminology.

Definition 2.1. We witl define the *optimal success probability*, denoted P_s^{OPT}, as the best possible probability of success that you can get given a set of possible density matrices $\{\sigma_i\}$ with corresponding probabolities of observing p_i.

Definition 2.2. We then denote the *optimal error probability*, denoted P_e^{OPT}, as the lowest possible probability of failure given the states as above.

One can easily see that the above two definitions are complementary in teh sense that $P_s^{\text{OPT}} + P_e^{\text{OPT}} = 1$. As we will see later,for $m \geq 3$ we **do not** have a way of finding a POVM that achieves these optimal probabilities. However, we **do** no of a measurement/POVM that does pretty well.

2.1 Pure States

To start to look at cases with $m \geq 3$ we will begin with the relatively easy case where each of the σ_i are pure states and are orthonormal to each other. To be precise, we will deal with the case where $\sigma_i = |v_i\rangle \langle v_i|$ where the set $\{|v_i\rangle\}$ is assumed to be an orthonormal basis set for the Hilbert space under consideration.

We then assert that the best measurement you can choose is then the POVM where $E_i = \sigma_i$. Note that this is in fact a POVM due to the nature of the σ_i and their creation from an orthonormal basis so that $\sum_i \sigma_i = I$. So we see that our probability of success is then:

$$\Pr[\text{success}] = \sum_i p_i \text{Tr}\left(\sigma_i^2\right) = \sum_i p_i \text{Tr}\left(\sigma_i\right) = \sum_i p_i = 1$$

Note we used the fact above that $\sigma_i^2 = (|v_i\rangle \langle v_i|)^2 = |v_i\rangle \langle v_i|v_i\rangle \langle v_i| = |v_i\rangle \langle v_i| = \sigma_i$. So that in this specific case we have a way of always being right! Well that's nice, but it is for quite a particular case. Let's generalize a bit.

2.2 Sum of Pure States

We now consider the case where each of the σ_i can be stated as a linear combination of pure states created from an orthonormal basis. For example we could have:

$$\sigma_1 = \frac{|v_1\rangle \langle v_1| + |v_2\rangle \langle v_2|}{2}, \quad \sigma_2 = |v_3\rangle \langle v_3|, \quad \sigma_3 = \frac{|v_1\rangle \langle v_1| + \cdots + |v_4\rangle \langle v_4|}{4}, \ldots$$

Where we note from the above that we require the density matrices to be properly normalized. So what is the best measurement strategy here? Well before coosing $E_i = \sigma_i$ seemed to work, how about we try that here? As you might be able to guess it's not that simple in this case. To see why consider the case where we have:

$$\sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$

WWhere tha above matrices are stated in terms of the aformentioned orthonormal basis $|v_i\rangle$. We can see that upon inspection we can do much better than simpy picking $E_i = \sigma_i$, as for the second density matrix this would result in a very low probability. So can we do a slight tweak on this strategy to get somthing better? It turns out that we can. Consider the following

$$E_i = \sigma_i^{-1/2} \sigma_i \sigma_i^{-1/2} \tag{3}$$

We will see that this proves useful but first we should clarify what we mean by $\sigma_i^{-1/2}$. As shown by [Pen55] as well as others, there is a general way to give a sort-of-inverse for all matrices, even if they are not square or of full rank. These matrices are called *pseudoinverses* the most general sense usually referring the the Moore-Penrose pseudoinverse. This is what we refer to when we write σ_i^{-1}, even though the density matrices are not in general invertible. In the context of these density matrices, which are diagonal (as they are sums of pure states created from an orthonormal basis), this simply corresponds to performing the inverse operation only on the diagonal elements. Additionally the square root in this case refers to taking the square root of the eigenvalues of the matrix or equivalently in this case the diagonal elements. For example we then have in terms of the matrix σ_2 above:

$$\sigma_2^{-1/2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \end{pmatrix}$$

So that in this case:

$$\sigma_2^{-1/2}\sigma_2\sigma_2^{-1/2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

So we see the true sense in which this pseudoinverse really acts as an inverse, but only on the part of the marix that it makes sense to try to take an inverse of. Note that this also works if instead of using $\sigma_i^{1/2}$ we use $S^{-1/2}$ where $S = \sigma_1 + \sigma_2$ (the reader is invited to check this for themselves).

In the more general case of more than two σ_i we generalize $S = \sum_i \sigma_i$ and use these as the elements of our POVM. However, this only works when we are dealing with the case that these σ_i are created as the linear combination of an orthonormal basis. In order to move to the most general case we must relax this assumption.

3 Pretty Good Measurement

We see here that the most general case, that is where we are simply given a set of σ_i is not far off from what we had in the previous section. WE try the reasonable first approximation of setting the $E_i = p_i\sigma_i$ where we are now weighting our POVM by the likelihood that we are going to be seeing a certain state. You might see notice that this doesn't work as:

$$S = \sum_i E_i \neq I \tag{4}$$

So that this is not a valid POVM. We see that in fact this can't be the case as:

$$\mathrm{Tr}(S) = \sum_i \mathrm{Tr}(E_i) = \sum_i p_i \mathrm{Tr}(\sigma_i) = \sum_i p_i = 1 \tag{5}$$

But it is quite clear that if $\text{Tr}(S) = 1$ then $S \neq I$. So maybe we can fix *this* to get a POVM that works. Okay, what if we keep S the same but now use (in analogy with the last section) $E_i = S^{-1/2} p_i \sigma_i S^{-1/2}$, we then see that we have:

$$\sum_i E_i = \sum_i S^{-1/2} p_i \sigma_i S^{-1/2} = S^{-1/2} \left(\sum_i p_i \sigma_i \right) S^{-1/2} = S^{-1/2} S S^{-1/2} = I \tag{6}$$

So we again regain the required property of a POVM. Note that we only truly get the identity if we know that the σ_i span the whole space, but we may assume this without loss of generality as we can simply extend the set while assigning probability zero to these additional σ_i. We will see that this POVM is exactly the measurement after which this lecture is named [HW94].

3.1 Example $m = 3$

We'll now go over a specific example where we hae three possible densities, that is $m = 3$. These three matrices are then:

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

We then assign the probabilities of measurement $p_1 = p_3 = p/2$ and $p_2 = 1 - p$. This then completely specifies the current problem. We note that if we consider the above density matrices we have know σ_1 will always output $|0\rangle$, σ_2 will output $|0\rangle$ with proabability $\frac{1}{2}$ and $|1\rangle$ with probability $\frac{1}{2}$, and σ_3 will always output $|1\rangle$. So we see that the best chance we have of getting this right is $P_s^{\text{OPT}} = \max\{p, 1 - p\}$ and we can actually give a way of achieving this optimal strategy.

To do this let's try simply measuring in the standard basis (this will work). So suppose that we use this measurement and the resulting state is $|0\rangle$, what should we do? Well it of course depends on p. We see that the probability that the state was σ_1 and we observe $|0\rangle$ is $\Pr(\sigma_1 \& |0\rangle) = p$ and correspondingly $\Pr(\sigma_2 \& |0\rangle) = p - 1$. This is of course onluy possible as $\Pr(\sigma_3 \& |0\rangle) = 0$. We then see that our choice is simply dependent on whether p of $1 - p$ is bigger, then we simply choose the state (σ_1 or σ_2) that is more likely.

So what happens if we implement the "Pretty Good Measurement" in this case? Well, from equation (2) we can find the probability of success in this situation. We first find what our POVM is by fist finding S.

$$S = \sum_i p_i \sigma_i = \frac{p}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{p}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + (p-1) \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

So that we have:

$$S^{-1/2} = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix}$$

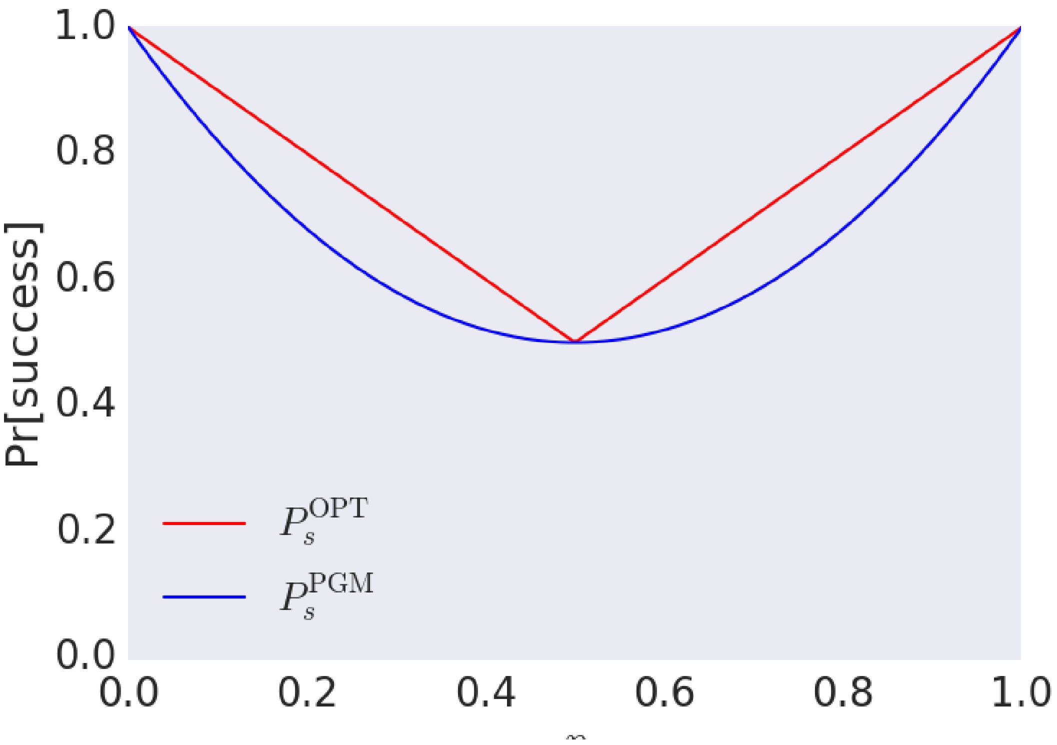
Computing $E_i = S^{-1/2} p_i \sigma_i S^{-1/2}$ we find:

$$E_1 = p \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_2 = (1-p) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad E_3 = p \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Finally, using equation (2) we have:

$$P_s^{\text{PGM}} = \frac{p}{2}\text{Tr}(\sigma_1 E_1) + (1-p)\text{Tr}(\sigma_2 E_2) + \frac{p}{2}\text{Tr}(\sigma_3 E_3) = p^2 + (1-p)^2 \tag{7}$$

We can see the comparison of the success probabilites of each model in the graph below. As we can see the Pretty Good Measurement seems to 'smooth' out the success rate in the optimal strategy.



So that we can see very clearly that the Pretty Good Measurement is **not** optimal. So why are we concerned with it? The thing is that while the PGM may not be the best that we can do in this specific case, it is often an optimal strategy. In order to quantify *just how often* we will introduce the following thoerem.

Theorem 3.1. *The error probability of the PGM is bounded above as:*

$$P_e^{\text{PGM}} \leq \sum_{i \neq j} \sqrt{p_i p_j} F(\sigma_i, \sigma_j) \tag{8}$$

This leads us to the question of the mysterious new function that appear in equation (8), this is what we will call the *Fidelity function.*

4 Fidelity Function

We can think of the Fidelity function as a way of measuring "how close" the individual states σ_i are to each other. We have seen similar concepts in past lectures such as the trace distance between two states ρ and σ:

$$d_{\text{Tr}}(\rho, \sigma) = ||\rho - \sigma||_1 \tag{9}$$

Where the subscript 1 above is to indicate that the map being applied is the $\mathcal{L}^1$ norm on the vector consisting of the diagonal elements of $\rho - \sigma$. Note that in this case the more similar the two matrices are, the smaller the trace distance. Though it seems from the placement of the Fidelity function within the above definition, we should have a function that gets larger the more similar the states are. We hence define the Fidelity function as:

Definition 4.1. The Fidelity function for two density matrices ρ and σ is given by:

$$F(\rho, \sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1 \tag{10}$$

Hence we define the fidelity function as the sum of the square roots of the eigenvalues of the matrix given by $\sqrt{\sigma}\rho\sqrt{\sigma}$.

To make more sense out of what exactly is going on here lets demonstrate this for some diagonal martrices. Say we have $\rho = \text{diag}(a_1, \dots a_d)$ and $\sigma = \text{diag}(b_1, \dots b_d)$. Then by $\sqrt{\rho}$ we mean $\sqrt{\rho} = \text{diag}(\sqrt{a_1}, \dots \sqrt{a_d})$ and by the $\mathcal{L}^1$ norm in this context we mean $||\rho||_1 = \sum_i |a_i|$. We then have that, in this context, $F(\rho, \sigma) = \sum_i \sqrt{a_i}\sqrt{b_i}$. So that we see the larger the fidelity, the more similar the states are, just as we wanted!

Note that since ρ and σ are density matrices we know they have trace 1, so that $F(\rho, \rho) = ||\sqrt{\rho}\sqrt{\rho}||_1 = ||\rho||_1 = 1$. It turns out that this can be made an if and only if statement, so let's go in to some general fidelity function properties:

1. The fidelity function is symmetric in its arguments $F(\rho, \sigma) = F(\sigma, \rho)$.

2. $F(\rho, \sigma) = 1$ if and only of $\rho = \sigma$.

3. If σ is a pure state ($|\psi\rangle\langle\psi|$ for some $|\psi\rangle$) then $\sqrt{\sigma} = \sigma$. In this case $F(\rho, \sigma)$, for some density matrix ρ, is then simply the sum of the square roots of the eigenvalues of the matrix:

   $$\sqrt{\sigma}\rho\sqrt{\sigma} = \sigma\rho\sigma = |\psi\rangle\langle\psi|\rho|\psi\rangle\langle\psi| = (\langle\psi|\rho|\psi\rangle)|\psi\rangle\langle\psi| = \chi|\psi\rangle\langle\psi|$$

   Where we have defined $\chi \equiv \langle\psi|\rho|\psi\rangle$, which is simply some scalar. Thus the fidelity is given by $F(\rho, \sigma) = \sqrt{\chi}$, which is quite a simplification.

4. If we have that both of the density matrices (say $\sigma = \langle\psi|\rho|\psi\rangle$ and $\rho = \langle\phi|\rho|\phi\rangle$) are pure states then the Fidelity function is given very easily by a similar analysis to above as:

   $$F(\rho, \sigma) = |\langle\phi|\psi\rangle| \tag{11}$$

5. We can actually develop a close realtionship between the trace distance and the fidelity as:

   $$1 - F \leq d_{\mathrm{Tr}} \leq \sqrt{1 - F^2} \tag{12}$$

   Note the quantity $1 - F^2$ id often referred to as the *infidelity.* This then gives that if F is very close to 1, that is $F = 1 - \epsilon$ for small $\epsilon > 0$, then we have a very tight bound on the trace distance as $1 - F = \epsilon \leq d_{\mathrm{Tr}} \leq \sqrt{1 - F^2} \approx \sqrt{2\epsilon}$.

6. We have a Theorem from Uhlman [Uhl75]:

   Theorem 4.2. *From Uhlman we have the following restatement of the fidelity as:*

   $$F(\rho, \sigma) = \max\{|\langle\varphi|\psi\rangle|\} \tag{13}$$

   where the maximum is taken over all possible purifications $|\varphi\rangle$ of ρ and $|\psi\rangle$ of σ.

To conclude we leave off with the general idea that Fidelity is small when your states are "far away" from each other and large when the states are "close" to each other. Returning to the relevance of the Fidelity with respect to the PGM error bound we see that if the fidelity is quite large we have and we have a uniform distribution on our states we have:

$$P_e^{\mathrm{PGM}} \leq \frac{1}{2}\sum_{i \neq j}\sqrt{p_i p_j}F(\sigma_i, \sigma_j) \leq \frac{1}{2}\sum_{i \neq j}\sqrt{p_i p_j} \leq \frac{1}{2}\sum_{i \neq j}\sqrt{m^{-2}} = \frac{m}{2} \tag{14}$$

Where we have assumed above, as the notation indicates earlier in the notes, that there are m possible states. But this is a terrible bound! We don't even know if this is less than one. However, we will see that in some important cases, specifically in the case of the Hidden Subgroup Problem, this will give quite a good bound.

References

[HW94] Paul Hausladen and William K. Wootters. A pretty good measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994.

[Pen55] R. Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:406–413, 7 1955.

[Uhl75] A. Uhlmann. The transition probability in the state space of a *-algebra. *Reports of Mathematical Physics*, 9:273–279, 10 1975.