CMSC 858L: Quantum Complexity

Instructor: Daniel Gottesman

Spring 2023

22 More About Quantum Supremacy

22.1 References

Supremacy of IQP is proven in Bremner, Jozsa, and Shepherd, "Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy," arXiv:1005.1407.

The argument and assumptions needed for hardness of weak simulation originate in Aaronson and Arkhipov, "The computational complexity of linear optics," arXiv:1011.3245. However, the version here is greatly simplified from their argument. I was unfortunately not able to find any paper with a good exposition of the full argument in a generic context without details specific to specific circuit families, even though it seems to be well-known to those working on the subject.

22.2 IQP

Let us examine one of these intermediate classes in more detail.

Definition 1. *IQP is a sampling class. A sampling problem is in IQP if it is the task of generating the output distribution of a quantum circuit of the following form:*

1. *Initialize all n qubits in the state $|+\rangle = |0\rangle + |1\rangle$.*
2. *Perform polynomially many diagonal gates, each acting on at most $O(\log n)$ qubits.*
3. *Measure each qubit in the X basis $|\pm\rangle = |0\rangle \pm |1\rangle$.*

IQP circuits are clearly non-universal quantum circuits, since the gates are all diagonal gates. They can't change the basis states at all, only the phases of the basis states. In fact, all the gates commute with each other! It doesn't matter what order they are performed in. This seems very weak, and it is very unlikely it is computationally universal for either classical or quantum computation.

However, surprisingly, once you add in post-selection (to get the class PostIQP), you get all of PostBQP:

Theorem 1. *PostIQP = PostBQP = PP.*

Proof. An approximately universal set of quantum gates is diagonal gates (or just controlled-Z and $R_{\pi/8}$) plus Hadamard. (Since the CNOT is $(I \otimes H)(C - Z)(I \otimes H)$.) We might as well assume (since the gate set is universal) that the initial states and measurements are in the X basis. We then just need to be able to simulate H within IQP plus post-selection.

This can be done as follows: For each H gate, introduce an ancilla (in the $|+\rangle$ state, of course), perform a $C - Z$ between the ancilla and the qubit on which we wish to perform H, and then measure the original data qubit in the X basis, post-selecting on the $|+\rangle$ outcome. The data has now shifted to the former ancilla

qubit but with a Hadamard on it: To see this, assume initial data state $|\psi\rangle = \alpha|0\rangle|\phi_0\rangle + \beta|1\rangle|\phi_1\rangle$ (where $|\phi_i\rangle$ is a state on the other qubits in the computer. Then the $C-Z$ gate gives us

$$C-Z|+\rangle|\psi\rangle = \alpha|00\rangle|\phi_0\rangle + \alpha|10\rangle|\phi_0\rangle + \beta|01\rangle|\phi_1\rangle - \beta|11\rangle|\phi_1\rangle \quad (1)$$

$$= \alpha|+\rangle|0\rangle|\phi_0\rangle + \beta|-\rangle|1\rangle|\phi_1\rangle \quad (2)$$

$$= \alpha|+\rangle|+\rangle|\phi_0\rangle + \beta|-\rangle|+\rangle|\phi_1\rangle + \alpha|+\rangle|-\rangle|\phi_0\rangle - \beta|-\rangle|-\rangle|\phi_1\rangle. \quad (3)$$

Post-selecting on $|+\rangle$ on the second qubit and then discarding the second qubit gives us

$$\alpha|+\rangle|\phi_0\rangle + \beta|-\rangle|\phi_1\rangle = H|\psi\rangle, \quad (4)$$

as desired. □

22.3 Approximate Quantum Supremacy

However, insisting on an *exact* weak simulation seems unreasonable. What about a simulation that generates *close* to the correct distribution? That is a more sensible notion of simulation.

There are multiple possible notions of what it means for a probability distribution to be close. One for which it is easy to extend supremacy results is to small *multiplicative error*.

Definition 2. *We say a random variable Y is close to random variable X up to* multiplicative error $1+\epsilon$ *if*

$$\text{Prob}(X=a)/(1+\epsilon) \leq \text{Prob}(Y=a) \leq (1+\epsilon)\text{Prob}(X=a) \quad (5)$$

for all a.

In other words, for every outcome, the probabilities of that outcome in the two distributions are within a multiplicative factor $1+\epsilon$ of each other.

Theorem 2. *If a quantum model $\mathcal{Q}$ with post-selection can decide PostBQP, then if there exists a weak classical simulation of $\mathcal{Q}$ with multiplicative error $1+\epsilon$ (for sufficiently small ϵ), then the polynomial hierarchy collapses to the third level.*

Proof. Suppose we have a weak classical simulation of $\mathcal{Q}$ with multiplicative error $1+\epsilon$. Consider an arbitrary language in PostBQP and consider a circuit Q in $\mathcal{Q}$ with post-selection that decides it. Then Q has an output qubit B and one or more post-selected qubits A. (If $\mathcal{Q}$ is arbitrary quantum circuits, we can combine post-selection on multiple qubits into a single qubit by letting that qubit be the AND of the other post-selected qubits. However, it might not be possible to do this in non-universal models $\mathcal{Q}$.)

The weak classical simulation of Q with the same measurements but no post-selection gives output distribution (A_C, B_C), whereas Q without post-selection has output distribution (A_Q, B_Q). We know that the C probabilities are within a factor $1+\epsilon$ of the Q probabilities.

With post-selection,

$$\text{Prob}(B_C=b|A_C=a) = \frac{\text{Prob}(A_C=a, B_C=b)}{\text{Prob}(A_C=a)} \quad (6)$$

$$\leq \frac{(1+\epsilon)\text{Prob}(A_Q=a, B_Q=b)}{\text{Prob}(A_Q=a)/(1+\epsilon)} \quad (7)$$

$$= \frac{(1+\epsilon)^2\text{Prob}(A_Q=a, B_Q=b)}{\text{Prob}(A_Q=a)} \quad (8)$$

$$\text{Prob}(B_C=b|A_C=a) \geq \frac{\text{Prob}(A_Q=a, B_Q=b)}{(1+\epsilon)^2\text{Prob}(A_Q=a)}. \quad (9)$$

Thus, there is a PostBPP simulation of PostBQP with multiplicative error $(1+\epsilon)^2$. Now, we can amplify the success probability of a PostBQP language, so that the probability of success in PostBQP is at least

$(2/3)(1+\epsilon)^2$ (for small enough ϵ that this is less than 1). Then if we have a "yes" instance, the probability of acceptance for the PostBPP algorithm is at least a fraction $(1+\epsilon^2)$ of the PostBQP acceptance probability, which is at least 2/3. Similarly, in a "no" instance, the probability of rejection in PostBPP is at least 2/3. □

However, this still seems too strong a condition. If we have two outcomes with probability $1/2 - 1/2^n$ and two with probability $1/2^n$, insisting that that a classical simulation have constant multiplicative error means that it has to get the two unlikely outcomes almost correct as well as the two likely ones. However, it would be very hard to tell the difference between the correct distribution and one that had only the two likely outcomes with probability 1/2 each.

A better condition is that the simulated distribution and the true distribution be close in statistical distance:

Definition 3. *We say a random variable Y is close to random variable X up to* additive error ϵ *if*

$$\frac{1}{2}\|X - Y\|_1 = \frac{1}{2}\sum_a |\text{Prob}(X = a) - \text{Prob}(Y = a)\| \leq \epsilon. \tag{10}$$

When two probability distributions are close up to additive error ϵ, then we can distinguish which of the two we have with $O(1/\epsilon)$ samples from the distribution, presuming we know precisely what the two distributions are.

Ideally, we would like to be able to say that if there is a weak classical simulation of certain quantum models up to additive error ϵ, then the polynomial hierarchy collapses. Unfortunately, this is challenging and has not been proven for any model, as far as I know, without additional assumptions. However, the assumptions needed have been systematized and it seems plausible that they are true for some of these models. In particular, in order to prove hardness (assuming PH doesn't collapse and the model with post-selection can decide PostBQP), we want a model to have a *worst-to-average-case reduction* and to satisfy an *anticoncentration* property. Anticoncentration has been for a number of different models, but in the few cases where we have a worst-to-average case reduction, those reductions are not strong enough for the desired result.

To understand why we want these two properties (which I will define in a little bit), let us first study what it means for a classical simulation to have additive error ϵ. Let $p_a(U) = \text{Prob}(X = a)$, where X is the distribution of outcomes for the quantum circuit U, and let $q_a(U) = \text{Prob}(Y = a)$, where Y is the distribution of outcomes for the classical simulation of the circuit U. We have that $\sum_a |p_a(U) - q_a(U)| \leq 2\epsilon$, but this doesn't necessarily say very much about a particular difference $|p_a(U) - q_a(U)|$ (although it does say it must be less than ϵ. In particular, if $p_a(U)$ is much smaller than ϵ, then it is possible that $q_a(U)$ is much larger than $p_a(U)$.

This is a problem for our argument, which goes through post-selection. Imagine that we have two output bits and post-select on the first one being 1. If we have the distribution $p_{10} = 2^{-n}$ and $p_{11} = 2^{-n/2}$, then conditioned on the postselection, the outcome 1 is extremely likely ($2^{n/2}$ times as likely as outcome 0). But if $q_{10} = 2^{-n/2}$ and $q_{11} = 2^{-n}$, while $q_{00} = p_{00}$ and $q_{01} = p_{01}$, then the simulation (non-post-selected) has additive error $2(2^{-n/2} - 2^{-n})$, which is quite small, but the post-selected simulation is completely wrong, with outcome 0 being $2^{n/2}$ as likely as outcome 1. Thus, even if the simulation is quite close in a context without post-selection, it may be completely wrong when post-selection is added. This is in contrast to multiplicative error.

One solution is to try to show that the existence of a classical simulation with additive error ϵ implies a classical simulation with multiplicative error $1 + O(\text{poly}(\epsilon))$. This is not quite true, but when we assume these additional properties, something sufficiently close to it is true. When there are 2^n possible outcomes, the *average* difference $\langle |p_a(U) - q_a(U)| \rangle \leq 2^{-n+1}\epsilon$, which means that most of the outcomes must have close probability. If we have a particular outcome a such that $|p_a - q_a| \leq 2^{-n}\epsilon$ and also $p_a \geq 2^{-n}\mu$, then it follows that we have a close multiplicative approximation to p_a:

$$(1 - \epsilon/\mu)p_a \leq q_a \leq (1 + \epsilon/\mu)p_a < \frac{1}{1 - \epsilon/\mu}p_a, \tag{11}$$

so q_a has multiplicative error $1/(1-\epsilon/\mu)$.

For a fixed U, we can't really control which outcomes are going to be remain after post-selection; that is determined by the hardness proof. So the strategy is going to be use a random U drawn from a family $\mathcal{F}$ so that the different U can mix up the different outcomes, effectively hiding the outcomes that we care most about from the simulation. We want typical $U \in \mathcal{F}$ to be hard to simulate (average case hardness), and that for all outcomes, for most $U \in \mathcal{F}$, the outcome is not too unlikely (in the sense that $p_a(U) \geq 2^{-n}\mu$; this is anticoncentration).

Definition 4. *We say that a family $\mathcal{F}$ of quantum circuits on n qubits satisfies an* anticoncentration bound *if, for all $\delta_A > 0$, $\exists \mu = \Omega(\text{poly}(1/\delta, n))$ such that for all x and for uniformly random $U \in \mathcal{F}$,*

$$\text{Prob}(|\langle x|U|0\rangle|^2 < \mu/2^n) < \delta_A. \tag{12}$$

#P is a class of counting problems containing problems of the following form: If $f(x,w)$ is a polynomial-time computable function, then, given x, how many values of w satisfy $f(x,w) = 1$?

We say that a family $\mathcal{F}$ of quantum circuits is average case #P-hard *if the existence of a classical randomized algorithm C which satisfies the conditions below implies the existence of a classical randomized algorithm to solve any #P problem. The conditions on C are that it runs in time $O(\text{poly}(n, 1/\epsilon_R, 1/\delta_R))$ and performs a strong simulation of a uniformly random U on n qubits from $\mathcal{F}$ with multiplicative error $1+\epsilon_R$ with probability at least $1-\delta_R$ over U.*

Note that the average case #P-hardness is to give a *strong* simulation, which is at least as hard as a weak simulation.

Also, note that #P is very closely related to PP. If $f \in \#\text{P}$ with a possible range of 2^n values of w, then the decision problem "is $g(x) > 2^{n-1}$?" is in PP. Also,

$$\text{P}^{\#\text{P}} = \text{P}^{\text{PP}}. \tag{13}$$

Thus, #P-hardness is essentially a functional version of PP-hardness. If we can solve a #P-hard problem, we can also decide any PP language.

There is also an additional property, not named or specifically defined in the literature, that lets you disguise the actual circuit U' of interest as another circuit U. Then the weak simulation must simulate U not knowing which circuit U' is really the concern, and so it is unlikely to fail for the specific outcome a we are concerned about. I will call this a *circuit obfuscation* property. In the proof, it lets you assume that *every* q_a has bounded multiplicative error, provided you pick a random U from the family $\mathcal{F}$.

Theorem 3. *Suppose a family $\mathcal{F}$ of quantum circuits satisfies an anticoncentration bound, is average case #P-hard, and has the circuit obfuscation property. Then if there exists a classical randomized algorithm to perform a weak simulation with additive error ϵ of a circuit U on n qubits randomly selected from $\mathcal{F}$ with probability $1-\delta$ in time $O(\text{poly}(n, 1/\epsilon, 1/\delta)$, then the polynomial hierarchy collapses to the third level.*

The proof basically puts together the thoughts from above and combines with the previous arguments about multiplicative and exact worst-case hardness.

For some circuit families, instead of using a circuit obfuscation property, the authors instead make a stronger hardness assumption, that it is hard to approximate p_a up to multiplicative error $1+\epsilon_R$ for a fraction $1-\delta_R$ (which is sometimes a smallish constant) of all pairs (a, U), $U \in \mathcal{F}$.

Proof. Suppose we have such a classical randomized algorithm. Then for random circuit U, with probability $1-\delta$, we have a classical algorithm C that outputs a probability distribution q_a for outcome a, with $1/2\|q_a - p_a\|_1 \leq \epsilon$, where p_a is the output distribution of U.

Now, by using C and Stockmeyer's approximate counting algorithm, we get a randomized classical algorithm using an NP oracle that, given a, can compute q_a to multiplicative error $1+\epsilon'$ with probability $1-\delta'$. Stockmeyer's approximate counting algorithm is a version of the algorithm we have already seen to place PostBPP in BPP^{NP}: That algorithm generated samples for the conditional probabilities of single bits

a conditioned on b. In the previous case, b was just a single bit as well, but that did not play any role in the algorithm; any efficiently computable condition would work the same way. Since it is just a conditional distribution over a single bit a, if $\text{Prob}(a|b) = O(1)$, it straightforward to estimate $\text{Prob}(a|b)$ to multiplicative accuracy $1 + \epsilon_1'$ with only a few samples. We can also bias the distribution $\text{Prob}(a|b)$ by adding m random bits and rejecting one value of a (say we reject $a = 1$) except when all the additional random bits are 1. This effectively magnifies $\text{Prob}(a = 0|b)$ by a factor 2^m, allowing us to estimate it to multiplicative error $1 + \epsilon_1'$ even when it is exponentially small. Chaining conditional probabilities of the single bits together, we can get the probability of a longer bit string, giving an estimate of q_a. The multiplicative errors multiply together $\prod_i (1 + \epsilon_i') = 1 + \epsilon'$, so we need each ϵ_i' to be about ϵ'/n.

Suppose Stockmeyer's algorithm gives output $\tilde{q}_a$, which is within a factor $1 + \epsilon'$ of q_a. Now, by the anticoncentration bound, with probability at least $1 - \delta_A$, $|\langle a|U|0\rangle|^2 = p_a \geq \mu/2^n$. When this happens, then q_a is within a factor $1/(1 - \epsilon/\mu)$ of p_a, and so $\tilde{q}_a$ is within a factor $1 + \epsilon_R = (1 + \epsilon')/(1 - 2\epsilon/\mu)$ of p_a. Stockmeyer's algorithm could also be used to calculate marginal probabilities $\tilde{q}_a$ for values of a on only a subset of bits, and the marginal probabilities $\tilde{q}_a$ have to be within a factor $1/(1 - \epsilon/\mu)$ of the quantum marginal probabilities p_a by the same argument.

But this algorithm is a strong simulation of U for at least a fraction $1 - \delta - \delta' - \delta_A = 1 - \delta_R$ of U's with multiplicative error $1 + \epsilon_R$ (using anticoncentration and the circuit obfuscation property), which by the average case #P-hardness assumption implies the existence of a classical randomized algorithm to solve any #P problem. This same algorithm can then be used to solve any problem in $\text{P}^{\#\text{P}}$, which includes the polynomial hierarchy. The algorithm is in BPP^{NP}, so this implies that the polynomial hierarchy collapses to the 3rd level. □