CSE 291 / Math 277A - Quantum Complexity Theory October 20, 2022

Lecture 8

Lecturer: Daniel Grier *Scribe: Daniel Grier*

1 Collision vs Simon's problem

A function $f\colon \{0,1\}^n \to \{0,1\}^n$ is *1-to-1* (also known as injective) if it maps distinct elements in its domain to distinct elements in its image, or in other words, $f(x) = f(y)$ implies that $x = y$. We say that $f\colon \{0,1\}^n \to \{0,1\}^n$ is *2-to-1* if every element in its image has exactly two preimages, or in other words, if $f(x) = c$, then there is exactly one other $y \neq x$ such that $f(x) = f(y) = c$.

Collision:
Oracle: $f\colon \{0,1\}^n \to \{0,1\}^n$
Goal: decide if f is 1-to-1 or 2-to-1, promised that one is true

Question. *What's the quantum query complexity of* Collision*?*

Notice that Collision is a generalization of Simon's problem, where we have removed the extra condition that if $f(x) = f(y)$ then $x \oplus y = s$ for some secret string $s \in \{0,1\}^n$. Can we use the same analysis? Let's try the same algorithm and see what happens:

$$|0^n\rangle \xrightarrow{H^{\otimes n}} \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle \xrightarrow{O_f} \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle\,|f(x)\rangle \xrightarrow{\text{measure}} \begin{cases} \frac{|x\rangle|c\rangle+|y\rangle|c\rangle}{\sqrt{2}} & \text{if 2-to-1} \\ |x\rangle\,|c\rangle & \text{if 1-to-1} \end{cases}$$

where $c \in \{0,1\}^n$ is the random outcome obtained by measuring the second register. So, dropping the measured register we just have to decide if we are given a superposition of basis states (x and y) or just a single basis state (x). Should such a task be easy? If we had a few copies of the state, then yes! Just measure each copy. When the function is 2-to-1, then we will see both x and y with high probability and so we know that there are multiple inputs that correspond to the outcome c. If the function were 1-to-1, then we would only see the single basis state x.

Of course, we don't actually have multiple copies of the state. We only have 1. What's worse, if we try to perform the same sequence of steps from the beginning, we measure the same c with exponentially small probability, and it's very unclear what information you would get until you measure the same outcome twice (i.e., see a collision). Moreover, the same picture holds if we follow the outline of Simon's algorithm and proceed to apply another layer of Hadamard gates. We seem to be stuck, which is good because...

Theorem 1 (Shi [Shi02])**.** *The quantum query complexity of* Collision *is at least* $\Omega(2^{n/3})$.

This reinforces our intuition that quantum computation is quite subtle, and to obtain a quantum speedup, the problem we are trying to solve requires some kind of structure that the quantum computer can take advantage of. Notice however, that it seems like we are actually getting a small improvement here over the $O(2^{n/2})$ classical randomized algorithm (which, once again, relies on the Birthday Paradox to see a collision). Indeed, there is a quantum algorithm that matches the lower bound of Shi.

Theorem 2 (Brassard, Høyer, and Tapp [BHT97])**.** *The quantum query complexity of* Collision *is* $O(2^{n/3})$.

We will see how to prove this later.

2 Forrelation

In the last lecture, we saw Simon's problem, which required $\Omega(2^{n/2})$ classical queries to solve (even with randomized algorithms), but only required $O(n)$ queries to solve quantumly. How far can we push these types of separations?

Question. *What is the largest separation between quantum and classical computers as measured by query complexity?*

Here's one possible approach: take the standard quantum algorithm we've used for the past few problems (e.g., Deutsch-Jozsa, Berstein-Vazirani, and Simon's problems) and define a problem based on that. To be clear, this algorithm is simply 1) start in the all-zeros state 2) apply a layer of Hadamard gates 3) apply the oracle 4) apply another layer of Hadamard gates 5) measure. Before we measure, the state looks like $H^{\otimes n} O_f H^{\otimes n} |0^n\rangle$. Moreover, we're usually only concerned about the amplitude on a particular state. Let Π be the projector onto that state (e.g., in Deutsch-Jozsa, we have $\Pi = |0^n\rangle\langle 0^n|$). In other words, the relevant probability is

$$\langle 0^n| H^{\otimes n} O_f H^{\otimes n} \, \Pi \, H^{\otimes n} O_f H^{\otimes n} |0^n\rangle$$

To define a problem based on the above quantity, we ask: promised that this quantity is either small or large, decide which is the case. While this would define a problem which is hard for classical computers, there is another choice which is has a bit more nice mathematical structure, which is related to the Boolean Fourier transform.

For any Boolean functions[1] $f, g\colon \{0,1\}^n \to \{\pm 1\}$, define their *forrelation* as

$$\phi_{f,g} := \langle 0^n| H^{\otimes n} O_g H^{\otimes n} O_f H^{\otimes n} |0^n\rangle$$

Actually, this is not the standard way to define the forrelation. As we will see, this quantity is equal to the correlation between one function and the Fourier transform to another:

Claim 3. $\phi_{f,g} = \frac{1}{2^{3n/2}} \sum_{x,y \in \{0,1\}^n} f(x)(-1)^{x \cdot y} g(y)$

Proof.

$$\begin{aligned}
|0^n\rangle &\xrightarrow{H^{\otimes n}} \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle \xrightarrow{O_f} \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} f(x)\,|x\rangle \xrightarrow{H^{\otimes n}} \frac{1}{2^n} \sum_{x,y \in \{0,1\}^n} f(x)(-1)^{x \cdot y}\,|y\rangle \\
&\xrightarrow{O_g} \frac{1}{2^n} \sum_{x,y \in \{0,1\}^n} f(x)(-1)^{x \cdot y} g(y)\,|y\rangle \xrightarrow{H^{\otimes n}} \frac{1}{2^{3n/2}} \sum_{x,y,z \in \{0,1\}^n} f(x)(-1)^{x \cdot y} g(y)(-1)^{y \cdot z}\,|z\rangle \\
&\xrightarrow{\langle 0^n|} \frac{1}{2^{3n/2}} \sum_{x,y \in \{0,1\}^n} f(x)(-1)^{x \cdot y} g(y)
\end{aligned}$$

□

So we know that $|\phi_{f,g}|^2$ can be estimated using a two-query quantum algorithm (repeat the sequence of operations in Claim 3 and compute the percentage of runs where you measure the all-zeros state). We're now ready to define the forrelation problem:

> FORRELATION:
> **Oracle:** $f, g\colon \{0,1\}^n \to \{\pm 1\}$
> **Goal:** decide if $|\phi_{f,g}|^2 \geq 2/3$ or $|\phi_{f,g}^2| \leq 1/3$ promised that one is true

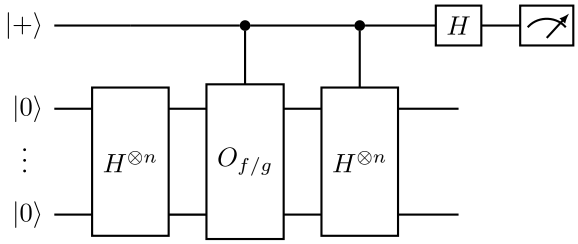
Unfortunately, proving a randomized classical lower bound is quite challenging, but nevertheless one can show that getting more than a quadratic speedup over the naive algorithm is impossible classically:

Theorem 4 (Aaronson and Ambainis [AA15])**.** *The classical randomized query complexity of* FORRELATION *is at least* $\tilde{\Omega}(2^{n/2})$.

This gives us a 2 vs $2^{n/2}$ separation. While nice, it feels like we've lost something from our original idea of using a generic quantum algorithm at the beginning of lecture. Namely, we started with a 1-query algorithm and ended with a 2-query algorithm by generalizing it to have this Fourier structure. Can we compute forrelation with a single query? It turns out that we can. First, we need to combine both functions

[1] For the most part in this class, we've consider Boolean functions where the image is in $\{0,1\}$. We switch to $\{\pm 1\}$ since it is more standard in the Boolean Fourier analysis literature. As we will see, it makes some equations easier to write.

f, g into a single function. This is actually quite simple, since we can design an $(n+1)$-bit controlled-(f, g) gate. If the first bit of the input is 1, apply f; otherwise, apply g. Still, it's a bit unclear how to adapt the original algorithm to this new oracle. The claim is that we measure 0 on the first qubit of the following circuit with probability $\frac{1+\phi_{f,g}}{2}$:



Analyzing the circuit layer-by-layer, we have (dropping n from everything for clarity):

$$|+\rangle|0\rangle \to \frac{|0\rangle + |1\rangle}{\sqrt{2}} H|0\rangle \to \frac{|0\rangle O_g H|0\rangle + |1\rangle O_f H|0\rangle}{\sqrt{2}} \to \frac{|0\rangle O_g H|0\rangle + |1\rangle HO_f H|0\rangle}{\sqrt{2}}$$
$$\to \frac{(|0\rangle + |1\rangle)O_g H|0\rangle + (|0\rangle - |1\rangle)HO_f H|0\rangle}{2}$$

Recall that the probability we measure 0 is equal to the ℓ_2-norm of the vector after we project the first qubit onto the 0 state. In our case, the projection leaves us with the state:

$$\frac{O_g H + HO_f H}{2}|0\rangle,$$

so, using the fact that all the individual matrices above are Hermitian, the probability we measure 0 is

$$\langle 0| \left(\frac{HO_g + HO_f H}{2}\right)\left(\frac{O_g H + HO_f H}{2}\right)|0\rangle = \frac{1+\phi_{f,g}}{2}$$

where we've used $H^2 = O_f^2 = O_g^2 = I$ and the definition of forrelation.

Using the above construction, we can give a (slightly different) version of the forrelation problem which only requires 1 quantum query to solve, but nevertheless require $\tilde{\Omega}(2^{n/2})$ classical randomized queries. Unfortunately, this is as far as we can push separations based on the forrelation problem. What happens if we move to a setting which require more quantum queries? To do this we will need to generalize forrelation.

3 k-fold forrelation

The k-fold forrelation generalizes the original forrelation from 2 oracle function calls to k calls:

$$\phi(f_1, f_2, \ldots, f_k) = \langle 0^n| H^{\otimes n} O_{f_k} H^{\otimes n} \cdots H^{\otimes n} O_{f_1} H^{\otimes n} |0^n\rangle$$

for functions $f_1, \ldots, f_k \colon \{0,1\}^n \to \{\pm 1\}$. Once again, the k-fold forrelation problem asks whether or not $|\phi(f_1, f_2, \ldots, f_k)|^2$ is greater than 2/3 or smaller than 1/3. Much like the original "2-fold forrelation" problem could be solved with one quantum query, the k-fold forrelation problem can be solved with $\lceil k/2 \rceil$ queries using the exact same proof idea.

Determining a lower bound on the randomized classical query complexity of k-fold forrelation remained elusive until it was settled by Bansal and Sinha:

Theorem 5 (Bansal and Sinha [BS21]). *The k-fold forrelation problem has randomized classical query complexity at least $\tilde{\Omega}(2^{n(1-1/k)})$.*

Independently, a similar lower bound was shown for a slightly different problem by Sherstov, Storozhenko, and Wu [SSW21]. Notice that when $k = 2$, we recover the $\tilde{\Omega}(\sqrt{2^n})$ lower bound of Aaronson and Ambainis for FORRELATION.

For this rest of this lecture, we will show that this is the best possible separation. There *is* a classical randomized algorithm that can solve k-fold forrelation with roughly $2^{n(1-1/k)}$ queries. In fact, we will show something slightly more general:

Theorem 6 (Bravyi, Gosset, Grier, Schaeffer [BGGS21]). *Every quantum algorithm that makes t queries can be simulated by a classical randomized algorithm that makes $\widetilde{O}(2^{n(1-1/2t)})$ queries.*

Recall that k-fold forrelation can be solved with $\lceil k/2 \rceil$ quantum queries, so this also gives an optimal algorithm for k-forrelation by setting $t = \lceil k/2 \rceil$.

The first question we need to ask is how to get a handle on an arbitrary quantum query algorithm. To do this, let's make some assumptions about how the quantum algorithm works. First, let's assume that we have a single oracle $\mathcal{O}$. If we have multiple oracles (as in the k-fold forrelation problem), we can combine them into a single oracle that has a control register that determines which oracle is applied. Second, let's assume that the quantum algorithm does not use any ancillary qubits. This will simply make the equations a bit simpler to write, but nothing will break by allowing ancillas. Finally, let's assume only gates with real (rather than complex) entries are used. We've seen before that this does not change the computational power of the circuit. Therefore, before measurement we can represent the state of the quantum algorithm as

$$|\psi\rangle := U_t \mathcal{O} U_{t-1} \cdots U_1 \mathcal{O} U_0 |0^n\rangle .$$

That is, apply an arbitrary unitary, query the oracle, apply another arbitrary oracle, and so on. As we saw earlier for our general 1-query quantum algorithm, we are concerned with the probability on some subset of states specified by a projector Π. We are therefore interested in estimating the quantity

$$\langle\psi| \Pi |\psi\rangle = \langle 0^n| U_0^\dagger \mathcal{O} U_1^\dagger \cdots \mathcal{O} U_t^\dagger \Pi U_t \mathcal{O} \cdots U_1 \mathcal{O} U_0 |0^n\rangle .$$

To reiterate, if a classical algorithm can estimate $\langle\psi| \Pi |\psi\rangle$ to a high accuracy, then it will be able to distinguish between the case where the probability of acceptance of the quantum algorithm is large (say greater than 2/3) or small (smaller than 1/3), and so it can solve the same problem that the quantum algorithm was solving. More generally, we will give a classical algorithm to estimate quantities that look like

$$q := \langle 0^n| M_{2t} \mathcal{O} \cdots M_1 \mathcal{O} M_0 |0^n\rangle$$

where the M_i are arbitrary linear operators that have bounded norm $\|M_i\| \leq 1$ (i.e., the magnitude of the largest eigenvector is bounded by 1). In particular, moving to arbitrary linear operators like this (instead of just unitary operators) allows us to handle the $U_t^\dagger \Pi U_t$ term in the expansion of $\langle\psi| \Pi |\psi\rangle$.

Our goal will be design an estimator $\hat{q}$ for q such that $\mathbb{E}[\hat{q}] = q$ and $\mathrm{Var}[\hat{q}] \leq \epsilon^2/100$. If we can do this, then by Chebyshev's inequality[2] we have $\Pr[|\hat{q} - q| \geq \epsilon] \leq 1/100$. So, setting ϵ to a small constant, this implies that our estimator $\hat{q}$ is good enough to determine the true acceptance probability of the original quantum algorithm (at least with high probability, which is all we care about since this is a randomized algorithm).

To do this, we create a sequence of estimators $|\psi_0\rangle, |\psi_1\rangle, \ldots, |\psi_{2t}\rangle$ where $|\psi_r\rangle$ is a classical representation of the state of the computation after r queries:

$$\mathbb{E}[|\psi_r\rangle] = M_r \mathcal{O} \cdots M_1 \mathcal{O} M_0 |0^n\rangle$$

We design the estimators adaptively, starting by setting $|\psi_0\rangle = M_0 |0^n\rangle$. Given our estimate $|\psi_{r-1}\rangle$ we construct $|\psi_r\rangle$ by sampling from the distribution of outcomes that arise from measuring the *possibly unnormalized* state $|\psi_{r-1}\rangle$. To this end, let's define a distribution

$$p_r(z) := \frac{|\langle z|\psi_r\rangle|^2}{\langle\psi_r|\psi_r\rangle}$$

over bit strings $z \in \{0,1\}^n$ that captures the probability of measuring the rth estimator. Since $|\psi_r\rangle$ might not have norm 1, we must divide by $\langle\psi_r|\psi_r\rangle$ to guarantee that p_r is a valid probability distribution.

[2] Chebyshev's inequality states that for random variable X, we have

$$\Pr\left[|X - \mathbb{E}[X]| \geq \epsilon\right] \leq \frac{\mathrm{Var}[X]}{\epsilon^2}$$

Concretely, to generate the next state estimator we sample L strings $z_1, \ldots, z_L$ independently from the previous estimator:

$$|\psi_r\rangle := M_r\mathcal{O}\left(\frac{1}{L}\sum_{i=1}^{L}\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}\right)|\psi_{r-1}\rangle$$

Intuitively, we are weighting our next estimator towards the basis states that have higher probability in the previous state. A key question is: *why can we simulate this classically?* First, it's important to recall that we are only keeping track of the number of classical queries that our algorithm makes, so in particular we can represent the state $|\psi_r\rangle$ as a giant 2^n-size vector. Furthermore, the M_r terms are known in advance (since they are derived from the quantum circuit), and so we can simulate them just by matrix-vector multiplication. Finally, since we have projected our state onto L basis states, to apply O_f, we only need to know what that oracle does to those L states. Of course, the larger we make L, the more accurate our estimator will be. However, increasing L means that we are increasing the number of classical queries per round, which we are trying to keep low. For now, let's just keep L as a variable. We will minimize it later.

Let's now see why our construction of $|\psi_r\rangle$ gives us a good estimator for our state. By linearity of expectation, it will suffice to analyze a single projection term:

$$\mathbb{E}\left[\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}\right] = \sum_{z\in\{0,1\}^n}\frac{|z\rangle\langle z|}{p_{r-1}(z)}\cdot p_{r-1}(z) = \sum_{z\in\{0,1\}^n}|z\rangle\langle z| = I,$$

and so we have

$$\mathbb{E}[|\psi_r\rangle] = M_r\mathcal{O}\left(\frac{1}{L}\sum_{i=1}^{L}\mathbb{E}\left[\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}\right]\right)|\psi_{r-1}\rangle = M_r\mathcal{O}\,|\psi_{r-1}\rangle$$

as desired. To avoid clutter in the equation, we have avoided explicitly specifying what the expectation is over in the equations. In this particular equation, the expectation is just over the randomness used to sample the z_i from p_{r-1}. Given this state estimator, we can easily define an estimator for the acceptance probability of the entire quantum algorithm:

$$q_r := \langle 0^n|\, M_{2t}\mathcal{O}\cdots M_{r+1}\mathcal{O}\,|\psi_r\rangle$$

represents the acceptance probability of the algorithm after r queries using the rth state estimator. Notice that $q_0 = q$ is the true acceptance probability of the quantum algorithm, and $\hat{q} := q_{2t}$ is our estimate of this acceptance probability after $2t$ rounds of sampling bit strings. To reiterate, q_{2t} is something that we have computed classically, while q_0 is just a quantity that we will use in our analysis (it is not known to the classical algorithm). Our goal will be to show a type of hybrid argument. Namely, we will show that q_r is close enough to q_{r-1} so that adding up all the differences between q_{2t} and q_0 is still small.

From the expectation calculation for each of our state estimators, we have

$$\mathbb{E}[q_r] = \langle 0^n|\, M_{2t}\mathcal{O}\cdots M_{r+1}\mathcal{O}\mathbb{E}[|\psi_r\rangle] = \langle 0^n|\, M_{2t}\mathcal{O}\cdots M_{r+1}\mathcal{O}M_r\mathcal{O}\cdots\mathcal{O}M_0\,|0^n\rangle = q$$

where the expectation is over all of the randomness used so far during the computation. For now let us simply state what the variance is and we will prove it later:

Lemma 7.

$$\mathrm{Var}[q_r] \le 2^{-n}\left(1+\frac{2^n}{L}\right)^r$$

To complete the proof of Theorem 6, recall that we need that $\mathrm{Var}[\hat{q}] = \mathrm{Var}[q_{2t}] \le \epsilon^2/100$ to apply Chebyshev's inequality. In other words, we get a sufficiently accurate randomized algorithm whenever the following holds:

$$2^{-n}\left(1+\frac{2^n}{L}\right)^{2t} \le \frac{\epsilon^2}{100}$$

Solving for L shows that $L = O(\epsilon^{-1/t}2^{n(1-\frac{1}{2t})})$ queries in each round are sufficient for a small variance. Since there are $2t$ many rounds, we get that there are $O(t\epsilon^{-1/t}2^{n(1-\frac{1}{2t})})$ total queries in the randomized algorithm, which finishes the proof.

proof of Lemma 7: Recall that for random variable X, we have $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. Therefore, we will need to compute

$$\mathbb{E}[q_r^2] = \mathbb{E}[q_r q_r^*] = \langle 0^n| M_{2t}\mathcal{O}\cdots\mathcal{O}M_{r+1}|\psi_r\rangle\langle\psi_r| M_{r+1}^\dagger\mathcal{O}\cdots\mathcal{O}M_{2t}^\dagger|0^n\rangle$$

where the first equality follows from the fact that the quantum computation only uses real numbers. This necessitates the following computation:

$$\mathbb{E}[|\psi_r\rangle\langle\psi_r|] = M_r\mathcal{O}\left(\frac{1}{L^2}\sum_{i,j=1}^{L}\mathbb{E}\left[\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}|\psi_{r-1}\rangle\langle\psi_{r-1}|\frac{|z_j\rangle\langle z_j|}{p_{r-1}(z_j)}\right]\right)\mathcal{O}M_r^\dagger.$$

There are two cases to consider for the inner expectation. If $i \neq j$, then samples z_i and z_j are independent, and so we just get the expectation of each term, which we previously calculated to be the identity:

$$\mathbb{E}\left[\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}|\psi_{r-1}\rangle\langle\psi_{r-1}|\frac{|z_j\rangle\langle z_j|}{p_{r-1}(z_j)}\right] = \mathbb{E}\left[\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}\right]|\psi_{r-1}\rangle\langle\psi_{r-1}|\,\mathbb{E}\left[\frac{|z_j\rangle\langle z_j|}{p_{r-1}(z_j)}\right] = |\psi_{r-1}\rangle\langle\psi_{r-1}|$$

If $i = j$, we get

$$\begin{aligned}
\mathbb{E}\left[\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}|\psi_{r-1}\rangle\langle\psi_{r-1}|\frac{|z_i\rangle\langle z_i|}{p_{r-1}(z_i)}\right] &= \sum_{z\in\{0,1\}^n}\left(\frac{|z\rangle\langle z|}{p_{r-1}(z)}|\psi_{r-1}\rangle\langle\psi_{r-1}|\frac{|z\rangle\langle z|}{p_{r-1}(z)}\right)p_{r-1}(z)\\
&= \sum_{z\in\{0,1\}^n}\frac{|z\rangle\langle z|\psi_{r-1}\rangle\langle\psi_{r-1}|z\rangle\langle z|}{p_{r-1}(z)}\\
&= \langle\psi_{r-1}|\psi_{r-1}\rangle\sum_{z\in\{0,1\}^n}|z\rangle\langle z|\\
&= \langle\psi_{r-1}|\psi_{r-1}\rangle I
\end{aligned}$$

Given that there are L terms with $i = j$ and $L^2 - L$ terms with $i \neq j$, we get

$$\mathbb{E}[|\psi_r\rangle\langle\psi_r|] = \frac{L-1}{L}M_r\mathcal{O}|\psi_{r-1}\rangle\langle\psi_{r-1}|\mathcal{O}M_r^\dagger + \frac{\langle\psi_{r-1}|\psi_{r-1}\rangle}{L}M_rM_r^\dagger$$

Putting everything together, we have

$$\begin{aligned}
\mathbb{E}[q_r^2] &= \frac{L-1}{L}q_{r-1} + \frac{\langle\psi_{r-1}|\psi_{r-1}\rangle}{L}\langle 0^n|M_{2t}\mathcal{O}\cdots\mathcal{O}M_{r+1}M_rM_r^\dagger M_{r+1}^\dagger\mathcal{O}\cdots\mathcal{O}M_{2t}^\dagger|0^n\rangle\\
&\leq q_{r-1}^2 + \frac{\langle\psi_{r-1}|\psi_{r-1}\rangle}{L}
\end{aligned}$$

where we have used that the M_i terms cannot increase the ℓ_2 norm (i.e., $\|M_i\| \leq 1$). Evidently, to compute the expectation of q_r^2, we need a bound on the magnitude of the $|\psi_r\rangle$ states. Fortunately, this can easily be computed as

$$\mathbb{E}[\langle\psi_r|\psi_r\rangle] = \mathrm{tr}\left(\mathbb{E}[|\psi_r\rangle\langle\psi_r|]\right) \leq \left(1 + \frac{2^n}{L}\right)\langle\psi_{r-1}|\psi_{r-1}\rangle$$

reusing our computation of $\mathbb{E}[|\psi_r\rangle\langle\psi_r|]$ above (and once again relying on the fact that $\|M_r\| \leq 1$). Now we recursively apply this identity in our equation for $\mathbb{E}[q_r^2]$ to obtain

$$\mathbb{E}[q_r^2] - \mathbb{E}[q_{r-1}^2] \leq \frac{1}{L}\left(1 + \frac{2^n}{L}\right)^{r-1}$$

where the expectations are over all the randomness used in the algorithm.

We are now finally in a position to apply our hybrid argument to bound the variance:

$$\mathrm{Var}[q_r] = \mathbb{E}[q_r^2] - \mathbb{E}[q_r]^2 = \mathbb{E}[q_r^2] - \mathbb{E}[q_0^2] \leq \sum_{i=1}^{r}\left(\mathbb{E}[q_i^2] - \mathbb{E}[q_{i-1}^2]\right) = \frac{1}{L}\sum_{i=0}^{r-1}\left(1 + \frac{2^n}{L}\right)^i$$

where we've used that $\mathbb{E}[q_r]^2 = q^2 = q_0^2 = \mathbb{E}[q_0^2]$. To bound the right hand side of the equation, we note that it just a geometric series $\sum_{i=0}^{r-1} g^i = \frac{g^r-1}{g-1}$ with common ratio $g := \left(1 + \frac{2^n}{L}\right)$. We get

$$\mathrm{Var}[q_r] \leq 2^{-n}\left(\left(1 + \frac{2^n}{L}\right)^r - 1\right) \leq 2^{-n}\left(1 + \frac{2^n}{L}\right)^r,$$

completing the proof. □

References

[AA15] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 307–316, 2015.

[BGGS21] Sergey Bravyi, David Gosset, Daniel Grier, and Luke Schaeffer. Classical algorithms for forrelation. *arXiv preprint arXiv:2102.06963*, 2021.

[BHT97] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum algorithm for the collision problem. *arXiv quant-ph/9705002*, 1997.

[BS21] Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1303–1316, 2021.

[Shi02] Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 513–519. IEEE, 2002.

[SSW21] Alexander A Sherstov, Andrey A Storozhenko, and Pei Wu. An optimal separation of randomized and quantum query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1289–1302, 2021.