Frontiers of Quantum complexity *Instructor: Anand Natarajan, Anurag Anshu*

Lecture 1: Quantum Merlin-Arthur class

Sept 4, 2025 *TF: Yeongwoo Hwang, Tina Zhang*

1 Quantum Merlin-Arthur protocol

We introduced a QMA protocol informally in Lecture slides. Let's make an attempt to define it more precisely. On input $|x\rangle$, some ancilla work register $|0\rangle_A$, and a register that holds the witness $|\psi\rangle_W$, the verifier runs a polynomial sized quantum circuit Q. A qubit O is measured in $\{|0\rangle, |1\rangle\}$ basis and the verifier outputs 1 iff the outcome is $|1\rangle$. Thus, the probability of outputting 1 is $\|(|1\rangle\langle 1|_O \otimes \mathbb{1})Q(|x\rangle |0\rangle_A \otimes |\psi\rangle_W)\|^2$. The maximum probability of success - over all possible witnesses $|\psi\rangle$ can be expressed as

$$\begin{aligned}&\max_{\psi} \left(\langle x| \otimes \langle 0|_A \otimes \langle\psi|_W\right) Q^\dagger (|1\rangle\langle 1|_O \otimes \mathbb{1}) Q \left(|x\rangle \otimes |0\rangle_A \otimes |\psi\rangle_W\right) \\ &= \|(\langle x| \otimes \langle 0|_A \otimes \mathbb{1}_W) Q^\dagger (|1\rangle\langle 1|_O \otimes \mathbb{1}) Q (|x\rangle \otimes |0\rangle_A \otimes \mathbb{1})\|_\infty. \end{aligned} \tag{1}$$

where $\|\cdot\|_\infty$ denotes the operator norm. Thus, the prover should simply send the eigenstate corresponding to the largest eigenvalue of the operator on the LHS, which we denote as

$$V_x^1 := (\langle x| \otimes \langle 0|_A \otimes \mathbb{1}_W) Q^\dagger (|1\rangle\langle 1|_O \otimes \mathbb{1}) Q (|x\rangle \otimes |0\rangle_A \otimes \mathbb{1}).$$

We also define

$$V_x^0 := (\langle x| \otimes \langle 0|_A \otimes \mathbb{1}_W) Q^\dagger (|0\rangle\langle 0|_O \otimes \mathbb{1}) Q (|x\rangle \otimes |0\rangle_A \otimes \mathbb{1}).$$

Note that $V_x^0 + V_x^1 = \mathbb{1}_W$.

The fact that quantum prover can send entangled states can make one nervous about the class QMA. Does QMA admit soundness amplification, like the nice classical classes such as BPP or the quantum class BQP? Does verifier need to do very non-trivial (still polynomial time) quantum computation to prevent being fooled by the prover? Remarkably, the 'lease' is not all in prover's hand and verifier has significant control. We will see that in the following section.

2 Verifier's power in QMA protocols

2.1 Soundness amplification of QMA protocols

In the above description of QMA, we set the probability of acceptance to be 2/3 (when $f(x) = 1$) and 1/3 (when $f(x) = 0$) somewhat arbitrarily. A common CS theme is soundness amplification - to repeat the protocol in parallel and take majority vote to improve these numbers much closer to 1 or 0. The view that the prover is simply trying to hit the largest eigenvalue of V_x^1 gives a simple proof that the amplification of QMA works. Lets introduce the majority vote protocol formally.

- Let k be a natural number. Verifier initializes k copies of the ancilla and k copies of the input $|x\rangle$ in the state $|x\rangle^{\otimes k} |0\rangle_A^{\otimes k}$.

- Prover sends a quantum state on k copies of W.
- Verifier runs k parallel copies of Q_x and then measures all the bits $O_1, \ldots O_k$. Let the outcomes be $o_1, \ldots o_k$.
- Verifier outputs 1 iff $o_1 + \ldots o_k \geq \frac{k}{2}$.

The success probability of the majority vote protocol across k runs on a witness $|\psi\rangle_{W_1 \ldots W_k}$ can be estimated as follows:

$$\sum_{\substack{o_1,o_2,\ldots o_k \text{ s.t.}\\ o_1+o_2+\ldots o_k \geq \frac{k}{2}}} \|(|o_1, \ldots o_k\rangle\langle o_1, \ldots o_k| \otimes \mathbb{1}) Q^{\otimes k} (|x\rangle^{\otimes k} |0\rangle_A^{\otimes k} \otimes |\psi\rangle_{W_1 \ldots W_k})\|^2$$

$$= \langle\psi|_{W_1,\ldots W_k} \left(\sum_{\substack{o_1,o_2,\ldots o_k \text{ s.t.}\\ o_1+o_2+\ldots o_k \geq \frac{k}{2}}} V_x^{o_1} \otimes \ldots V_x^{o_k} \right) |\psi\rangle_{W_1,\ldots W_k}.$$

Thus, we need to understand the largest eigenvalue of the operator

$$\left(\sum_{o_1,\ldots o_k \in \{0,1\} : \sum_i o_i \geq k/2} V_x^{o_1} \otimes \ldots V_x^{o_k} \right).$$

Since V_x^0, V_x^1 mutually commute, all the operators in the sum commute. So the maximum success probability is

$$\left\| \sum_{\substack{b_1,\ldots b_k \in \{0,1\} \text{ s.t.}\\ \sum_i b_i \geq k/2}} V_x^{b_1} \cdots V_x^{b_k} \right\|_\infty = \sum_{\substack{b_1,\ldots b_k \in \{0,1\} \text{ s.t.}\\ \sum_i b_i \geq k/2}} \|V_x^{b_1}\|_\infty \otimes \ldots \|V_x^{b_k}\|_\infty.$$

Lets consider the case $f(x) = 0$, where recall that $\|V_x^1\| \leq 1/3$. So the maximum success probability is

$$\begin{aligned}
\left\| \sum_{\substack{b_1,\ldots b_k \in \{0,1\} \text{ s.t.}\\ \sum_i b_i \geq k/2}} V_x^{b_1} \otimes \ldots V_x^{b_k} \right\| &\leq \sum_{\substack{b_1,\ldots b_k \in \{0,1\} \text{ s.t.}\\ \sum_i b_i \geq k/2}} \|V_x^{b_1}\| \cdots \|V_x^{b_k}\| \\
&= \sum_{\substack{b_1,\ldots b_k \in \{0,1\} \text{ s.t.}\\ \sum_i b_i \geq k/2}} (\|V_x^1\|)^{\sum_i b_i} (\|V_x^0\|)^{n - \sum_i b_i} \\
&= e^{-\Omega(k)},
\end{aligned}$$

where the last line uses a Chernoff bound and the fact that $\|V_x^0\| \geq 2/3$. In the yes case, the largest eigenvalue can be lower bounded in a similar manner.

Note that we are crucially using the fact that V_x^0, V_x^1 commute. This would not be needed if $\|V_x^0\| \leq \frac{1}{5}$ in the no case. However, amplification should also work if $\|V_x^0\| \approx \frac{1}{2} - \frac{1}{\text{poly}(n)}$, in which case the above approach seems necessary.

2.2 Soundness amplification from one witness

A remarkable observation due to Marriot and Watrous is that a sequential repetition using only one witness suffices for amplification.

Define the projectors $\Pi_1 = Q^\dagger(|1\rangle\langle 1|_O \otimes \mathbb{1})Q$ and $\Pi_2 = \mathbb{1}_A \otimes |x\rangle\langle x| \otimes |0\rangle\langle 0|_A$ (dropping the label x for convenience). Then comparing with eq. (1), the maximum success probability - denoted p_x - is equivalently the largest eigenvalue of $\Pi_2\Pi_1\Pi_2$. Invoking Jordan's lemma (see Section 1, Pages 2-4 of Regev's lecture), consider the two dimensional subspace where the component of Π_2 is $|w\rangle := |\psi_x\rangle_W \otimes |x\rangle \otimes |0\rangle_A$ (where $|\psi_x\rangle_W$ achieves the eigenvalue p_x for $\Pi_2\Pi_1\Pi_2$). Let the vector orthogonal to it be $|w^\perp\rangle$ and $|v\rangle$ be the component of Π_1, with the orthogonal vector $|v^\perp\rangle$ (Figure 1).

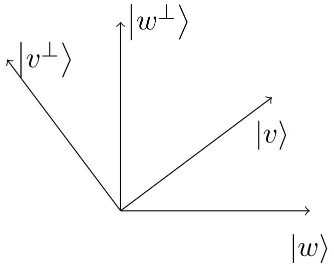

Figure 1: Vectors appearing in the Marriott-Watrous amplification. The overlap between $|v\rangle, |w\rangle$ is $\sqrt{p_x}$.

Consider the following algorithm $\mathcal{A}_1$.

- Measure according to $\{\Pi_1, \mathbb{1} - \Pi_1\}$ and then according to $\{\Pi_2, \mathbb{1} - \Pi_2\}$.

The action of the algorithm is depicted in Figure 2, which is immediate from Figure 1.

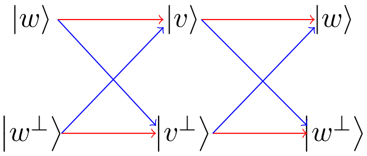

Figure 2: Action of the algorithm $\mathcal{A}_1$ on the input states $|w\rangle$ and $|w^\perp\rangle$. The red transitions occur with probability p_x and the blue transitions occur with probability $1 - p_x$.

The important point is that because Jordan's Lemma breaks down the full space into subspaces of dimension at most 2, the operators Π_1 and Π_2 keep us within that 2 dimensional space. At each round, the algorithm $\mathcal{A}_1$ essentially flips a $\{p_x, 1 - p_x\}$-biased coin (with heads being the first outcome). We repeat it $N = O(\frac{1}{(c-s)^2})$ times and accept if there are more than $N\frac{c+s}{2}$ "heads". Else we reject. Once again applying a Chernoff bound, this achieves the desired amplification.

2.3 QMA verification in logarithmic depth

A nice fact about famous quantum algorithms - such as Shor's algorithm and quantum Fourier transform - is that they can run in relatively low depth (logarithmic) with clever parallelization. But its not known if all problems in BQP can be solved in log depth.

We will see here that a QMA verifier can run in logarithmic depth and in fact there much more structure to the verification. This is the power of the soundness condition. We will use Rosgen's protocol (and later in the course look at another way to do so, using the detectability lemma). The idea is to consider any verification protocol and ask the prover to share all the quantum states at all times of the computation. The states at time t and $t+1$ are compared via a multi-qubit swap test (conjugating out the action of depth one unitary between these times). The multi-qubit swap test can itself be performed by preparing a CAT state, so the overall protocol requires only logarithmic depth.