Problem Set 1:
Quantum Learning and Complexity Theory (Summer 2025)

Junseo Lee

July 21, 2025

Problem 1. To better prepare for future topics on quantum learning algorithms, it is important to become familiar with how classical learning models extend to the quantum setting. The following problems are designed to help you reflect on and articulate the key differences and implications of quantum learning models.

For each of the following classical learning models, we provide a brief description. Your task is to define the corresponding **quantum learning model**, highlighting the key differences. Specifically, describe:

- the type of oracle access (classical vs. quantum),
- the nature of the data provided to the learner,
- how hypotheses are produced,
- any differences in complexity notions (e.g., sample or query complexity).

We provide the example of exact learning to illustrate how to answer the question.

Exact Learning: The learner has access to a membership oracle $\mathsf{MQ}(c)$ for the target concept $c \in C$. Given input $x \in \{0,1\}^n$, the oracle returns $c(x)$. The learner must output a hypothesis h such that $h(x) = c(x)$ for all x, with probability at least $2/3$.

Quantum analogue: In the quantum setting, the learner has access to a quantum membership oracle $\mathsf{QMQ}(c)$, which performs the unitary transformation $\mathsf{QMQ}(c) : |x, b\rangle \mapsto |x, b \oplus c(x)\rangle$ for all $x \in \{0,1\}^n$ and $b \in \{0,1\}$. The learner starts with an initial quantum state, makes a sequence of queries to the quantum oracle, and finally performs a measurement to produce a classical hypothesis h. The quantum query complexity is the number of calls to the QMQ oracle needed to exactly identify c (i.e., produce $h = c$ with high probability), optimized over all quantum learners.

(a) **PAC Learning:** The learner has access to a random example oracle $\mathsf{PEX}(c, D)$ that returns pairs $(x, c(x))$, where x is drawn from an unknown distribution D over $\{0,1\}^n$. The learner must output a hypothesis h such that with probability at least $1-\delta$, $\Pr_{x\sim D}[h(x) \neq c(x)] \leq \varepsilon$. *Define the quantum analogue of this model.*

(b) **Agnostic Learning:** The learner has access to a labeled example oracle $\mathsf{AEX}(D)$ that returns samples (x, b) drawn from a distribution D over $\{0,1\}^{n+1}$. The learner must output a hypothesis h such that with probability at least $1-\delta$, $\mathsf{err}_D(h) \leq \min_{c\in C} \mathsf{err}_D(c) + \varepsilon$, where $\mathsf{err}_D(h) := \Pr_{(x,b)\sim D}[h(x) \neq b]$. *Define the quantum analogue of this model.*

Problem 2. In class, we discussed the fundamental difference between *fully entangled measurements* and *incoherent measurements* in quantum state tomography. Fully entangled measurements (e.g., Schur sampling on n copies) achieve optimal copy complexity $n = \Theta(d^2/\epsilon^2)$, but require joint access to all n copies of the state, which is infeasible on near-term devices. Incoherent measurements, on the other hand, measure one copy at a time and are experimentally feasible, but require a higher copy complexity of $n = \Theta(d^3/\epsilon^2)$. Recent works suggest that a more realistic and practical model lies between these extremes: *t-entangled measurements*, which allow measurements on t copies at a time. Surprisingly, it was not known until recently whether such intermediate levels of entanglement could yield meaningful improvements in sample complexity. (Note: I recommend that you independently search for and study research papers related to this topic.)

(a) Explain why incoherent measurements are easier to implement experimentally but lead to worse copy complexity. Provide two reasons based on physical and algorithmic considerations.

(b) Describe one reason why it is difficult to adapt techniques from fully entangled tomography (e.g., Schur sampling) to the intermediate t-entangled regime.

(c) Suppose you can only measure $t = 2$ copies of the state ρ at a time. Why is it nontrivial to determine whether this offers an advantage over incoherent ($t = 1$) measurements?

(d) Consider the following interpolation question: For fixed dimension d and error ϵ, how do you expect the copy complexity n of t-entangled tomography to scale with t in the range $1 \leq t \leq d^2$? Justify your intuition qualitatively, not quantitatively.

(e) In your own words, discuss why understanding the intermediate t-entangled regime is important both theoretically and practically for near-term quantum computing platforms.

Problem 3. To prepare for future lectures, it is important to become familiar with the basics of quantum query complexity. In class, we also discussed success probabilities of quantum algorithms and how they can be amplified using standard boosting techniques. The following problem is designed to help reinforce these foundational ideas.

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Suppose we have a randomized (or quantum) algorithm that computes $f(x)$ correctly with constant success probability, say at least $2/3$, using $R(f)$ (or $Q(f)$) queries. Now, we want to reduce the error probability of the algorithm to some small value ε, i.e., we want the success probability to be at least $1 - \varepsilon$ on all inputs. The question is: *how many queries are required to achieve this higher accuracy?* Prove the following:

(a) For randomized algorithms, the number of queries needed to reduce the error to ε is $\mathcal{O}(R(f)\log(1/\varepsilon))$.

(b) For quantum algorithms, the number of queries needed to reduce the error to ε is $\mathcal{O}(Q(f)\log(1/\varepsilon))$.

In other words, by repeating the base algorithm $\mathcal{O}(\log(1/\varepsilon))$ times and appropriately aggregating the outputs (e.g., via majority vote or other amplification techniques), we can achieve error probability ε with only a logarithmic overhead in query complexity.

Problem 4. In class, we have discussed the basic ideas behind deriving lower bounds in learning settings. Building on those ideas, as well as the references provided during the lectures, you are encouraged to solve the following problem. While the specific technical tools will be covered in more detail in later lectures, attempting this problem using the concepts from class and exploring relevant literature will be a valuable exercise. (Note: No specific paper is pointed out intentionally, as part of the learning process is to explore and identify useful references on your own.)

Let $O_1, \ldots, O_M$ be a set of traceless Hermitian observables acting on n-qubit systems, each with operator norm $\|O_i\|_\infty = 1$, and assume the symmetry condition: $O_i = -O_{i+M/2}$, for all $i = 1, \ldots, M/2$. Moreover, each O_i has eigenvalues in $\{\pm 1\}$. Define the *correlation complexity* of this observable set as:

$$\delta(O_1, \ldots, O_M) := \sup_{|\phi\rangle} \frac{2}{M} \sum_{i=1}^{M/2} \langle \phi | O_i | \phi \rangle^2. \tag{1}$$

Suppose a learning algorithm *without quantum memory* attempts to learn the expectation values $\mathrm{tr}(O_i \rho)$ to within additive error ε, for all $i \in \{1, \ldots, M\}$, with success probability at least $2/3$. Let the number of copies of ρ used by the algorithm be T.

(a) Show that for the special family of states

$$\rho_x = \frac{1 + 3\varepsilon O_x}{2^n}, \tag{2}$$

the expectation value $\mathrm{tr}(O_x \rho_x) = 3\varepsilon$, while $\mathrm{tr}\left(O_x \cdot \frac{I}{2^n}\right) = 0$.

(b) Show that, under the assumption $\varepsilon < 0.29$, any learning algorithm *without quantum memory* requires at least

$$T \geq \Omega\left(\frac{1}{\varepsilon^2 \delta(O_1, \ldots, O_M)}\right) \tag{3}$$

copies of ρ, by performing a many-versus-one distinguishing reduction and applying Jensen's inequality.

(c) Intuitively explain the role of $\delta(O_1, \ldots, O_M)$ in characterizing the hardness of shadow tomography.