PSET 2

Due on **Friday, October 31st, 2025 at 11:59 pm PT**

Ph 220: Quantum Learning Theory, Fall 2025 (Hsin-Yuan Huang) Version: **1.0**

This problem set covers concepts from learning and predicting properties of a quantum state. The questions are designed to be challenging, so please do not feel discouraged if you get stuck or are unable to solve some of them. If you find yourself running low on time, we recommend aiming for breadth rather than depth: it is better to complete a few parts of each question than to solve one question completely while skipping others. Hints are provided. While these hints may help you solve the problems more easily, you are not required to follow them as long as your proofs are correct.

1 (25 pts.) Learning a low-rank state

Motivation: In this problem, we will develop the optimal single-copy quantum state tomography for learning n-qubit states up to ϵ trace distance using $\mathcal{O}(8^n/\epsilon^2)$ random measurements. We will establish some facts about subexponential random variables, and refine this bound under the additional assumption that the unknown state has rank r.

Setup: We measure N copies of the unknown state ρ with a randomized measurement: evolve under a Haar-random unitary U and measure the evolved state $U\rho U^\dagger$ in the Z basis to obtain $|b\rangle \in \{0,1\}^n$. Let each randomized measurement be labeled by $U_i \in \mathbb{U}(2^n)$ and $|b_i\rangle \in \{0,1\}^n$, and define

$$\hat{\rho} \triangleq \frac{1}{N}\sum_{i=1}^{N}\left((2^n+1)U_i^\dagger |b_i\rangle\langle b_i| U_i - \mathbb{I}\right).$$

In the first part, we will prove the concentration inequality in Eq. (2) using tools from probability theory and use the concentration inequality to derive the $\mathcal{O}(8^n/\epsilon^2)$ sample complexity bound.

A random variable Z with mean zero is said to be $O(1)$-subexponential if $\mathbb{E}[|Z|^k] \leqslant O(k)^k$ for all powers $k \geqslant 1$. You may use the *subexponential Bernstein inequality*, which states that any collection of independent $O(1)$-subexponential random variables $X_1, \ldots, X_N$ with mean zero and $\epsilon < 1$ satisfies the tail bound

$$\mathbf{Pr}\Big[\Big|\frac{1}{N}\sum_i X_i\Big| > \epsilon\Big] \leqslant \exp\left(-N\Omega\left(\epsilon^2\right)\right).$$

1.A. (5 pts.) Prove the inequality

$$\mathbb{E}_{U,b}\left[\left((2^n+1)|\langle u|U^\dagger|b\rangle|^2\right)^k\right] \leqslant O(k)^k, \tag{1}$$

where $|u\rangle$ is an arbitrary fixed unit vector, and the expectation is taken with respect to U sampled from the Haar measure and $|b\rangle \in \{0,1\}^n$ obtained from measuring $U\rho U^\dagger$ in the Z basis.

Hint: Analyze the nt-qubit density matrix $\int (U|u\rangle\langle u|U^\dagger)^{\otimes t} d\mu_H(U)$ using the Schur-Weyl duality. Prove that $\int (U|u\rangle\langle u|U^\dagger)^{\otimes t} d\mu_H(U) = \frac{1}{t!}\binom{2^n+t-1}{t}^{-1}\sum_\sigma \sigma$, where σ is a permutation over t copies of n-qubit systems.

1.B. (5 pts.) Prove that Eq. (1) implies that for any unit vector $|\psi\rangle$, each of the random variables

$$\langle\psi|\left((2^n+1)U_i^\dagger|b_i\rangle\langle b_i|U_i - \mathbb{I} - \rho\right)|\psi\rangle$$

for the measurement outcomes $U_i, |b_i\rangle$ is $O(1)$-subexponential.

1.C. (5 pts.) Conclude from Part **1.B.** and the subexponential Bernstein inequality that for any unit vector $|\psi\rangle$ and any $0 < \varepsilon_1 < 1$,

$$\mathbf{Pr}\Big[|\langle\psi|(\rho-\hat{\rho})|\psi\rangle| > \varepsilon_1\Big] \leqslant \exp\left(-N\Omega\left(\varepsilon_1^2\right)\right). \tag{2}$$

Recall from PSET 1: 3.C., we have $N\big(\mathbb{CP}^{2^n-1}, d_{\mathrm{tr}}, \varepsilon_2\big) \leqslant (C/\varepsilon_2)^{2\cdot 2^n-2}$, this implies that

$$\mathbf{Pr}\Big[\exists|\psi\rangle \in N\big(\mathbb{CP}^{2^n-1}, d_{\mathrm{tr}}, \varepsilon_2\big), |\langle\psi|(\rho-\hat{\rho})|\psi\rangle| > \varepsilon_1\Big] \leqslant (C/\varepsilon_2)^{2\cdot 2^n-2}\exp\left(-N\Omega\left(\varepsilon_1^2\right)\right).$$

From the definition of covering net $N\big(\mathbb{CP}^{2^n-1}, d_{\mathrm{tr}}, \varepsilon_2\big)$, we have

$$\mathbf{Pr}\Big[\|\rho-\hat{\rho}\|_\infty > \frac{\varepsilon_1}{1-\varepsilon_2}\Big] = \mathbf{Pr}\Big[\exists|\psi\rangle \in \mathbb{CP}^{2^n-1}, |\langle\psi|(\rho-\hat{\rho})|\psi\rangle| > \frac{\varepsilon_1}{1-\varepsilon_2}\Big] \leqslant (C/\varepsilon_2)^{2\cdot 2^n-2}\exp\left(-N\Omega\left(\varepsilon_1^2\right)\right).$$

Hence for any $0 < \eta \leqslant 1$, by choosing $\varepsilon_1, \varepsilon_2$ appropriately, with probability at least 0.99,

$$\|\rho - \hat{\rho}\|_\infty \leqslant \eta \qquad \text{if} \qquad N \geqslant \Omega(2^n/\eta^2)\,. \tag{3}$$

By taking $\eta = \epsilon/2^n$, we have

$$\|\rho - \hat{\rho}\|_1 \leqslant 2^n \|\rho - \hat{\rho}\|_\infty \leqslant 2^n \cdot \eta = \epsilon \qquad \text{if} \qquad N \geqslant \Omega(2^n/\eta^2) \, = \Omega(8^n/\epsilon^2). \tag{4}$$

So we concluded that $\tilde{O}(8^n/\epsilon^2)$ samples suffice to learn ρ to trace distance ϵ.

In the second part of this problem, we will show how to refine the sample complexity bound in the special case where the state has bounded rank. Henceforth, assume that ρ has rank r for some $1 \leqslant r \leqslant 2^n$. Let us define $\hat{\rho}_{\mathrm{LR}} \triangleq \mathrm{proj}(\hat{\rho})$, where proj is the projection to the space of rank-r density matrices given by retaining only the r largest eigenvalues. Concretely, if $\hat{\rho}$ has eigendecomposition $U \operatorname{diag}(\lambda_1, \ldots, \lambda_{2^n}) U^\dagger$ with $\lambda_1 \geqslant \cdots \geqslant \lambda_{2^n}$, then $\mathrm{proj}(\hat{\rho}) = U \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0) U^\dagger$.

1.D. (4 PTS.) Show that $\|\rho - \hat{\rho}_{\mathrm{LR}}\|_{\mathrm{op}} \leqslant 2\|\rho - \hat{\rho}\|_{\mathrm{op}}$.

Hint: Use Eckart-Young theorem: $\mathrm{proj}(\rho)$ is the best rank-r approximation to ρ in the operator norm.

1.E. (4 PTS.) If ρ has rank r, deduce from Eq. (4), for an appropriate choice of η, that $\|\rho - \hat{\rho}_{\mathrm{LR}}\|_{\mathrm{tr}} \leqslant \epsilon$ provided $N \geqslant \Omega(2^n r^2/\epsilon^2)$.

2 (30 PTS.) ESTIMATING MANY PAULI EXPECTATIONS IN PARALLEL

Motivation: Suppose we are given m Pauli operators $C_1, \ldots, C_m$ on n qubits, each of weight at most k. We want to estimate all expectations $\mu_i = \mathrm{tr}(C_i \rho)$ to additive accuracy ε, with failure probability at most δ, using as few copies of ρ as possible. Consider a randomized measurement scheme where one measures a random full-weight Pauli string from $\{X, Y, Z\}^{\otimes n}$ on each copy of ρ and reuses these outcomes to estimate every C_i supported on that string. A preliminary version of this measurement scheme is known as *quantum overlap tomography*. You will prove that

$$N = O\left(\frac{3^k}{\varepsilon^2} \log \frac{m}{\delta}\right)$$

samples suffice.

Setup: On each of N copies of ρ, draw $P^{(t)} \in \{X, Y, Z\}^{\otimes n}$ uniformly at random and measure it, obtaining outcomes $o_j^{(t)} \in \{\pm 1\}$. For $C_i = \bigotimes_{j=1}^n \sigma_{i,j}$ with support $\mathrm{supp}(C_i) = \{j : \sigma_{i,j} \neq I\}$, set

$$\begin{aligned} \mathbb{1}_i^{(t)} &= \mathbf{1}\left[\forall j \in \mathrm{supp}(C_i) : P_j^{(t)} = \sigma_{i,j}\right], \\ Y_i^{(t)} &= \prod_{j \in \mathrm{supp}(C_i)} o_j^{(t)}. \end{aligned}$$

If $\mathbb{1}_i^{(t)} = 1$, then $Y_i^{(t)}$ is a valid single-shot outcome for C_i. Let

$$S_i = \sum_{t=1}^N \mathbb{1}_i^{(t)}, \qquad \widehat{\mu}_i = \frac{1}{\max\{1, S_i\}} \sum_{t:\, \mathbb{1}_i^{(t)}=1} Y_i^{(t)}.$$

2.A. (3 PTS.) **Hit probability.** Show that for each i, $p_i = \mathbf{Pr}[\mathbb{1}_i^{(t)} = 1] = 3^{-\mathrm{wt}(C_i)}$. Deduce the uniform lower bound $p_i \geqslant 3^{-k}$ for all i.

2.B. (4 PTS.) **Enough hits.** Use the multiplicative Chernoff bound for $S_i \sim \mathrm{Bin}(N, p_i)$ to provide a value of N such that with probability at least $1 - \delta/3$, $S_i \geqslant (1-\varepsilon) N p_{\min}$, for all i, where $p_{\min} = \min_i p_i \geqslant 3^{-k}$.

2.C. (8 PTS.) **Accuracy given enough hits.** Condition on $S_i \geqslant T$. Show that Hoeffding's inequality implies

$$\mathbf{Pr}\left[|\widehat{\mu}_i - \mu_i| > \varepsilon \;\middle|\; S_i \geqslant T\right] \leqslant \frac{\delta}{3m}$$

provided $T \geqslant \frac{2}{\varepsilon^2} \log(6m/\delta)$.

2.D. (5 PTS.) **Putting it all together.** Combine the previous parts and a union bound over all i to conclude that if

$$N \geqslant \frac{2}{\varepsilon^2(1-\varepsilon)\,p_{\min}} \log\Big(\tfrac{6m}{\delta}\Big),$$

then with probability at least $1-\delta$, all estimates satisfy $|\widehat{\mu}_i - \mu_i| \leqslant \varepsilon$. Simplify using $p_{\min} \geqslant 3^{-k}$ to obtain the stated $O(3^k \varepsilon^{-2} \log(m/\delta))$ scaling.

2.E. (10 PTS.) **Extension: commutators.** Define $K_{jk} = \mathrm{tr}\big(i[P_j, P_k]\rho\big)$. Show that if each P_j has weight at most k, then any nonzero commutator has weight at most $2k-1$. Deduce the number of samples needed so that all nonzero entries of K are estimated to accuracy ε with probability at least $1-\delta$.

3 (45 (+10 OPTIONAL) PTS.) ADVANCED PAULI EXPECTATION ESTIMATION

Motivation: In earlier problems, we analyzed methods for efficiently estimating expectation values of *low-weight* Pauli operators. We now extend our analysis to the more general and challenging case where the Pauli observables are not guaranteed to be low-weight.

3.A. (7.5 PTS.) **Simultaneous Diagonalization.** Consider a set of m distinct n-qubit Pauli observables $S = \{C_1, \ldots, C_m\}$ that pairwise commute, i.e. $[C_i, C_j] = 0$ for all $i \neq j$. Show that there exists a unitary U such that for every i, $U^\dagger C_i U$ is a tensor product of I and Z operators.

Hint: Try to construct U in a sequence of steps that gradually transform the C_i's into products of I and Z operators, as if you were performing Gaussian elimination. Represent Pauli strings with symplectic vectors.

3.B. (7.5 PTS.) **Sample Complexity for Commuting Sets.** Using the result from (a), prove that $O(\log(m/\delta)/\varepsilon^2)$ measurements on individual copies of an unknown state ρ are sufficient to estimate all expectation values $\mathrm{tr}(C_i\rho)$ for $C_i \in S$ to additive error ε with total failure probability at most δ.

3.C. (7.5 PTS.) **Partitioned Sets.** Now consider an arbitrary set S of m Pauli observables that can be partitioned into M disjoint commuting subsets $S = S_1 \cup S_2 \cup \cdots \cup S_M$, where observables within each S_j mutually commute. Prove that $O(M\log(m/\delta)/\varepsilon^2)$ total measurements on individual copies of ρ are sufficient to estimate all expectation values in S to the desired accuracy.

3.D. (7.5 PTS.) **Fractional Coloring and Sample Complexity.** Consider an arbitrary set $S = \{C_1, \ldots, C_m\}$ of Pauli observables that may not commute with one another. To analyze this case, we introduce the concept of *fractional coloring* of a graph.

Definition: Let $G = (V, E)$ be a graph. An independent set in a graph is a set of vertices where no two vertices are connected by an edge. A **fractional coloring** of G with parameter χ is a probability distribution q over the independent sets $I \subseteq V$ of the graph such that every vertex $v \in V$ has probability at least $1/\chi$ of being included in an independent set randomly sampled from the distribution q:

$$\forall v \in V: \quad \mathop{\mathbf{Pr}}_{I\sim q}[v \in I] \geqslant 1/\chi.$$

The smallest possible value of χ for which such a distribution exists is called the **fractional chromatic number** of G, denoted $\chi_f(G)$.

Define the *anticommutation graph* $G(S)$ where each Pauli observable in S corresponds to a vertex, and an edge connects any two observables that **anticommute**. Prove that if $G(S)$ admits a fractional coloring with parameter χ, then a total of $O(\chi\log(m/\delta)/\varepsilon^2)$ measurements on individual copies of ρ suffice to estimate all expectation values in S to additive error ε with total failure probability at most δ.

3.E. (5 PTS.) **Two-Copy Commutativity.** We now explore an alternative strategy using two-copy measurements. Prove that for any set S of n-qubit Pauli observables, the corresponding set of operators $\{P \otimes P \mid P \in S\}$ acting on $2n$ qubits forms a collection of pairwise commuting observables.

3.F. (10 PTS.) **Magnitude Estimation.** Using the result from the previous part, prove that one can estimate the magnitude $|\mathrm{tr}(P\rho)|$ to additive error ε for all $P \in S$ using $O(\log|S|/\varepsilon^4)$ measurements on two-copy states $\rho \otimes \rho$. Explain why the sample complexity scales as ε^{-4} rather than the more familiar ε^{-2}.

3.G. (10, OPTIONAL PTS.) **Sign Recovery via Gentle Measurement.** The previous parts allow us to identify observables P where $|\mathrm{tr}(P\rho)| \leqslant \varepsilon$, for which an estimate of 0 provides sufficient accuracy. For the remaining subset $S' \subseteq S$ where $|\mathrm{tr}(P\rho)| > \varepsilon$ for all $P \in S'$, prove that one can determine the full expectation values $\mathrm{tr}(P\rho)$ to additive error ε by performing entangled measurements on additional $O(\log|S'|/\varepsilon^2)$ copies of ρ.

Hint: For each operator $P \in S'$, the sign of $\mathrm{tr}(P\rho)$ can be determined with small but non-negligible success probability. Use majority voting over multiple independent trials (done coherently) to amplify this success probability to a constant. Then invoke the gentle measurement lemma to bound the disturbance to the quantum state throughout this process.