Lecture 10: October 8, 2024

Scribe: Jin Ming Koh *State learning and tomography*

10.1 Recap & Introduction

The **random-access-code** (RAC) no-go theorem from last lecture says that any mechanism storing m bits of information in n qubits, conceptualized as a quantum map $\{0,1\}^m \to n$ qubits, that allows the retrieval of any bit with probability $\geq 1-\epsilon$ must satisfy $n \geq m\,[1-H(\epsilon)]$.

From a resource perspective, we can write $\mathcal{N} \geq C(\mathcal{N})$ for a channel $\mathcal{N}$ with capacity $C(\mathcal{N})$. This means that each use of the channel sends $C(\mathcal{N})$ bits of classical information (cbits). Intuitively, the $\geq$ sign here signifies "power" in the resource sense—the left-hand side can be used to achieve all the same things as the right-hand side can.

The **reverse Shannon theorem** says that $C(\mathcal{N}) + [\text{some}\,\mathsf{rbits}] \geq \mathcal{N}$, where rbits are shared pairs of random bits. The $\geq$ sign tells us we can simulate the channel $\mathcal{N}$ with the resources on the left-hand side. For example, $[1-H(\epsilon)]\,\mathsf{cbits} + [\text{some}\,\mathsf{rbits}] \geq \mathsf{BSC}_\epsilon$, where BSC_ϵ is the binary symmetric channel with error probability ϵ.

There is a hierarchy of resource inequalities. An rbit is the weakest. Sharing a maximally entangled pair (ebit), and transmission of a cbit, are stronger than an rbit. The transmission of a qubit is the strongest.

10.2 Quantum State Learning

Definition 10.2.1 (Quantum state learning task)**.** *We are given an unknown state ρ on n qubits, and an unknown distribution D of 2-outcome measurements $\{M, I-M\}$. We want to learn ρ, that is, to be able to predict the outcomes of measurements in D.*

Theorem 10.2.1 (Quantum state learning sample complexity)**.** *The quantum state learning task can be accomplished with $\mathcal{O}(1/\epsilon^2)$ samples, for an acceptable error tolerance ϵ.*

To be clear, in the quantum state learning task, we are given the data

$$(M_1, O_1 \sim \mathrm{tr}[M_1\rho]), \qquad (M_2, O_2 \sim \mathrm{tr}[M_2\rho]), \qquad \ldots \tag{10.1}$$

to learn from. We can think of ρ as a map $M \to [0,1]$, where M is a measurement operator.

10.2.1 Proof of sample complexity

To simplify our analysis, we shall assume that the measurement expectation values $\mathrm{tr}[M\rho] \approx 0$ or $\mathrm{tr}[M\rho] \approx 1$. More precisely, either $\mathrm{tr}[M\rho] \leq \epsilon$ or $\mathrm{tr}[M\rho] \geq 1-\epsilon$.

Definition 10.2.2 (VC dimension)**.** *The VC dimension of a set of states S is the largest set of measurements that is shattered by S*

Definition 10.2.3 (Shattering)**.** *We say that $M_1, \ldots, M_d$ is shattered if for all $O_1, \ldots, O_d \in \{0,1\}$ there exists ρ such that $\mathrm{tr}[\rho M_j] \approx O_j$ for all $j \in [d]$.*

Example 10.2.1. *We ask given measurement operators $0 \leq M_1, \dots, M_d \leq I$, for all $O_1, \dots, O_d \in \{0,1\}$, does there exist a ρ such that $\mathrm{tr}[\rho M_j] \approx O_j$ for all $j \in [d]$. The answer is in the positive for*

$$M_1 = \frac{I + Z_1}{2}, \qquad M_2 = \frac{I + Z_2}{2}, \qquad \dots. \tag{10.2}$$

That is, this example of $M_1, \dots, M_d$ is shattered.

Remark 10.2.1. *Observe that if $M_1, \dots, M_d$ is shattered then it defines a RAC. We consider the map $\{0,1\}^d \to \rho$, and we can concoct a ρ that returns the bitstring to be accessed upon measurement. This tells us*

$$d \leq \frac{n}{1 - H(\epsilon)}. \tag{10.3}$$

This implies that ρ can be learnt with $\mathcal{O}\big(n/\epsilon^{o(1)}\big)$ samples. To formally show this, one can use a theorem that says a concept class can be learnt with $\mathcal{O}$(VC dimension of class) samples.

10.3 State Tomography

Definition 10.3.1 (State tomography task)**.** *Let ρ be a $d \times d$ density matrix. Given $\rho^{\otimes n}$, we want to output a description of $\hat{\rho}$ such that*

$$\frac{1}{2}\|\rho - \hat{\rho}\|_1 \leq \epsilon, \tag{10.4}$$

for an acceptable error tolerance ϵ.

Remark 10.3.1. *The trace distance used in the quality criteria of state tomography means that we want $\hat{\rho}$ to be accurate across all measurements. This is stringent and is essentially a worst-case assurance. In comparison, quantum state learning concerns average-case error over the distribution of measurements D.*

Example 10.3.1. *In the $d = 2$ case,*

$$\rho = \frac{I + \sum_{j=1}^{3} \alpha_j \sigma_j}{2}, \tag{10.5}$$

where coefficients $\alpha_j = \mathrm{tr}[\rho \sigma_j]$. So we measure each σ_j $n/3$ times, to get estimates

$$\hat{\alpha}_j = \alpha_j + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \tag{10.6}$$

Then

$$\rho - \hat{\rho} = \sum_{j=1}^{3} (\alpha_j - \hat{\alpha}_j)\, \frac{\sigma_j}{2} \Longrightarrow \|\rho - \hat{\rho}\|_1 \sim \frac{1}{\sqrt{n}}. \tag{10.7}$$

Therefore we need number of samples $n \sim 1/\epsilon^2$ to get within error tolerance ϵ.

Theorem 10.3.1 (State tomography sample complexity)**.** *The state tomography task can be accomplished with $\mathcal{O}\big(d^2/\epsilon^2\big)$ samples.*

Remark 10.3.2. *In d dimensions, a density matrix has $d^2 - 1$ real degrees of freedom, which matches the d^2 in the sample complexity. Note that this is much worse than quantum state learning, which had sample complexity going as $\mathcal{O}(\log d)$.*

Theorem 10.3.2 (State tomography single-copy sample complexity)**.** *The state tomography task can be accomplished with $\mathcal{O}\big(d^3/\epsilon^2\big)$ samples using single-copy measurements only.*

10.3.1 Proof of sample complexity

Today we show $n \gtrsim d^2/\epsilon^2$. The idea is to construct $\rho_1, \dots, \rho_M$ satisfying the following properties:

1. *Well-separated.*
$$\frac{1}{2}\|\rho_x - \rho_y\|_1 \geq \frac{\epsilon}{10} \quad \forall \quad x \neq y. \tag{10.8}$$

2. *High entropy.*
$$S(\rho_x) \geq \log d - \mathcal{O}(\epsilon^2). \tag{10.9}$$

3. *Many states.*
$$M = \exp(cd^2), \tag{10.10}$$
where $c > 0$ is a constant.

Remark 10.3.3. *Can we really obtain $M = \exp(cd^2)$ many states? The volume of an ϵ-ball in d dimensions goes as ϵ^{d-1}, so maybe we can.*

The implication of $\rho_1, \dots, \rho_M$ is that, if we can perform state tomography on them with accuracy $\epsilon/20$ and failure probability $< \delta$, then we can distinguish the different $\rho_x^{\otimes n}$. We can take the description $\hat{\rho}$ to be the x labels, and this satisfies the quality criteria of the state tomography task.

We should imagine the pipeline
$$x \in [M] \longrightarrow \rho_x^{\otimes n} \longrightarrow \hat{\rho} \longrightarrow x. \tag{10.11}$$

We start by writing
$$I(X;\hat{X}) \geq (1-\delta)\log M - 1 \geq cd^2, \tag{10.12}$$
where the second inequality is due to Fano's inequality. Also, supposing that $\hat{\rho}^{\otimes n}$ lives on $Q^n = Q_1 Q_2 \dots Q_n$, we can write
$$I(X;\hat{X}) \leq I(X;Q^n) \leq nI(X;Q_1). \tag{10.13}$$

But
$$I(X;Q_1) = \underbrace{S\left(\frac{1}{M}\sum_x \rho_x\right)}_{\leq \log d} - \frac{1}{M}\sum_x S(\rho_x) \leq \mathcal{O}(\epsilon^2). \tag{10.14}$$

Putting the inequalities together, we conclude $n \gtrsim d^2/\epsilon^2$ as desired.

Remark 10.3.4. *What about the high entropy condition? Can that really be satisfied? Yes it can, by a short argument below.*

Lemma 10.3.1. *There exists $d \times d$ unitaries $U_1, \dots, U_M$ where $M = \exp(cd^2)$, and a fixed projector Π with rank $d/2$, such that for all $x \neq y$,*
$$\left\| U_x \frac{\Pi}{d/2} U_x^\dagger - U_y \frac{\Pi}{d/2} U_y^\dagger \right\|_1 \geq \frac{1}{10}. \tag{10.15}$$

Proof. Simply choose the $U_1, \dots, U_M$ randomly. □

Then, using this lemma, we choose
$$\rho_x = (1-\epsilon)\frac{I}{d} + \epsilon U_x \frac{\Pi}{d/2} U_x^\dagger. \tag{10.16}$$

The eigenvalues of ρ_x are then $(1 \pm \epsilon)/d$, and $S(\rho_x) = \log d - \mathcal{O}(\epsilon^2)$ as desired.