CHAPTER 15

Lower Bounds for Pauli Shadow Tomography

In this chapter, we will apply the learning tree formalism to obtain lower bounds for shadow tomography of *Pauli observables*. Concretely, the task here is to estimate $\mathrm{tr}(P\rho)$ for all n-qubit Pauli operators, given copies of unknown state ρ. Like purity testing, this turns out to be a problem for which two-copy measurements are exponentially more statistically efficient than single-copy measurements. In Section 1, we give a protocol that solves this task with two-copy measurements on only $O(n/\epsilon^4)$ copies of ρ. In Section 2.1, we give an exponential lower bound for this problem when one can only perform single-copy measurements. Intriguingly, the two-copy measurement protocol requires *adaptive* choice of measurements; in Section 3 we show that without adaptivity, there is an exponential lower bound even for two-copy measurements. Finally, in Section 4, we prove a qualitative strengthening of the lower bound from Section 2.1 demonstrating that even if one has a small amount of additional quantum memory, but not enough to perform two-copy measurements, then there is still an exponential lower bound for Pauli shadow tomography.

1. Upper bound using two-copy measurements

The protocol here involves two stages: first learning the *absolute values* $|\mathrm{tr}(P\rho)|$ and then learning the *signs* of the observable values. The first stage is based on a protocol due to [**HKP21, CCHL22**], and the second stage is based on a protocol due to [**CGY24, KGKB25**].

1.1. Learning the absolute values

The starting point for the first stage of the protocol is an idea from the chapter on learning stabilizer states, namely that even though the Pauli operators P do not commute, their tensor squares $P^{\otimes 2}$ do commute. Indeed, recall from Chapter 10 that these $\{P^{\otimes 2}\}$ can be simultaneously diagonalized, and their joint eigenbasis is the *Bell basis*.

By measuring in the Bell basis, we can thus obtain estimates for the *two-copy* observables $\mathrm{tr}(P^{\otimes 2}\rho^{\otimes 2}) = \mathrm{tr}(P\rho)^2$. This gives rise to the following guarantee:

Theorem 229 (Learning absolute values)**.** *There is a protocol which takes as input $O(n/\epsilon^4)$ copies of ρ, performs two-copy measurements in the Bell basis, and outputs estimates for all quantities $\{|\mathrm{tr}(P\rho)|\}$ up to additive error ϵ with high probability.*

PROOF. Using Bell basis measurements on $O(n/\epsilon^4)$ copies, we can simultaneously obtain ϵ^2-accurate estimates for all quantities $\{\mathrm{tr}(P\rho)^2\}$ with high probability. It remains to convert these to estimates for $\{|\mathrm{tr}(P\rho)|\}$. For this, observe that for any scalars $0 \leq x, y \leq 1$,

$$(\sqrt{x} - \sqrt{y})^2 = \frac{|x-y|^2}{(\sqrt{x}+\sqrt{y})^2} = |x-y| \cdot \frac{|x-y|}{x+y+\sqrt{2xy}} \leq |x-y|\,,$$

so if we simply output the square roots of our estimates for $\{\mathrm{tr}(P\rho)^2\}$, these will be ϵ-accurate. $\square$

1.2. Resolving the signs

Theorem 230. *Given ϵ-accurate estimates of $\{|\mathrm{tr}(P\rho)|\}$, and given the ability to perform* single-copy *measurements on $O(n/\epsilon^4)$ additional copies of ρ, there is a protocol which estimates $\{\mathrm{tr}(P\rho)\}$ to additive error $O(\epsilon)$ with high probability.*

We will consider Algorithm 10 below.

Algorithm 10: LEARNPAULISIGNS($\epsilon, \{f_P\}, \rho$)

Input: Accuracy $\epsilon > 0$, estimates $\{f_P\}$ of absolute values $\{|\mathrm{tr}(P\rho)|\}$, copies of ρ

Output: Estimates $\{\hat{E}_P\}$ of true values $\{\mathrm{tr}(P\rho)\}$

Find a state σ such that $|f_P - |\mathrm{tr}(P\sigma)|| \leq \epsilon$ for every P.

Perform Bell measurements on $O(n/\epsilon^4)$ copies of $\sigma \otimes \rho$ to estimate all $\mathrm{tr}(P^{\otimes 2}(\sigma \otimes \rho)) = \mathrm{tr}(P\sigma)\mathrm{tr}(P\rho)$ to error ϵ^2 with high probability.

Denote each estimate of $\mathrm{tr}(P\sigma)\mathrm{tr}(P\rho)$ by g_P.

If $f_P < 2\epsilon$, set $\hat{E}_P = 0$. Otherwise if $f_P > 2\epsilon$, set $\hat{E}_P = g_P/\mathrm{tr}(\sigma P)$.

return $\{\hat{E}_P\}_{P \in A}$

PROOF. In the first step, we can always find a σ such that $|f_P - |\mathrm{tr}(P\rho)|| \leq \epsilon$ for all P, as ρ itself satisfies the condition. Once we have an explicit description of this σ, we get exact access to $\mathrm{tr}(P\sigma)$.

In the second step, we obtain ϵ^2-accurate estimates g_P of $\mathrm{tr}(P \otimes P\ \sigma \otimes \rho)$ for every P. Observe that performing a Bell measurement on $\sigma \otimes \rho$ given explicit access to σ can be done using a single-copy measurement on ρ. If $f_P < 2\epsilon$, we set $\hat{E}_P = 0$ and this is trivially an 3ϵ-accurate estimate of $\mathrm{tr}(P\rho)$:

$$|\hat{E}_P - \mathrm{tr}(P\rho)| = |\mathrm{tr}(P\rho) - f_P + f_P| \leq |\mathrm{tr}(P\rho) - f_P| + |f_P| \leq \epsilon + 2\epsilon = 3\epsilon.$$

If $f_P > 2\epsilon$, we set $\hat{E}_P = g_P/\mathrm{tr}(P\sigma)$, in which case

$$|\hat{E}_P - \mathrm{tr}(P\rho)| = \frac{|g_P - \mathrm{tr}(P\sigma)\mathrm{tr}(P\rho)|}{|\mathrm{tr}(P\sigma)|} \leq \frac{\epsilon^2}{f_P - \epsilon} \leq \epsilon. \square$$

2. Lower bound using single-copy measurements

The protocol in the previous section uses two-copy measurements. It turns out that with only single-copy measurements, it is not possible to solve Pauli shadow tomography with such a small number of copies of ρ. The proof of this will use the learning tree framework from the previous chapter.

2.1. General lower bound based on second moment

Here we give a general recipe for proving lower bounds for shadow tomography using single-copy measurements; the lower bound on the number of copies needed will depend on a certain quantity that characterizes the extent to which the observables in question approximately commute.

Consider m traceless observables $O_1, \ldots, O_m$ with $\|O_i\|_\infty = 1$ which additionally satisfy

$$O_i = -O_{i+m/2}, \text{ and all eigenvalues of } O_i \text{ are } \pm 1, \quad \forall i = 1, \ldots, m/2\,. \tag{66}$$

We will characterize the hardness of estimating these observable values using single-copy measurements via the following quantity,

$$\delta(O_1, \ldots, O_m) = \sup_{|\phi\rangle} \frac{2}{m} \sum_{i=1}^{m/2} \langle\phi| O_i |\phi\rangle^2 \,, \tag{67}$$

where $\sup_{|\phi\rangle}$ is taken over all n-qubit pure states. We have the following general lower bound, which leverages a convexity argument first used in [**BCL20**]:

Theorem 231 (General shadow tomography lower bound)**.** *Let $O_1, \ldots, O_m$ be traceless observables satisfying Eq. (66). Then any learning protocol which only performs single-copy measurements on copies of unknown state ρ and outputs ϵ-accurate estimates for $\{\mathrm{tr}(O_i\rho)\}$ with high probability requires $\Omega\left(\frac{1}{\epsilon^2\delta(O_1,\ldots,O_m)}\right)$ copies.*

Proof. Following the framework in Chapter 14, we will show a lower bound for the easier task of distinguishing between whether the unknown state is maximally mixed or whether it is given by

$$\rho_i \coloneqq \frac{\mathrm{Id} + 3\epsilon O_i}{2^n}\,.$$

Note that if one had an algorithm for estimating $\mathrm{tr}(O_i\rho)$ for all i to error ϵ, then we could distinguish between these two ensembles.

Now consider any learning protocol using T copies of ρ, and consider its tree representation $\mathcal{T}$. For any leaf ℓ of $\mathcal{T}$, denote the sequence of edges on the path from root r to leaf ℓ by $\{e_{u_t,s_t}\}_{t=1}^T$, where $u_1 = r$ and $s_T = \ell$. If the POVM element corresponding to edge e_{u_t,s_t} is $\{2^n w_{s_t}^{u_t} |\psi_{s_t}^{u_t}\rangle \langle\psi_{s_t}^{u_t}|\}$, then the probability of reaching leaf ℓ for a given state ρ is given by

$$p^\rho(\ell) = \prod_{t=1}^T w_{s_t}^{u_t} 2^n \langle\psi_{s_t}^{u_t}|\rho|\psi_{s_t}^{u_t}\rangle\,.$$

We then have the following calculation:

$$\begin{aligned}
\frac{\left(\mathbb{E}_{i\sim[m]} p^{\rho_i}(\ell)\right)}{p^{\mathrm{Id}/2^n}(\ell)} &= \mathbb{E}_i \prod_{t=1}^{T} \left(\frac{w_s^u 2^n + 3\epsilon w_s^u 2^n \langle \psi_{s_t}^{u_t} | \, O_i \, | \psi_{s_t}^{u_t} \rangle}{w_s^u 2^n} \right) \\
&= \mathbb{E}_i \exp\left(\sum_{t=1}^{T} \log\left(1 + 3\epsilon \langle \psi_{s_t}^{u_t} | \, O_i \, | \psi_{s_t}^{u_t} \rangle\right)\right) \\
&\geq \exp\left(\sum_{t=1}^{T} \mathbb{E}_i \log\left(1 + 3\epsilon \langle \psi_{s_t}^{u_t} | \, O_i \, | \psi_{s_t}^{u_t} \rangle\right)\right) && (68) \\
&\geq \exp\left(\sum_{t=1}^{T} \frac{1}{m} \sum_{i=1}^{m/2} \log\left(1 - 9\epsilon^2 \langle \psi_{s_t}^{u_t} | \, O_i \, | \psi_{s_t}^{u_t} \rangle^2\right)\right) && (69) \\
&\geq \exp\left(- \sum_{t=1}^{T} \frac{18}{m} \sum_{i=1}^{m/2} \epsilon^2 \langle \psi_{s_t}^{u_t} | \, O_i \, | \psi_{s_t}^{u_t} \rangle^2 \right) && (70) \\
&\geq \exp\left(-9T\epsilon^2 \delta(O_1, \ldots, O_m)\right) && (71) \\
&\geq 1 - 9T\epsilon^2 \delta(O_1, \ldots, O_m)\,.
\end{aligned}$$

Eq. (68) uses Jensen's inequality, Eq. (69) uses the fact that the set of observables is closed under negation from Eq. (66), Eq. (70) uses the elementary inequality $\log(1-x) \geq -2x, \forall x \in [0, 3/4]$ which is satisfied given $\epsilon < 1/4$, and Eq. (71) uses the definition of δ in Eq. (67).

As long as $T \leq c/(\epsilon^2 \delta(O_1, \ldots, O_m))$ for sufficiently small constant $c > 0$, we obtain a sufficiently strong one-sided bound on the likelihood ratio for all leaves ℓ to deduce that one cannot distinguish between maximally mixed and $\{\rho_i\}$ with sufficiently good advantage. □

2.2. Instantiating the bound for Pauli observables

It remains to compute the $\delta(O_1, \ldots, O_m)$ quantity in the case of Pauli observables. We will take $m = 2(4^n - 1)$ and take $O_1, \ldots, O_m$ to be the set of nontrivial n-qubit Pauli observables together with their negations. This collection certainly satisfies Eq. (66). We now compute δ:

Lemma 232 (δ for Pauli observables)**.**

$$\delta\left(P_1, \ldots, P_{2(4^n-1)}\right) = \sup_{|\phi\rangle} \frac{1}{4^n - 1} \sum_{i=1}^{4^n-1} \langle \phi | \, P_i \, | \phi \rangle^2 = \frac{1}{2^n + 1}.$$

Proof. For any n-qubit pure state $|\phi\rangle$, we have

$$\begin{aligned}
\frac{1}{4^n - 1} \sum_{i=1}^{4^n-1} \langle \phi | \, P_i \, | \phi \rangle^2 &= \frac{1}{4^n - 1} \mathrm{tr}\left(\left(\sum_{i=1}^{4^n-1} P_i \otimes P_i\right) |\phi\rangle \langle \phi|^{\otimes 2}\right) \\
&= \frac{1}{4^n - 1} \mathrm{tr}\left(\left(2^n \mathrm{SWAP}_n - \mathrm{Id}^{\otimes 2n}\right) |\phi\rangle \langle \phi|^{\otimes 2}\right) \\
&= \frac{2^n - 1}{4^n - 1} = \frac{1}{2^n + 1}\,.
\end{aligned}$$

□