Frontiers of Quantum Complexity and Cryptography Spring 2022

Lecture 5 - Shadow Tomography

Lecturer: Henry Yuen *Scribes: Asa Kosto*

1 Overview

In this lecture we discussed shadow tomography. Shadow tomography allows us to get specific information about a quantum state using fewer copies of the state than full scale tomography. In contrast, with "standard" tomography you get a complete classical description of a quantum state, which is a very long vector/matrix whose dimension is exponential in the number of qubits. Performing "standard" tomography is too inefficient for anything except for very small quantum systems. However at the end of the day we're not interested in recovering a full vector/matrix description of a quantum state, but learning about certain properties of it (for example, measurement statistics on different subsets of qubits). This motivates shadow tomography.

Fun Fact: The record for quantum state tomography, as of 2018, was a 10 qubit system. [1, 2]

Main Goal: Can we use fewer copies of a state in order to obtain specific information about it that we care about?

2 Shadow Tomography

Warmup. Fix an observable A (i.e. a Hermitian matrix).

Suppose we are given copies of an unknown quantum state $|\psi\rangle$, and we are interested in the expectation value of that observable with respect to $|\psi\rangle$ within some error bounds. In particular, we would like to estimate $\langle\psi|A|\psi\rangle \pm \epsilon$. We can do this with k copies of $|\psi\rangle$ where $k = O(\|A\|^2/\epsilon^2)$, where $\|A\|$ is the spectral norm of A.

To see this, first observe that there is a simple quantum algorithm that, given one copy of $|\psi\rangle$, outputs a real number whose expectation is $\langle\psi|A|\psi\rangle$: let $A = \sum_j \lambda_j P_j$ denote the spectral decomposition of A, with P_j's forming orthogonal projections and λ_j's being the associated eigenvalues. Then the algorithm performs the projective measurement $P = \{P_j\}_j$, obtains outcome j, and then outputs the real number λ_j. By definition the expected value of the output is

$$\sum_j \langle\psi|P_j|\psi\rangle \cdot \lambda_j = \langle\psi| \sum_j \lambda_j P_j |\psi\rangle = \langle\psi|A|\psi\rangle \ .$$

We now invoke Chebyshev's inequality: by repeating this k times and taking an average, we get that the deviation from the mean $\langle\psi|A|\psi\rangle$ is more than ϵ with probability $\frac{\langle\psi|A^2|\psi\rangle}{\epsilon^2 k} \leq \frac{\|A\|^2}{\epsilon^2 k} = \frac{1}{100}$ if we set $k = 100\|A\|^2/\epsilon^2$.

Thus, if we know we are interested in only one observable A, then we can estimate $\langle\psi|A|\psi\rangle$ using a number of copies that is *independent* of the dimension of $|\psi\rangle$, which is great.

Estimating many observables What if we have a long list of observables $A_1 \dots A_T$ – how many samples of an unknown state $|\psi\rangle$ do we need then? Can we do better than repeating the procedure above T times (which would require $O(T \cdot \max_i \|A_i\|^2/\epsilon^2)$ samples). Here, let's think of T as being *exponentially large*, say 2^n. If we just repeated the above procedure T times, this would require an exponential number of copies of $|\psi\rangle$, in which case we might as well use "standard" tomography.

It turns out that shadow tomography allows us to estimate a large number of observables using a very small number of copies of the unknown state. Specifically, the setup of shadow tomography is this: given some number of copies $|\psi\rangle^{\otimes k}$ of an unknown state, we perform shadow tomography on it to extract a piece of data known as a *classical shadow* S. This shadow S is going to be a much more efficient representation of $|\psi\rangle$ that allows us to estimate $\langle\psi|A_j|\psi\rangle \pm \epsilon$ for the observables that we're interested in (but we will not be able to estimate *all* observables). The shadow tomography procedure may depend on the observables $\{A_j\}$.

$$|\psi\rangle^{\otimes k} \to \boxed{\text{Shadow Tomography}} \to \text{Classical Shadow } S \to \langle\psi|A_i|\psi\rangle$$

In order for shadow tomography to be interesting, we would like the following to be minimized:

1. The number of copies of $|\psi\rangle$ needed.
2. The complexity of producing the classical shadow S from copies of $|\psi\rangle$.
3. The complexity of estimating the observables $\langle\psi|A_j|\psi\rangle$ given the classical shadow S.

Ideally, we would like each of these to be as small as possible – polynomial in the number of qubits, even. It is interesting and non-trivial to get any one of these items smaller than polynomial in the dimension d of $|\psi\rangle$ (which is, remember, exponential in the number of qubits) and the number of observables T, which would represent a savings over "standard" tomography.

The following result of Aaronson was the first to give a shadow tomography procedure where the number of copies of $|\psi\rangle$ needed can be poly*logarithmic* in the dimension and in the number of observables one is interested in.

Theorem 1 (Aaronson's shadow tomography procedure [1])**.** *There exists a quantum algorithm that, given T projection matrices $P_1, \dots, P_T$ (which do not, in general, sum to the identity) and $|\psi\rangle^{\otimes T}$, outputs a classical shadow S from which one can estimate $\langle\psi|P_i|\psi\rangle \pm \epsilon$ as long as*

$$k = \widetilde{O}(\epsilon^{-4} \cdot \log^4 T \cdot \log d)$$

where $\widetilde{O}(\cdot)$ hides $\log\log T, \log\log d, \log\frac{1}{\epsilon}$ factors. The algorithm runs in time that is polynomial in T and d.

Features of Aaronson's shadow tomography algorithm: We can estimate the probabilities of exponentially many measurements using only a polynomial number of copies of $|\psi\rangle$.

Drawbacks: The time complexity of the algorithm is still exponential, and the shadow is still quite large, around size poly(2^n).

2.1 A (more) time-efficient shadow tomography method

Recently, Huang, Kueng and Preskill presented another shadow tomography procedure that is more time-efficient than Aaronson's procedure.

Theorem 2 (Huang, Kueng, Preskill's shadow tomography procedure [3])**.** *Fix $\epsilon, \delta > 0$ and integer $B > 0$. There exists a quantum algorithm that, given $k = O(\log(\frac{1}{\delta})\frac{B}{\epsilon^2})$ copies of an unknown state $|\psi\rangle$, generates a classical shadow S that has the following property: for all observables M such that $\|M\|_{shadow} \leq B$, we have that with probability at least $1-\delta$ over S that the quantity $\langle\psi|M|\psi\rangle$ can be estimated to within $\pm\epsilon$. The algorithm to generate the shadow S runs in polynomial time; however the procedure to estimate $\langle\psi|M|\psi\rangle$ may not necessarily be.*

There are several things to note about Huang, Kueng and Preskill's shadow tomography procedure:

1. The shadow tomography procedure does *not* depend on the observables being estimated (unlike with Aaronson's procedure). However, which observables can be estimated accurately depend on the shadow S that is generated, which is probabilistic.

2. This theorem can be easily extended to handle a large number of observables. Let $M_1, \ldots, M_T$ denote observables. By the above theorem, if we generate a shadow S given $|\psi\rangle^{\otimes T}$, then for all $j \in [T]$ we have

$$\Pr_S[\text{ can estimate } \langle\psi|M_j|\psi\rangle \text{ to within } \pm\epsilon] \geq 1 - \delta \ .$$

   Then by the union bound we have

$$\Pr_S[\text{ can estimate } \langle\psi|M_j|\psi\rangle \text{ for all } j \text{ to within } \pm\epsilon] \geq 1 - \delta T \ .$$

   Thus provided that $\delta T \ll 1$, we can accurately estimate *all* of the observables $M_1, \ldots, M_T$ with high probability over S provided that $k = O(\log T \cdot B/\epsilon^2)$.

3. The quantity $\|M\|_{shadow}$ is known as the *shadow norm* of an observable M, which can be upper bounded by $\mathrm{Tr}(M^2)$. Intuitively it is a measure of complexity of the observable. An example of observables that have small shadow norm are *pure state observables*, i.e., $M = |\phi\rangle\langle\phi|$ for some $|\phi\rangle$. Such observables are useful if one wants to check whether an unknown state $|\psi\rangle$ is one of a collection of different pure states $|\phi_1\rangle, \ldots, |\phi_T\rangle$.

What do these shadows look like? The Huang, Kueng, Preskill procedure is quite simple. Assume that $|\psi\rangle$ is an n-qubit state. For the j'th copy of $|\psi\rangle$, the algorithm will sample an independent, uniformly random *Clifford unitary* C_j on n qubits, apply it to $|\psi\rangle$, and then measure it in the standard basis to obtain an n-qubit string $b_j \in \{0,1\}^n$. The shadow S simply consists of the sequence of pairs

$$(C_1, b_1), (C_2, b_2), \ldots, (C_k, b_k) \ .$$

This is efficient because a random Clifford circuits are easy to sample and to describe (they have a canonical form that consists of at most $O(n^2)$ gates).

We won't discuss it here, but it is not that much harder to describe how to take such a shadow to estimate the value of an observable M. If you're interested, you should take a look at [3] to see how it's done!

References

[1] S. Aaronson, *Shadow Tomography of Quantum States*, Proceedings of STOC'2018, ECCC TR17-164, arXiv:1711.01053, 2018

[2] C. Song, K. Xu, W. Liu, C. Yang, S.-B. Zheng, H. Deng, Q. Xie, K. Huang, Q. Guo, L. Zhang, P. Zhang, D. Xu, D. Zheng, X. Zhu, H. Wang, Y.-A. Chen, C.-Y. Lu, S. Han, and J.-W. Pan, *10-qubit entanglement and parallel logic operations with a superconducting circuit*, Phys. Rev. Lett., 119(180511), 2017. arXiv:1703.10302

[3] H-Y. Huang, R. Keung, J. Preskill *Predicting many properties of a quantum system from very few measurements*, Nature Physics, pages 1050–1057, 2020