Frontiers of Quantum Complexity and Cryptography Spring 2022

Lecture 3 - A simple tomography algorithm

Lecturer: Henry Yuen *Scribes: Shuhan Zhang*

1 Overview

In the last lecture we learned about state tomography. In this lecture we talked about a simple tomography procedure and covered some proven and proposed bounds for tomography. Finally, we introduced the Haar distribution and discussed its construction and complexity.

2 Review of Tomography

Given k copies of a n qubit mixed state $\rho \in \mathbb{C}^{d \times d}$ with $d = 2^n$, the task of tomography is to output a classical description "σ" such that $D(\rho, \sigma) \leqslant \epsilon$.

In the last lecture, we have seen that in order to perform tomography, we need at least $\Omega(d/\log d) = \Omega(2^n/n)$ copies of ρ. We obtained this lower bound by:

1. Calculating the maximum number of quantum states that can be packed into a d-dimensional space such that the distance between any two states is at least 2ϵ; this number R gives rise to a "quantum codebook" and $\log R$ determines the number of classical bits that can be conveyed using this codebook.
2. Using *Holevo's theorem*, which states that at most n classical bits can be reliably encoded into n qubits, to obtain the minimum number of copies of ρ needed.

3 A Simple Tomography Algorithm

We introduced a simple tomography algorithm using $\tilde{O}(d^6)$ copies of ρ ($\tilde{O}(\cdot)$ denotes the upper bound omitting log factors). This algorithm leverages the fact that every Hermitian matrix can be expressed as a linear combination of Pauli matrices, which is proven in problem set 1. Using this fact, ρ can be written as:

$$\rho = \sum_{W \in \{I,X,Y,Z\}^n} \alpha_W W_1 \otimes \cdots \otimes W_n \ .$$

The idea of this algorithm is to estimate α_W for all W's and use these α_W's to obtain an approximation of ρ.

3.1 Algorithm of Pauli Tomography

This simple algorithm of Pauli tomography takes the following steps:

1. Divide the input state $\rho^{\otimes k}$ into 4^n groups of $\rho^{\otimes t}$, where $k = 4^n \cdot t$.

2. For each $W \in \{I, X, Y, Z\}^n$, use t copies to estimate α_W. Recall from Problem set 1 that

$$\alpha_W = 2^{-n}\mathrm{Tr}(\rho W) \ .$$

   Call this estimate $\tilde{\alpha}_W$.

3. Compute the closest density matrix σ to $\tilde{\sigma}$, where $\tilde{\sigma}$ is obtained from the $\tilde{\alpha}_W$'s by

$$\tilde{\sigma} = \sum_W \tilde{\alpha}_W W_1 \otimes \cdots \otimes W_n \ .$$

   Note that $\tilde{\sigma}$ is Hermitian, but it is not necessarily true that $\tilde{\sigma}$ is positive semidefinite and $\mathrm{Tr}(\tilde{\sigma}) = 1$, so $\tilde{\sigma}$ might not be a valid quantum state. Thus, the classical description "σ" is obtained by computing the closest density matrix to $\tilde{\sigma}$.

3.2 Analysis

Suppose that a large enough t has been chosen so that $|\tilde{\alpha}_W - \alpha_W| \leqslant \eta$ for all of the 4^n different W's. We want to find out how good the tomography result is by computing how far $\tilde{\sigma}$ is from ρ. In particular, we want to calculate the trace norm $\| \tilde{\sigma} - \rho \|_1$, which is a measure of the maximum probability of distinguishing between $\tilde{\sigma}$ and ρ.

$$\begin{aligned} \| \tilde{\sigma} - \rho \|_1 &= \| \sum_W \tilde{\alpha}_W W - \sum_W \alpha_W W \|_1 = \| \sum_W (\tilde{\alpha}_W - \alpha_W) W \|_1 \leqslant \sum_W \| (\tilde{\alpha}_W - \alpha_W) W \|_1 \\ &\leqslant \sum_W |\tilde{\alpha}_W - \alpha_W| \, \| W \|_1 \leqslant \eta \sum_W \| W \|_1 = 4^n \cdot 2^n \cdot \eta = \epsilon \ . \end{aligned} \tag{1}$$

Here, the first inequality follows from the triangle inequality, the second inequality follows from the fact that $(\tilde{\alpha}_W - \alpha_W)$ is a number, and the last equality follows from the fact that there are 4^n different W's in the sum and $\| W \|_1 = \sum_j |\lambda_j| = \sum_j |\pm 1| = 2^n$, because W has 2^n eigenvalues $\lambda_j = \pm 1$.

However, the output of this simple tomography algorithm is σ rather than $\tilde{\sigma}$. Therefore, we want to compute how far σ is from ρ.

$$\| \sigma - \rho \|_1 = \| \tilde{\sigma} - \tilde{\sigma} + \tilde{\sigma} - \rho \|_1 \leqslant \| \sigma - \tilde{\sigma} \|_1 + \| \tilde{\sigma} - \rho \|_1 \leqslant \epsilon + \epsilon = 2\epsilon \ . \tag{2}$$

Here, the first inequality follows from triangle inequality. For the second inequality, we know from (1) that $\| \tilde{\sigma} - \rho \|_1 \leqslant \epsilon$, and since σ is the closest quantum state to $\tilde{\sigma}$, $\| \sigma - \tilde{\sigma} \|_1 \leqslant \| \rho - \tilde{\sigma} \|_1 \leqslant \epsilon$, so $\| \sigma - \tilde{\sigma} \|_1 + \| \tilde{\sigma} - \rho \|_1 \leqslant \epsilon + \epsilon = 2\epsilon$.

Now we have obtained a target distance of 2ϵ from (2), we want to work backwards to see how large t needs to be in order to guarantee $|\tilde{\alpha}_W - \alpha_W| \leqslant \eta$ for all of the 4^n different W's. We start by looking at the probability that $|\tilde{\alpha}_W - \alpha_W| \leqslant \eta$ for a fixed W. Recall that

$$\alpha_W = 2^{-n}\mathrm{Tr}(\rho W) \ . \tag{3}$$

In Problem set 1, we have calculated that in order to output an estimate β such that

$$|\beta - \mathrm{Tr}(\rho W)| \leqslant \tau \ . \tag{4}$$

with probability at least $1-\gamma$, we need to input $O(\frac{1}{\tau^2}\ln\frac{1}{\gamma})$ copies of ρ. This means that we need

$$t = O(\frac{1}{\tau^2}\ln\frac{1}{\gamma}) \ . \tag{5}$$

Now we want to calculate how small τ and γ can be. From (3) and (4), we can see that in order for

$$|\tilde{\alpha}_W - \alpha_W| = |\tilde{\alpha}_W - 2^{-n}\mathrm{Tr}(\rho W)| \leqslant \eta \ ,$$

we need

$$\tau \leqslant 2^n \cdot \eta = \frac{\epsilon}{4^n} \ . \tag{6}$$

On the other hand, we can figure out how small γ can be in terms of the total probability. Suppose that we want to estimate $|\tilde{\alpha}_W - \alpha_W| \leqslant \eta$ for all W's with probability $1-\delta$.

$$\Pr[\exists W \text{ such that } |\tilde{\alpha}_W - \alpha_W| > \eta] \leqslant \sum_W \Pr[|\tilde{\alpha}_W - \alpha_W| > \eta] = 4^n \cdot \gamma = \delta \ . \tag{7}$$

Here, the inequality follows from the union bound. The first equality follows from the fact that there are 4^n W's in the sum. From (7), we can obtain an expression of γ in terms of δ:

$$\gamma = \frac{\delta}{4^n} \ . \tag{8}$$

Therefore, from (5), (6), and (8), we get

$$t = O\left(\frac{1}{\tau^2}\ln\frac{1}{\gamma}\right) = O\left(\left(\frac{4^n}{\epsilon}\right)^2 \ln\frac{4^n}{\delta}\right) = O\left((4^n)^2\frac{1}{\epsilon}n\ln\frac{1}{\delta}\right) \ ,$$

and

$$k = 4^n \cdot t = O\left((4^n)^3\frac{1}{\epsilon}n\ln\frac{1}{\delta}\right) = \tilde{O}(d^6) \ .$$

3.3 Complexity

We have just shown that the *sample complexity* (i.e. number of copies of the input state ρ needed) of the aforementioned tomography algorithm is $\tilde{O}(d^6)$. What about its *time complexity* (i.e. how much time it takes to process those copies of ρ)? The estimation procedure for α_W clearly takes $\mathrm{poly}(t,n)$ time, because it simply requires measuring each of the n qubits of t copies of ρ in a different Pauli basis, and averaging the outcomes (as you determined in Problem Set 1). There are 4^n different α_W's to estimate, so the estimation process takes $\mathrm{poly}(4^n,t,n)$ time.

The only potentially tricky part is the part about computing the closest density matrix σ to the estimate $\tilde{\sigma}$. Actually, here we will argue how to find a "close-enough" density matrix σ (not necessarily the closest). Essentially the basic idea is to truncate the negative eigenvalues of $\tilde{\sigma}$ and rescale. This can be done using basic linear algebra: since $\tilde{\sigma}$ is a Hermitian matrix, it can be diagonalized as

$$\tilde{\sigma} = \sum_j \tilde{\lambda}_j |b_j\rangle\langle b_j|$$

for some basis $\{|b_j\rangle\}_j$ and some eigenvalues (not necessarily all positive) $\{\lambda_j\}_j$. Since there exists a density matrix ρ that is ϵ-close to $\tilde{\sigma}$, this means that (a) the trace of $\tilde{\sigma}$ is close to 1 and (b) the total mass of the negative eigenvalues of $\tilde{\sigma}$ cannot be large.

Lemma 1. $\sum_{j\in N} |\tilde{\lambda}_j| \leq 2\epsilon$.

Proof. Let $N = \{j : \tilde{\lambda}_j < 0\}$ denote the indices of the negative eigenvalues of $\tilde{\sigma}$. Now, define P to be the projector onto the span of $\{|b_j\rangle\}_{j\in N}$, in other words, the eigenvectors whose eigenvalues are negative.

A useful variational characterization of the trace norm $\|X\|_1$ of a matrix X is the following:

$$\|X\|_1 = \max_{M:\|M\|_\infty \leq 1} \mathrm{Tr}(MX)$$

Here the maximization is over all matrices M whose maximum singular value (also known as the *operator norm*) is at most 1. The projection P is such a matrix, so therefore

$$\epsilon \geq \frac{1}{2}\|\rho - \tilde{\sigma}\|_1 = \frac{1}{2} \max_{M:\|M\|_\infty \leq 1} \mathrm{Tr}(M(\rho - \tilde{\sigma})) \geq \frac{1}{2}\mathrm{Tr}(P(\rho - \tilde{\sigma})) = \frac{1}{2}\mathrm{Tr}(P\rho) - \frac{1}{2}\mathrm{Tr}(P\tilde{\sigma}) \ .$$

Now $\mathrm{Tr}(P\rho)$ is a nonnegative number (because trace of the product of two positive semidefinite matrices is nonnegative). We also have

$$\mathrm{Tr}(P\tilde{\sigma}) = \sum_{j\in N} \tilde{\lambda}_j < 0 \ .$$

Putting the previous two lines together we conclude that $\sum_{j\in N} |\tilde{\lambda}_j| \leq 2\epsilon$ as desired. □

Lemma 2. $1 - 2\epsilon \leq \sum_{j\notin N} \tilde{\lambda_j} \leq 1 + 2\epsilon$.

Proof. We again use the variational characterization of the trace norm.

$$\epsilon \geq \frac{1}{2}\|\rho - \tilde{\sigma}\|_1 = \frac{1}{2} \max_{M:\|M\|_\infty \leq 1} \mathrm{Tr}(M(\rho - \tilde{\sigma})) \geq \frac{1}{2}\mathrm{Tr}(\rho - \tilde{\sigma}) \ .$$

Since ρ is a density matrix, $\mathrm{Tr}(\rho) = 1$. We then have $\mathrm{Tr}(\tilde{\sigma}) = \sum_{j\notin N} \tilde{\lambda}_j - \sum_{j\in N} |\tilde{\lambda}_j|$. Rearranging, we have

$$\sum_{j\notin N} \tilde{\lambda}_j \geq 1 - 2\epsilon + \sum_{j\in N} |\tilde{\lambda}_j| \geq 1 - 2\epsilon$$

where we used our bound on the total mass of negative eigenvalues of $\tilde{\sigma}$. The proof of the upper bound proceeds similarly. □

Assuming these bounds, we can define our nearby density matrix σ by truncating the negative eigenvalues of $\tilde{\sigma}$ and rescaling:

$$\sigma := \frac{1}{\beta} \sum_{j\notin N} \tilde{\lambda}_j |b_j\rangle\langle b_j|$$

where $\beta = \sum_{j \notin N} |\tilde{\lambda}_j|$. This is clearly a density matrix: it is positive semidefinite (all its eigenvalues are nonnegative) and it has trace 1. We just need to calculate how far σ is from $\tilde{\sigma}$:

$$\begin{aligned}
\|\sigma - \tilde{\sigma}\|_1 &\leq \left\| \frac{1}{\beta} \sum_{j \notin N} \tilde{\lambda}_j |b_j\rangle\langle b_j| - \sum_{j \notin N} \tilde{\lambda_j} |b_j\rangle\langle b_j| \right\|_1 + \left\| \sum_{j \notin N} \tilde{\lambda_j} |b_j\rangle\langle b_j| - \tilde{\sigma} \right\|_1 \\
&= \sum_{j \notin N} \tilde{\lambda_j} \left(1 - \frac{1}{\beta}\right) + \sum_{j \in N} |\tilde{\lambda}_j| \\
&\leq 4\epsilon \ .
\end{aligned}$$

Performing this truncation and rescaling can be done in time that is polynomial in the dimension of these matrices (this is something you can do in, say, Matlab).

Thus the overall time complexity of the tomography procedure is poly(d).

4 Tomography Bounds

We have already seen in class some upper and lower bounds on the number of copies, k, of a quantum state ρ which are required for tomography. Specifically we have proven the lower bound that we need at least

$$k \geq \Omega\left(\frac{d}{\log d}\right)$$

copies. And we have from our simple algorithm that we can achieve tomography for error parameters ϵ, δ with no more than

$$O\left(d^6 \log d \frac{1}{\epsilon^2} \log \frac{1}{\delta}\right)$$

copies.

But much tighter bounds are known, and indeed we can say confidently that in general $k \approx O\left(\frac{d^2}{\epsilon^2}\right)$ copies are both required and sufficient, based on a lower bound from O'Donnell and Wright [1] and an upper bound from Haah et al. [2]. However, the algorithm which achieves tomography using this bound is quite complicated, and the best known runtime is exponential in the dimension (so *doubly-exponential* in the number of qubits!). This leads us to a natural question:

Open Question: Can we achieve a poly(d) runtime with this optimal number of copies?

The answer to this is not currently known.

How do we account for the difference of a factor of d between our lower bound and the lower bound provided by [1]? Recall that our proof of the lower bound relied only on pure states, giving us d degrees of freedom. However, there is nothing to stop us from encoding messages in mixed states: $m \Longleftrightarrow \rho_m \in \mathbb{C}^{d \times d}$. Allowing this flexibility increases our message space by a factor of d, and this idea is used in the proof of the optimal lower bound.

Open Question: What is the true bound if we restrict ourselves to only pure states?

We have a proven upper bound of $O(\frac{d}{\epsilon^2})$, but this is not tight. Whether a tighter bound exists is unknown (and a potential project topic!).

Practical Note: These bounds make a significant difference to physicists who are actually performing tomography in the lab. Tomography of two qubits typically takes hours, and three or more qubits may take days. So even small improvements to these bounds can have huge impacts.

References

[1] R. O'Donnell, J. Wright, *Efficient quantum tomography*, STOC 2016: Proceedings of the forty-eighth annual ACM symposium on Theory of Computing:899-912, 2016.

[2] J. Haah, A. Harrow, Z. Ji, X. Wu, N. Yu, *Sample-Optimal Tomography of Quantum States*, IEEE Transactions on Information Theory, 63(9):5628 - 5641, 2017.