Supplementary Information

Zachary Kokot

August 12, 2024

Contents

1 Measurement in Different Bases with only Computational Basis Measurements

1.1 Motivation

On existing quantum devices it is often the case that measurements can only be performed in the computational basis. However, it is possible to perform measurements in different bases by applying a unitary transformation to the state before measurement. In this section, we will show how to perform measurements in different bases using only computational basis measurements.

1.2 Method

Given a quantum state $|\psi\rangle$ of a two level system (qubit), we can mathematically represent a measurement which results in one of the two basis states $|v_1\rangle$ or $|v_2\rangle$ as

$$|v_1\rangle \langle v_1|\psi\rangle = v_1 |v_1\rangle ,$$
$$|v_2\rangle \langle v_2|\psi\rangle = v_2 |v_2\rangle .$$

These two measurement outcomes, v_1 and v_2, are the eigenvalues of my measurement basis. Say that instead I want to measure in a different basis with eigenstates $|u_1\rangle$ and $|u_2\rangle$. I need to then perform some unitary operation on my measurement operator so

$$|v_1\rangle \langle v_1| \xrightarrow{\mathbf{U}} |u_1\rangle \langle u_1| ,$$
$$|v_2\rangle \langle v_2| \xrightarrow{\mathbf{U}} |u_2\rangle \langle u_2| .$$

This can be done by simply applying the unitary operator, $\mathbf{U}$, to the ket and bra which make up the measurement operator and thus the measurement results are given by

$$\mathbf{U} |v_1\rangle \langle v_1| \mathbf{U}^\dagger |\psi\rangle = |u_1\rangle \langle u_1|\psi\rangle = u_1 |u_1\rangle ,$$
$$\mathbf{U} |v_2\rangle \langle v_2| \mathbf{U}^\dagger |\psi\rangle = |u_2\rangle \langle u_2|\psi\rangle = u_2 |u_2\rangle .$$

This is not so helpful for us as our measurement operator is fixed but using the associative property we can place brackets to "change" the order operators are applied. This allows us to write the measurement results as

$$\mathbf{U}(|v_1\rangle \langle v_1| (\mathbf{U}^\dagger |\psi\rangle)) = u_1 |u_1\rangle ,$$
$$\mathbf{U}(|v_2\rangle \langle v_2| (\mathbf{U}^\dagger |\psi\rangle)) = u_2 |u_2\rangle .$$

This is useful as we can see that the measurement results are given by the application of the unitary operator to the state before measurement. This allows us to perform

measurements in different bases using only computational basis measurements as long as we can fine a unitary operator which performs the desired transformation. Additionally, notice the left most operator $\mathbf{U}$ only serves to transform the state after measurement to the corresponding basis state but as measurement is usually the final step in a quantum circuit it can usually be ignored.

1.3 Example

Consider the following example. We wish to measure in the X basis but only have access to computational (Z) basis measurements. Following the method above We must find a unitary operator which transforms the Z basis states to the X basis states. Conveniently, We can write the X basis states as a linear combination of the Z basis states as follows

$$|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle,$$
$$|-\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle.$$

Thus the unitary operator must act as follows

$$\mathbf{U}|0\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle,$$
$$\mathbf{U}|1\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle.$$

The states of single qubit systems can be described by a two element vector and the unitary operator can be described by a 2×2 matrix. Thus, the unitary operator acting on the computational basis states can be written in the computational basis as

$$\mathbf{U}\begin{pmatrix}1\\0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},$$
$$\mathbf{U}\begin{pmatrix}0\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

If we express the unitary operator as a matrix, we have that

$$\mathbf{U} = \begin{pmatrix}a & b\\c & d\end{pmatrix}.$$

Then, we can solve for the elements of the matrix by solving the following system of equations

$$\begin{pmatrix}a & b\\c & d\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},$$
$$\begin{pmatrix}a & b\\c & d\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Solving the above system of equations we find that

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Thus, the unitary operator which transforms the Z basis states to the X basis states is simply the Hadamard gate. Thus, to measure in the X basis we must apply the adjoint of the Hadamard gate to the state before measurement. Conveniently the Hadamard gate is its own adjoint so just the Hadamard gate must be applied to the state before measurement.

2 Proof any Density Matrix can be Expressed as a Linear Combination of Pauli Matrices

2.1 Motivation

2.2 Theorem

Let $\hat{\rho}$ be an $n \times n$ density matrix. Then, $\hat{\rho}$ can be expressed as a linear combination of tensor products of Pauli matrices. That is, there exists a set of coefficients $\{a_{i_1,i_2,\ldots,i_n}\}$ such that:

$$\hat{\rho} = \sum_{i_1,i_2,\ldots,i_n=0}^{3} a_{i_1,i_2,\ldots,i_n} \hat{\sigma}_{i_1,i_2,\ldots,i_n}, \tag{1}$$

where $\hat{\sigma}_{i_1,i_2,\ldots,i_n} = \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \cdots \otimes \hat{\sigma}_{i_n}$ and $\hat{\sigma}_i$ corresponds to the i-th Pauli matrix (where σ_0 is the 2×2 identity matrix.). Additionally, the coefficients $a_{i_1,i_2,\ldots,i_n}$ are the expectation values of the corresponding Pauli matrices. That is, $a_{i_1,i_2,\ldots,i_n} = \mathrm{tr}(\hat{\sigma}_{i_1,i_2,\ldots,i_n}\hat{\rho})$.

2.3 Proof

We will begin by proving the theorem for a single qubit. Then, we will extend the proof to n qubits. For a single qubit, the density matrix can be expressed as:

$$\hat{\rho} = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \tag{2}$$

Where to satisfy the Hermiticity condition, $\rho_{01} = \rho_{10}^*$, and ρ_{00} and ρ_{11} are real. Additionally, to satisfy the trace condition, $\mathrm{tr}(\hat{\rho}) = 1$, we have that $\rho_{00} + \rho_{11} = 1$.
We know that density matrices are Hermitian and positive semi-definite. Thus, the following theorem is useful.

Theorem 1

3 Proof of the Sample Complexity for Complete State Tomography

3.1 Motivation

If we wish to compare the sample complexity of shadow tomography to full state tomography, we must first determine the sample complexity of full state tomography. In this section, we will prove that the sample complexity of full state tomography scales exponentially with the number of qubits.

3.2 Theorem

Fix a system of n qubits and maximum error ϵ. Then, the number of measurements required to perform full state tomography scales with the number of qubits as follows:

$$N = \mathcal{O}\left(\frac{4^n}{\epsilon^2}\right).$$

3.3 Proof

In quantum mechanics the outcomes of measurements are probabilistic. Start with the classic example of

4 Proof of the Sample Complexity for the Random Local Pauli Measurement Primitive

4.1 Motivation

Part of what makes shadow tomography appealing is that it requires a smaller number of measurements to predict an observable within the same error when compared to full state tomography. In this section, we will prove that the random local Pauli measurement primitive requires a number of measurements that scales logarithmically with the number of target observables and exponentially in the locality of the observables. This is a significant improvement over full tomography, which requires a number of measurements that scales exponentially with the number of qubits.

4.2 Theorem

Fix a collection of M observables, maximum error ϵ, and failure probability δ. Set

$$K = 2\log\left(\frac{2M}{\delta}\right) \quad \text{and} \quad N \leq \frac{34}{\epsilon^2} \max_{1\leq i\leq M} 4^{L_i} ||O_i||_\infty^2,$$

Where L_i is the locality of the i-th observable, $||O||_\infty$ denotes the matrix norm. Then, a collection of NK independent classical shadows allow for the prediction of M observables via median of means prediction such that

$$|\hat{o}_i(N,K) - \mathrm{tr}(O_i\rho)| \leq \epsilon \quad \text{for all} \quad 1 \leq i \leq M$$

with probability at least $1-\delta$.

4.3 Proof

To estimate the expectation value of M observables by median of means estimation with maximum error ϵ and success probability $1-\delta$ we turn to the results of [1]. In which they state the following theorem:

Theorem 2

Let X be a random variable with variance σ^2. Then, K independent sample means of size $N = 34\frac{\sigma^2}{\epsilon^2}$ suffice to construct a median of means estimator $\hat{\mu}(N,K)$ that obeys

$$\mathbb{P}[|\hat{\mu}(N,K) - \mathbb{E}[X]| \geq \epsilon] \leq 2e^{-K/2}. \tag{3}$$

They also give the following lemma:

Lemma 1

Fix an observable O and set $\hat{o} = tr(O\hat{\rho})$, where $\hat{\rho}$ is a classical shadow. Then

$$Var[\hat{o}] = \mathbb{E}\left[(\hat{o} - \mathbb{E}[\hat{o}^2])\right] \leq \left|\left|O - \frac{tr(O)}{2^n}\mathbb{I}\right|\right|^2_{shadow}. \tag{4}$$

From the above theorem and lemma, we can conclude that to estimate a single observable the size of each estimator is given by

$$N \leq \frac{34}{\epsilon^2}\left|\left|O - \frac{\mathrm{tr}(O)}{2^n}\mathbb{I}\right|\right|^2_{\text{shadow}}. \tag{5}$$

To estimate M observables, we need to take the maximum of the above expression over all observables. This will ensure that for all observables the maximum error is less than ϵ. Thus, the size of each estimator to estimate M observables is given by

$$N \leq \frac{34}{\epsilon^2}\max_{1\leq i\leq M}\left|\left|O_i - \frac{\mathrm{tr}(O_i)}{2^n}\mathbb{I}\right|\right|^2_{\text{shadow}}. \tag{6}$$

Finally, for the arbitrary Pauli basis measurement primitive we have that the shadow norm has the following property

$$\left|\left|O - \frac{\mathrm{tr}(O)}{2^n}\mathbb{I}\right|\right|^2_{\text{shadow}} \leq 4^{L(O)}||O||^2_\infty, \tag{7}$$

Where $L(O)$ is the locality of the observable and $||O||_\infty$ denotes the matrix norm. This implies that the size of each estimator to estimate M observables is given by

$$N \leq \frac{34}{\epsilon^2} \max_{1\leq i\leq M} 4^{L_i}||O_i||^2_\infty. \tag{8}$$

From the above theorem, we have that for a single random variable the probability the error bound is violated is given by

$$\mathbb{P}[|\hat{\mu}(N,K) - \mathbb{E}[X]| \geq \epsilon] \leq 2e^{-K/2}. \tag{9}$$

We would like to ensure that the probability that the error bound is for any of the M observables is less than δ. To determine an expression for K which satisfies this requirement we can use Booles inequality. Booles inequality states that for any countable set of events $A_1, A_2, \ldots, A_n$ we have

$$\mathbb{P}\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \mathbb{P}[A_i]. \tag{10}$$

In English this means that the probability any event in the set occurs is no greater than the sum of the probabilities of the individual events. Assume the estimation of each observable has a probability of violating the error bound of δ_0. Then, by Booles inequality, the probability that the error bound is violated for any of the M observables is less than

$$\sum_{i=1}^{M} \mathbb{P}[A_i] = \sum_{i=1}^{M} \delta_0 = M\delta_0 = \delta.$$

Thus, we have that

$$2e^{-K/2} \leq \frac{\delta}{M}.$$

Rearranging the above expression we find that

$$K \geq 2\log\left(\frac{2M}{\delta}\right).$$

Hence, a collection of NK independent classical shadows allow for the prediction of M observables via median of means prediction such that

$$|\hat{o}_i(N,K) - \mathrm{tr}(O_i\rho)| \leq \epsilon \quad \text{for all} \quad 1 \leq i \leq M$$

with probability at least $1-\delta$. □

5 Proof of the Computational Complexity for the Estimation of Pauli Observables from shadows generated by the Random Local Pauli Measurement Primitive

5.1 Motivation

The computational complexity of the estimation of Pauli observables from shadows generated by the random local Pauli measurement primitive is an important consideration when determining the feasibility of the primitive. In this section, we will prove that the computational complexity of the estimation of Pauli observables from shadows generated by the random local Pauli measurement primitive scales linearly with the number of qubits, the number of target observables, the size of the shadow.

5.2 Theorem

Fix a collection of M Pauli observables, a collection of NK independent classical shadows, and a system size (number of qubits) n. Then, the computational complexity of the estimation of M observables from the shadows scales with

$$\mathcal{O}\left(MNKn\right).$$

5.3 Proof

The estimation of Pauli observables using a collection of NK independent classical shadows generated by the random local Pauli measurement primitive can be broken down into the following steps for a single observable:

1. Split the NK independent classical shadows into K groups of size N.
2. Iterate over each group of shadows and estimate the expectation value of the Pauli observable.
   (a) Iterate over each shadow in the group.
   (b) For each shadow in the group, obtain the single shot expectation value for the Pauli observable.
   (c) Compute the mean of the single shot expectation values.
3. Compute the median of the means of the estimates of the Pauli observable.

It is clear to see how the terms N and K arise as the estimation protocol simply loops over all NK terms in the collection of shadows.
The term n arises from the way in which each single shot expectation value is obtained. See Section 6 for more details.
Finally, the term M arises from the fact that the estimation protocol must be repeated for each of the M Pauli observables. This is because the estimation of each Pauli observable not dependent of the others in general.

6 Computing the Expectation for a Single Pauli Observable

6.1 Motivation

In this section, we will outline how to compute the expectation value of a single Pauli observable using the random local Pauli measurement primitive. This implementation avoids the need for reconstructing a single shot density matrix representation of the state and the computation of a possibly very large matrix product. This offers the potential for a computational speedup by avoiding more expensive operations. Additionally, this implementation does not require the exponentially growing density matrix to be stored in memory.

6.2 Method

The expectation value of a single Pauli observable can be computed from a collection of shadows using the following equation. Let O be a Pauli observable and $\hat{\rho}$ be a classical shadow. Then, the single shot expectation value of the Pauli observable is given by:

$$\hat{o} = \prod_{j=1}^{n} \left[\text{tr}\{3U_j P_j U_j^\dagger \left|\hat{b}_j\right\rangle \left\langle\hat{b}_j\right|\} - \text{tr}\{P_j\} \right]. \tag{11}$$

Where P_j is the j-th Pauli operator, U_j is the j-th unitary operator, and $|\hat{b}_j\rangle$ is the j-th bit of the bitstring of the shadow. Where the above expression evaluates to $\pm 3^{\text{Locality}(O)}$ if for each j, U_j transforms P_j to Z (the Pauli-Z operator). Otherwise the expression evaluates to 0. Then the expectation value of the Pauli observable is given by the mean of the single shot expectation values.

6.3 Proof

Starting with the general expression for the expectation value of an observable, we have:

$$\hat{o} = \text{tr}(O\hat{\rho})$$

Where O is the observable and $\hat{\rho}$ is the classical shadow. For a Pauli observable, O, we have that $O = \bigotimes_{j=1}^{n} P_j$ where $P_j \in \{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$. Then, the expectation value of the Pauli observable is given by:

$$\hat{o} = \text{tr}\left\{\left(\bigotimes_{j=1}^{n} P_j\right)\hat{\rho}\right\}.$$

From [1], we have that:

Theorem 3

In the random local Pauli measurement primitive, we have the single shot density matrix representation of the state is given by:

$$\hat{\rho} = \bigotimes_{j=1}^{n}\left(3U_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right| U_j - \mathbb{I}\right). \tag{12}$$

Thus, the expectation value of the Pauli observable is given by:

$$\hat{o} = \text{tr}\left\{\left(\bigotimes_{j=1}^{n} P_j\right)\left(\bigotimes_{j=1}^{n}\left(3U_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right| U_j - \mathbb{I}\right)\right)\right\},$$

$$\hat{o} = \text{tr}\left\{\bigotimes_{j=1}^{n} P_j\left(3U_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right| U_j - \mathbb{I}\right)\right\},$$

$$\hat{o} = \text{tr}\left\{\bigotimes_{j=1}^{n} 3P_jU_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right| U_j - P_j\right\},$$

$$\hat{o} = \prod_{j=1}^{n}\left[\text{tr}\{3P_jU_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right| U_j - P_j\}\right],$$

$$\hat{o} = \prod_{j=1}^{n}\left[\text{tr}\{3P_jU_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right| U_j\} - \text{tr}\{P_j\}\right],$$

$$\hat{o} = \prod_{j=1}^{n}\left[\text{tr}\{3U_jP_jU_j^\dagger \left|\hat{b}_j\right\rangle\left\langle\hat{b}_j\right|\} - \text{tr}\{P_j\}\right].$$

References

[1] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". en. In: *Nature Physics* 16.10 (2020), pp. 1050–1057. DOI: 10.1038/s41567-020-0932-7. URL: https://www.nature.com/articles/s41567-020-0932-7.