Classical Shadows Tomography

A very brief overview

Antonio Ruiz-Molero

International Iberian Nanotechnology Laboratory – INL

June 15, 2022

Overview

1. Motivation of the talk
2. Context
3. Main idea
4. Procedure
5. Main results
6. Overview of the proofs using global Clifford unitaries
7. Application to quantum process tomography

References

[1] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting Many Properties of a Quantum System from Very Few Measurements". In: *Nature Physics* 16.10 (Oct. 2020), pp. 1050–1057. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/s41567-020-0932-7. arXiv: 2002.08953. URL: http://arxiv.org/abs/2002.08953 (visited on 02/04/2022).

[2] Ryan Levy, Di Luo, and Bryan K. Clark. *Classical Shadows for Quantum Process Tomography on Near-term Quantum Computers*. arXiv:2110.02965. type: article. arXiv, Oct. 6, 2021. arXiv: 2110.02965[cond-mat,physics:physics,physics:quant-ph]. URL: http://arxiv.org/abs/2110.02965 (visited on 05/13/2022).

Motivation

- Very pedagogical paper.
- Ideas about algorithms can be easy to produce but not so easy to justify and prove their performance (at least for me).
- Learn the way of thinking and path to follow when facing these problems.
- You might learn something.
- You don't want to miss this cool and trendy technique.

Context

- Quantum state tomography. Learn about an unknown state $\in \mathcal{H}$ of dimension 2^n where n is the number of qubits.
- Measuring only allows accessing to a part of the system each time, needing several copies of the state.
- Curse of dimensionality. Number of parameters scales exponentially with size.

Shadow tomography

We are interested in certain properties of the state:

$$o_i = \langle\psi|O_i|\psi\rangle \, ,$$

for a certain set of observables $\{O_1, ..., O_M\}$.

Related work[1]

General quantum state tomography

- Reconstructs full density matrix.
- Information-theoretic bounds assures that, at least, $d \operatorname{rank} \rho$ copies are needed.
- Saturating this bounds imply circuits with entanglement that acts on each qubit simultaneously.
- More tractable techniques allow to $\operatorname{rank}(\rho)^2 d$.
- For full generality, $d = 2^n$ copies are needed.
- Do not forget the classical power needed to store and process data.

[1] More information can be found in Section 3 in Supplementary information of [1]

Related work

Matrix Product State tomography

- In certain many-body systems, a MPS with low bond dimension can approximate the state.
- Number of samples scales polynomially.
- In the general case, similar to general QST.

Neural Network tomography

- A deep neural network is trained by feeding in quantum measurement outcomes. The internal network configuration shapes to provide a classical description of the system.
- Still not well understood the class of systems that can be efficiently represented.

Direct fidelity estimation

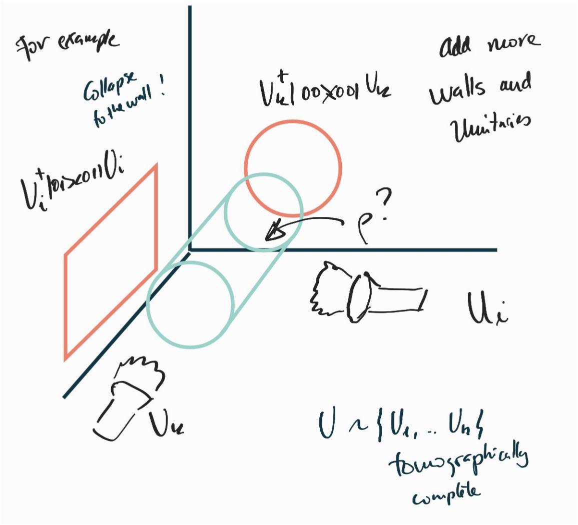
- Tailor-made to learn $\langle\psi|\rho|\psi\rangle$ up to accuracy ϵ
- Can vary from $\frac{1}{\epsilon^2}$ to $\frac{2^n}{\epsilon^4}$

Related work

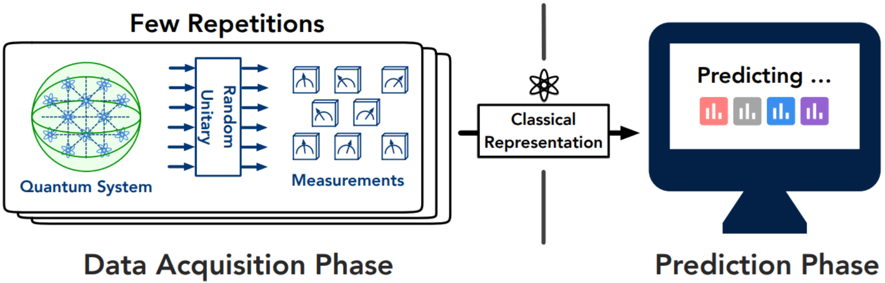
Shadow tomography

- Aims at simultaneously estimating the outcome probabilities associated with M 2-outcome measurements: $p_i(\rho) = \mathrm{tr}(E_i \rho)$ where the maximum operator norm of E_i is 1.
- The best result need a number of copies of $\tilde{\mathcal{O}}(\log(M)^2 log(d)^2/\epsilon^2)$

Main idea


for example
Collapse to the wall !
$V_i^\dagger |0\rangle\langle 0| V_i$
$V_k^\dagger |0\rangle\langle 0| V_k$
Add more walls and Unitaries
ρ?
U_i
U_k
$U \sim \{U_1, .. U_n\}$
tomographically complete

Procedure


Few Repetitions
Quantum System
Random Unitary
Measurements
Classical Representation
Predicting ...
Data Acquisition Phase
Prediction Phase

Data acquisition

Snapshot. Repeat N times

- Select a random unitary U from the set $\mathcal{U}$ that is assumed to be tomographically complete, i.e., for different states $\rho, \sigma, \exists U \in \mathcal{U}$ $s.t.$ $\langle b|U\rho U^\dagger|b\rangle \neq \langle b|U\sigma U^\dagger|b\rangle$, with computational basis $\{|b\rangle : b \in \{0,1\}^n\}$
- Apply this U to the unknown state $\rho \to U\rho U^\dagger$
- Measure in the computational basis.
- Invert U on the outcome of the measurement: $U^\dagger|b\rangle\langle b|U$
- Store it in a computer.

Classical shadow of ρ

- ρ is collapsed to $|\hat{b}\rangle\langle\hat{b}|$ with prob $Pr[\hat{b} = b] = \langle b|U\rho U^\dagger|b\rangle$, $b \in \{0,1\}^n$
- Average over all the unitaries of our set:

$$\mathbb{E}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U) = \mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|U\rho U^\dagger|b\rangle \; U^\dagger|b\rangle\langle b|U = \mathcal{M}(\rho)$$

- Think about this like a quantum channel $\mathcal{M}$. We are interested in the inverse:

$$\hat{\rho} = \mathcal{M}^{-1}(U^\dagger|b\rangle\langle b|U)$$

- Tomographic completeness ensures that this inverse is unique.
- We call $\hat{\rho}$ the classical shadow of ρ. Not semi-definite, in general.
- $\mathbb{E}(\hat{\rho}) = \rho$ by design.

Classical representation and prediction

Using the stored data, we can efficiently apply $\mathcal{M}^{-1}$ to each snapshot, constructing a classical snapshot of ρ. The whole set is called classical shadow of ρ:

$$S(\rho, N) = \left\{\hat{\rho}_1 = \mathcal{M}^{-1}(U_1^\dagger|\hat{b}_1\rangle\langle\hat{b}_1|U_1), ..., \hat{\rho}_N = \mathcal{M}^{-1}(U_N^\dagger|\hat{b}_N\rangle\langle\hat{b}_N|U_N)\right\}.$$

Median of means

We want to predict $O_1, ..., O_M$ observables.

- Import the classical shadow $S(\rho, N)$
- Split the shadow into K equally-sized sets and compute the mean of each set: $\hat{\rho}_{(k)} = \frac{1}{\lfloor N/K \rfloor}\sum_{i=(k-1)\lfloor N/K\rfloor+1}^{k\lfloor N/K\rfloor} \hat{\rho}_i$
- Predict each expected value o_i as: $o_i = \text{median}\left\{\text{tr}\left(O_i\hat{\rho}_{(1)}\right), \ldots, \text{tr}\left(O_i\hat{\rho}_{(K)}\right)\right\}$ for $i = 1, ..., M$

On the predictions of linear functions

Theorem 1

Fix a measurement primitive $\mathcal{U}$*, a collection of* $O_1, \dots, O_M$ *of* $2^n \times 2^n$ *Hermitian matrices and accuracy parameters* $\epsilon, \delta \in [0,1]$. *Set*

$$K = 2\log(2M/\delta) \text{ and } N = \frac{34}{\epsilon^2} \max_{1 \le i \le M} \left\| O_i - \frac{tr(O_i)}{2^n} \mathbb{I} \right\|_{shadow}^2 .$$

Then, a collection of NK independent classical shadows allow for accurately predicting all features via median of means:

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

with probability at least $1 - \delta$.

On the predictions of linear functions

Shadow norm

$$\|O\|_{\text{shadow}} = \max_{\sigma:\text{ state}} \left(\mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|U\sigma U^\dagger|b\rangle \langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle^2 \right)^{1/2}$$

The norm and, therefore, the convergence and number of samples needed depend on $\mathcal{M}$, which depends on the set of unitaries $\mathcal{U}$.

The dimension of the system does not appear!

Generalization of Theorem 1 without constants

$$N_{tot} = \left(\frac{\log(2M)}{\epsilon^2} \max_{1\leq i\leq M} \left\| O_i - \frac{tr(O_i)}{2^n}\mathbb{I} \right\|^2_{\text{shadow}} \right)$$

Performance depending on $\mathcal{U}$

Random Clifford measurements

With $\mathcal{U} = \mathrm{Cl}(2^n)$:

$$\hat{\rho} = (2^n + 1)U^\dagger |\hat{b}\rangle\langle\hat{b}| U - \mathbb{I} \text{ and } \left\| O - \frac{tr(O)}{2^n} \right\|_{\text{shadow}}^2 \leq 3tr(O^2).$$

Very powerful, difficult to implement in practice.

Random Pauli measurements

Only tensor product of single-qubit Clifford gates. $U = U_1 \otimes \ldots \otimes U_n \sim \mathcal{U} = \mathrm{Cl}(2)^{\otimes n}$

$$\hat{\rho} = \bigotimes_{j=1}^{n} (3U^\dagger |\hat{b}\rangle\langle\hat{b}| U - \mathbb{I}) \text{ and } \left\| O - \frac{tr(O)}{2^n} \right\|_{\text{shadow}}^2 \leq 4^{\text{locality}(O)} \|O\|_\infty^2$$

Matching information-theoretic lower bounds

- Random Clifford measurements:

Theorem 2

Any procedure based on a fixed set of single-copy measurements that can predict, with additive error ϵ, M arbitrary linear functions $tr(O_i \rho)$, requires at least $\Omega(\log(M) \max_i tr(O_i^2)/\epsilon^2)$ copies of the state ρ.

- Random Pauli measurements:

Theorem 3

Any procedure based on a fixed set of single-copy measurements that can predict, with additive error ϵ, M arbitrary k-linear functions $tr(O_i \rho)$, requires at least $\Omega(\log(M) 3^k/\epsilon^2)$ copies of the state ρ.

Predicting linear functions with classical shadows

Lemma 1

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

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

Handwritten proof

Predicting linear functions with classical shadows

- Once we know the variance, which depends on the specific set of unitaries, we want to see the convergence of the method.
- We recall classical concentration arguments: Chernoff, Hoeffding inequalities (they all derive from Markov inequality).
- Applied to this scheme, samples scale as $N = Var[\hat{o}_i]/(\delta\epsilon^2)$. Bad scaling with ϵ, our additive error.
- Use median of means.

Proof of theorem 1

Handwritten proof

Proof of scaling for random global Clifford unitaries

Handwritten proof

- Unitary t-designs.
- Application to our case of interest.

Quantum process tomography

- Need of characterizing quantum devices.
- Focus on the dynamics of the physical system.

Approaches

- Maximum Likelihood estimation
- Linear inversion method
- Bayesian methods
- Compressed sensing methods

Quantum channel

A quantum channel is a linear map from the space of operators on a Hilbert space to the space of operators of another Hilbert space that is $\mathcal{CP}$ and preserves trace.

Quantum channel

A quantum channel is a linear map

$$\Phi : L(\mathcal{X}) \rightarrow L(\mathcal{Y}),$$

i.e., $\Phi \in T(\mathcal{X}, \mathcal{Y})$, with $\mathcal{X}, \mathcal{Y}$ Hilbert spaces, satisfying that Φ is completely positive and trace preserving.

If $\mathcal{X} = \mathcal{Y} = \mathcal{H}$, it can map the state $\rho \rightarrow \Phi(\rho)$
Example: U unitary, $\Phi(\rho) = U\rho U^{\dagger}$ is called a unitary channel.
How can we learn a quantum channel using techniques like Classical shadows tomography?

Choi–Jamiołkowski isomorphism

Choi–Jamiołkowski isomorphism

Given $\mathcal{E} : L(\mathcal{X}) \to L(\mathcal{Y})$ a linear map. Fixed an orthonormal basis for $\mathcal{X}$ and define $P_+ = \sum_{i,j=1}^{d} |ii\rangle\langle jj|$. Then the mapping

$$\Lambda_{\mathcal{E}} = (\mathcal{I} \otimes \mathcal{E})(P_+)$$

defines a isomorphism between the vector spaces of linear maps on a d-dimensional system and linear operators on a $\dim(\mathcal{X}) \times \dim(\mathcal{Y})$-dimensional Hilbert space $L(\mathcal{Y} \otimes \mathcal{X})$.
The inverse is given by:

$$\mathcal{E}(\rho) = \mathrm{tr}_1 \left[\left(\rho^T \otimes \mathcal{I} \right) (\Lambda_{\mathcal{E}}) \right] \ \rho \in L(\mathcal{X})$$

If $\mathcal{E}$ is $\mathcal{CP}$, the obtained operator is positive on $L(\mathcal{Y} \otimes \mathcal{X})$.
Quantum process tomography can be thought as quantum state tomography!!

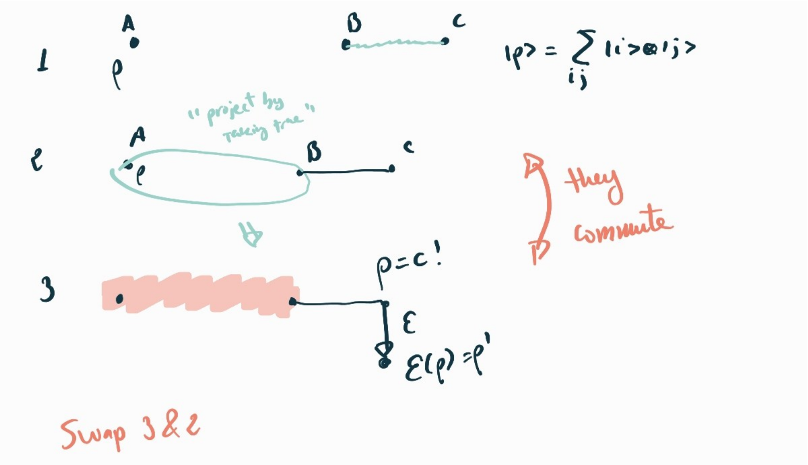
Choi–Jamiołkowski isomorphism

Check of the inverse function

$$\begin{aligned}
\mathrm{tr}_1\left[\left(\rho^T \otimes \mathcal{I}\right)(\Lambda_{\mathcal{E}})\right] &= \mathrm{tr}_1\left[\left(\rho^T \otimes \mathcal{I}\right)(\mathcal{I} \otimes \mathcal{E}) \sum_{i,j=1}^{d} |jj\rangle\langle kk|\right] \\
&= \sum_{i,j=1}^{d} \mathrm{tr}_1\left[\left(\rho^T \otimes \mathcal{I}\right)(|j\rangle\langle k| \otimes \mathcal{E}(|j\rangle\langle k|))\right] = \sum_{i,j=1}^{d} \mathrm{tr}\left(\rho^T |j\rangle\langle k|\right) \otimes \mathcal{E}(|j\rangle\langle k|) \\
&= \sum_{i,j=1}^{d} \langle k|\rho^T j\rangle\, \mathcal{E}(|j\rangle\langle k|) = \mathcal{E} \sum_{i,j=1}^{d} (|j\rangle\langle k|\, \langle j|\rho|k\rangle) = \mathcal{E}(\rho)
\end{aligned}$$

$\mathrm{tr}(\Lambda_{\mathcal{E}}) = d \to \rho_{\Lambda_{\mathcal{E}}} = \Lambda_{\mathcal{E}}/d = \Lambda_{\mathcal{E}}/2^n$

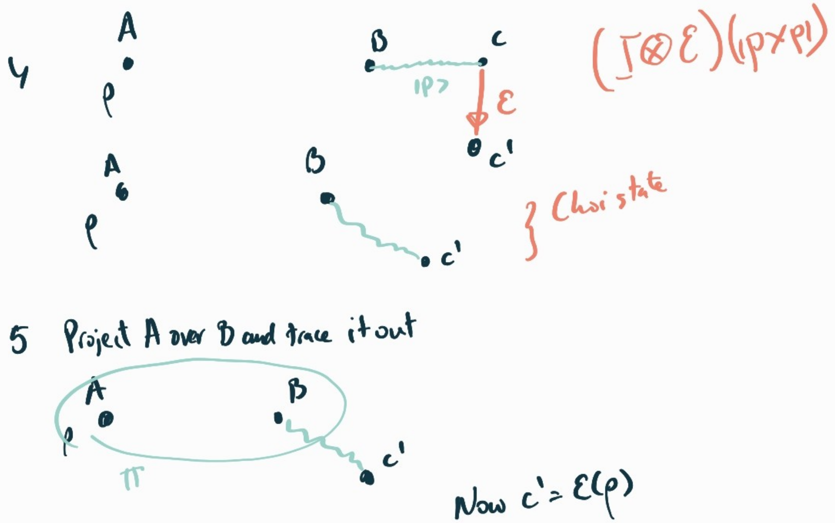
Intuition



2

[2]Source: https://bit.ly/3xflPkY

Intuition


4
A
ρ
B
C
$|\varphi\rangle$
$\mathcal{E}$
C'
$(I \otimes \mathcal{E})(|\varphi\rangle\langle\varphi|)$
A
ρ
B
C'
} Choi state
5 Project A over B and trace it out
A
B
ρ
π
C'
Now $C' = \mathcal{E}(\rho)$

Theoretical analysis of classical shadows for quantum process tomography

Theorem 4

For a n-qubit quantum process ρ_Λ and $\epsilon, \delta \in (0,1)$, given a set of density matrix pairs $\{(\rho_1^{in}, \sigma_1), ..., (\rho_M^{in}, \sigma_M)\}$, the number of measurements N that suffices to predict $\mathrm{tr}((\rho_i^{in} \otimes \sigma_i)\rho_\Lambda)$ for any i up to error ϵ with probability $1-\delta$ is of order:

$$\frac{\log(2M/\delta)}{\epsilon^2} \max_{1 \leq i \leq M} \left\| O_i - \frac{tr(O_i)}{2^n} \mathbb{I} \right\|_{shadow}^2$$

with $O_i = (\rho_i^{in} \otimes \sigma_i)$ and the shadow norm introduced before.

Theoretical analysis of classical shadows for quantum process tomography

Global random Clifford measurements

The number of copies needed scales as $\frac{\log(2M/\delta)}{\epsilon^2} \max_{1\leq i\leq M} \operatorname{tr}(O_i^2)$. If ρ_i^{in}, σ_i are all pure states, $\operatorname{tr}(O_i^2) = 1$

1-qubit random Clifford measurements

The number of copies needed scales as $\frac{\log(2M/\delta)}{\epsilon^2} \max_{1\leq i\leq M} 4^{k_i} \|O_i\|_\infty^2$. If ρ_i^{in}, σ_i are all pure states, $\|O_i\|_\infty^2 = 1$. k_i defines the locality of the operator O_i

Estimating these expectation values is equivalent to compute the overlap between σ_i and ρ_i: $\operatorname{tr}(\mathcal{E}(\rho_i^{in})\sigma_i = 2^n \operatorname{tr}((\rho_i^{in} \otimes \sigma_i)\rho_\Lambda)$. The scaling is the same except for the dependence on the number of qubits due to normalization.

Full quantum process tomography using global Clifford unitaries

Not only we want to estimate the overlap between two states, one of it evolved under the channel, but to characterize the Choi matrix completely. Which will be the scaling?

Theorem 5

For a n-qubit quantum process ρ_Λ *and* $\epsilon, \delta \in (0,1)$*, the number of random global Clifford measurements N that suffices to simultaneously predict any reduced k-qubit process Choi matrix* $\rho_{\Lambda^{(k)}}$ *such that*

- *Frobenius norm error up to* ϵ *with probability* $1-\delta$ *is of order* $\frac{4^{k+n}}{\epsilon^2}\log(2(8n)^{2k}/\delta)$.

For full quantum process tomography, k=n. It scales exponentially with the number of qubits. Sad news, but it was expected.

Classical Shadows Tomography

A very brief overview

Antonio Ruiz-Molero

International Iberian Nanotechnology Laboratory – INL

June 15, 2022