Haar measure applications

Myeongjin Shin

QISCA Summer School 2025
Quantum Learning and Complexity Theory – **Lecture 5**

Aug 9, 2025

Introduction

Lecture 4 explores diverse examples and applications where the **Haar measure** plays a fundamental role in quantum information. We derive well-known formulas that reduce to computing moments over the Haar measure, including the **twirling of quantum channels** and the **average gate fidelity**. These formulas lay the foundation for various applications, such as Randomized Benchmarking.

Introduction

Furthermore, we provide detailed insights into two notable examples showcasing the applications of the theory of unitary design. We examine **Barren Plateaus** in Variational Quantum Algorithms, shedding light on the optimization landscapes encountered in such algorithms. Additionally, we delve into **Classical Shadow tomography**, where the theory of unitary design aids in designing efficient measurement strategies for reconstructing properties of unknown quantum states.

Recap: Haar measure

Definition 1 (Haar measure)

The Haar measure on the unitary group $\mathrm{U}(d)$ is the unique probability measure μ_H that is both left and right invariant over the group $\mathrm{U}(d)$, i.e., for all integrable functions f and for all $V \in \mathrm{U}(d)$, we have:

$$\int_{\mathrm{U}(d)} f(U)\, d\mu_H(U) = \int_{\mathrm{U}(d)} f(UV)\, d\mu_H(U) = \int_{\mathrm{U}(d)} f(VU)\, d\mu_H(U). \tag{1}$$

The Haar measure is a probability measure, satisfying:

- $\int_S 1\, d\mu_H(U) \geq 0$
- $\int_{\mathrm{U}(d)} 1\, d\mu_H(U) = 1$
- $\underset{U \sim \mu_H}{\mathbb{E}} [f(U)] := \int_{\mathrm{U}(d)} f(U)\, d\mu_H(U)$

Recap: Computing moments

Theorem 2 (Computing moments)

Let $O \in \mathcal{L}\left((\mathbb{C}^d)^{\otimes k}\right)$. The moment operator can then be expressed as a linear combination of permutation operators:

$$\mathop{\mathbb{E}}_{U \sim \mu_H}\left[U^{\otimes k} O U^{\dagger \otimes k}\right] = \sum_{\pi \in S_k} c_\pi(O) V_d(\pi), \tag{2}$$

where the coefficients $c_\pi(O)$ can be determined by solving the following linear system of $k!$ equations:

$$\mathbf{Tr}\left(V_d^\dagger(\sigma) O\right) = \sum_{\pi \in S_k} c_\pi(O) \mathbf{Tr}\left(V_d^\dagger(\sigma) V_d(\pi)\right) \quad \textit{for all } \sigma \in S_k. \tag{3}$$

This system always has at least one solution.

Example 3 (First and second moment)

Given $O \in \mathcal{L}\left(\mathbb{C}^d\right)$, we have:

$k=1$

$\mathbb{I}$

$$\mathop{\mathbb{E}}_{U\sim\mu_H}\left[UOU^\dagger\right] = \frac{\mathbf{Tr}(O)}{d} I. \tag{4}$$

$\mathrm{Tr}\left(\mathop{\mathbb{E}}_{U\sim\mu_H}[UOU^\dagger]\right) = \mathrm{Tr}(O)$

$= \mathrm{Tr}(c\mathbb{I}) = cd$

$c = \frac{\mathrm{Tr}(O)}{d}$

Given $O \in \mathcal{L}((\mathbb{C}^d)^{\otimes 2})$, we have:

$k=2$

$\mathbb{I}, \mathbb{F}$

$$\mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^{\otimes 2}OU^{\dagger\otimes 2}\right] = \underline{c_{\mathbb{I},O}}\mathbb{I} + \underline{c_{\mathbb{F},O}}\mathbb{F}, \tag{5}$$

$\mathrm{Tr}(\mathbb{I}\cdot \quad)$

$\mathrm{Tr}(\mathbb{F}\cdot \quad)$

where:

$$c_{\mathbb{I},O} = \frac{\mathbf{Tr}(O) - d^{-1}\mathbf{Tr}(\mathbb{F}O)}{d^2-1} \quad \text{and} \quad c_{\mathbb{F},O} = \frac{\mathbf{Tr}(\mathbb{F}O) - d^{-1}\mathbf{Tr}(O)}{d^2-1}. \tag{6}$$

For instance, consider a distribution ν where the set of unitaries S is discrete and each unitary has an equal probability of being chosen. In this case, we have:

$$\mathop{\mathbb{E}}_{V\sim\nu}\left[V^{\otimes k} O V^{\dagger\otimes k}\right] = \frac{1}{|S|}\sum_{V\in S} V^{\otimes k} O V^{\dagger\otimes k}. \tag{7}$$

Observation 4

A probability distribution ν is a unitary k-design if and only if:

$$\mathop{\mathbb{E}}_{V\sim\nu}\left[V^{\otimes k}\otimes V^{*\otimes k}\right] = \mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^{\otimes k}\otimes U^{*\otimes k}\right]. \tag{8}$$

k=3 → unitary 3-design → clifford group
{CNOT, H, S}

Examples of moment calculations

Example 5 (Twirling of a quantum channel is a depolarizing channel)

Let ν a unitary 2-design distribution. Consider a quantum channel $\Phi : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ and a quantum state $\rho \in \mathcal{S}\left(\mathbb{C}^d\right)$. Then:

$$\mathop{\mathbb{E}}_{U\sim\nu}\left[U^\dagger \Phi\left(U\rho U^\dagger\right) U\right] = p_\Phi \rho + (1 - p_\Phi)\,\mathbf{Tr}(\rho)\frac{I}{d}, \tag{9}$$

where the left-hand side represents the so-called twirling of Φ, and we define:

$$p_\Phi := \frac{d^2 F_e(\Phi) - 1}{d^2 - 1}. \tag{10}$$

Here, $F_e(\Phi)$ denotes the entanglement fidelity given by $F_e(\Phi) := \frac{1}{d^2}\langle\Omega|\,\Phi\otimes\mathcal{I}(|\Omega\rangle\langle\Omega|)\,|\Omega\rangle$.

$|\Omega\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d} |i\rangle\otimes|i\rangle$

Proof of Example 4

Considering a Kraus decomposition for Φ with operators $\{K_i\}_{i=1}^{d^2}$, we have:

$\Phi(\rho) = \sum_{i=1}^{d^2} K_i \rho K_i^\dagger$

$$
\begin{aligned}
\mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^\dagger \Phi\left(U\rho U^\dagger\right) U\right] &= \sum_{i=1}^{d^2} \mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^\dagger K_i U \rho U^\dagger K_i^\dagger U\right] && (11)\\
&= \sum_{i=1}^{d^2} \mathop{\mathbb{E}}_{U\sim\mu_H} \mathbf{Tr}_2\left[(I\otimes\rho) U^{\dagger\otimes 2}(K_i\otimes K_i^\dagger)U^{\otimes 2}\,\mathbb{F}\right] && (12)\\
&= \sum_{i=1}^{d^2} \mathbf{Tr}_2\left[(I\otimes\rho) \mathop{\mathbb{E}}_{U\sim\mu_H}\left(U^{\otimes 2}(K_i\otimes K_i^\dagger)U^{\dagger\otimes 2}\right)\mathbb{F}\right], && (13)
\end{aligned}
$$

k=2 moment

Swap test

where in the second equality we used that $AB = \mathbf{Tr}_2\left(A\otimes B\,\mathbb{F}\right)$, and in the third equality that $\mathop{\mathbb{E}}_{U\sim\mu_H}[f(U)] = \mathop{\mathbb{E}}_{U\sim\mu_H}\left[f(U^\dagger)\right]$ for all integrable functions f.

Proof of Example 4

Using the property of the second moment, we have:

$$\mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^{\otimes 2}\left(\sum_{i=1}^{d^2} K_i\otimes K_i^\dagger\right)U^{\dagger\otimes 2}\right] = c_{\mathbb{I}}\mathbb{I} + c_{\mathbb{F}}\mathbb{F}, \tag{14}$$

where the coefficients $c_{\mathbb{I}}$ and $c_{\mathbb{F}}$ are given by:

$$c_{\mathbb{I}} = \frac{\sum_{i=1}^{d^2}\mathbf{Tr}\left(K_i\otimes K_i^\dagger\right) - d^{-1}\mathbf{Tr}\left(\sum_{i=1}^{d^2}K_i\otimes K_i^\dagger\mathbb{F}\right)}{d^2-1} = \frac{\sum_{i=1}^{d^2}|\mathbf{Tr}(K_i)|^2 - 1}{d^2-1} \tag{15}$$

$$c_{\mathbb{F}} = \frac{\mathbf{Tr}\left(\mathbb{F}\sum_{i=1}^{d^2}K_i\otimes K_i^\dagger\right) - d^{-1}\mathbf{Tr}\left(\sum_{i=1}^{d^2}K_i\otimes K_i^\dagger\right)}{d^2-1} = \frac{d - d^{-1}\sum_{i=1}^{d^2}|\mathbf{Tr}(K_i)|^2}{d^2-1}, \tag{16}$$

where we used the *swap-trick* and the fact that $\sum_{i=1}^{d^2}K_i^\dagger K_i = I$.

Proof of Example 4

Therefore:

$$\mathop{\mathbb{E}}_{U\sim\mu_H}\left[U^\dagger\Phi\left(U\rho U^\dagger\right)U\right] = \mathbf{Tr}_2\left[(I\otimes\rho)\left(c_{\mathbb{I},}\mathbb{I}+c_{\mathbb{F}}\mathbb{F}\right)\mathbb{F}\right] \tag{17}$$

$$= c_{\mathbb{I}}\mathbf{Tr}_2\left(((I\otimes\rho)\mathbb{F}\right) + \frac{1}{d}(1-c_{\mathbb{I}})\mathbf{Tr}_2((I\otimes\rho)) \tag{18}$$

$$= c_{\mathbb{I}}\rho + (1-c_{\mathbb{I}})\mathbf{Tr}(\rho)\,\frac{I}{d}\,. \tag{19}$$

Proof of Example 4

To conclude the proof, we observe that $c_{\mathbb{I}} = p_{\Phi}$, as defined in Eq.(10). This follows from the relationship $\sum_{i=1}^{d^2} |\mathbf{Tr}(K_i)|^2 = d^2 F_e(\Phi)$, as it can be easily seen:

$$F_e(\Phi) := \frac{1}{d^2} \langle\Omega| \Phi \otimes \mathcal{I}(|\Omega\rangle \langle\Omega|) |\Omega\rangle \tag{20}$$

$$= \sum_{i=1}^{d^2} \frac{1}{d^2} \langle\Omega| K_i \otimes I |\Omega\rangle \langle\Omega| K_i^{\dagger} \otimes I |\Omega\rangle \tag{21}$$

$$= \frac{1}{d^2} \sum_{i=1}^{d^2} |\mathbf{Tr}(K_i)|^2 . \tag{22}$$

Examples of moment calculations

Example 6 (Average gate fidelity)

Let ν be a state 2-design distribution. Consider a quantum channel $\Phi : \mathcal{L}(\mathbb{C}^d) \to \mathcal{L}(\mathbb{C}^d)$ and a unitary channel $\mathcal{U}(\cdot) = U(\cdot)U^\dagger$. Then, the average gate fidelity is given by:

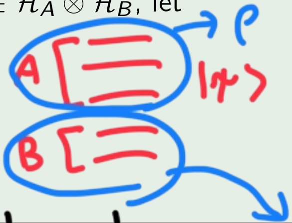
$$\mathop{\mathbb{E}}_{|\psi\rangle\sim\nu}\left[\langle\psi|\mathcal{U}^\dagger \circ \Phi(|\psi\rangle\langle\psi|)|\psi\rangle\right] = \frac{dF_e(\mathcal{U}^\dagger \circ \Phi) + 1}{d+1}, \tag{23}$$

where $\mathcal{U}^\dagger(\cdot) = U^\dagger(\cdot)U$ represents the adjoint channel of $\mathcal{U}$, and $F_e(\Phi) := \frac{1}{d^2}\langle\Omega|\Phi\otimes\mathcal{I}(|\Omega\rangle\langle\Omega|)|\Omega\rangle$ corresponds to the *entanglement-fidelity*.

Examples of moment calculations

Example 7 (Purity)

Consider the complex Hilbert space of two-qudit systems $\mathcal{H}_A \otimes \mathcal{H}_B$ of dimensions respectively $d_A = \dim(\mathcal{H}_A)$ and $d_B = \dim(\mathcal{H}_B)$. Given $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, let $\rho_A := \mathbf{Tr}_B\left(|\psi\rangle\langle\psi|\right)$. We have:



$$\mathop{\mathbb{E}}_{|\psi\rangle\sim\nu} \mathbf{Tr}\left(\rho_A^2\right) = \frac{d_B + d_A}{d_A d_B + 1}, \tag{24}$$

where ν is a 2-design distribution.

$$\simeq \frac{1}{d_A} + \frac{1}{d_B}$$

$U \sim \nu \quad U^{\otimes 2} |\phi\rangle \Big] =$

$|\psi\rangle \sim \nu \quad |\psi\rangle$

We can express the expected value as follows:

$$\mathop{\mathbb{E}}_{|\psi\rangle\sim\nu} \mathbf{Tr}(\rho_A^2) = \mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H} \mathbf{Tr}(\rho_A^{\otimes 2}\mathbb{F}_A) = \mathop{\mathbb{E}}_{|\psi\rangle\sim\nu} \mathbf{Tr}\left(|\psi\rangle\langle\psi|^{\otimes 2}(\mathbb{F}_A \otimes \mathbb{I}_B)\right), \tag{25}$$

Swap trick

Since ν is a 2-design, we have:

$$\mathop{\mathbb{E}}_{|\psi\rangle\sim\nu} \mathbf{Tr}(\rho_A^2) = \mathbf{Tr}\left(\mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H}\left[|\psi\rangle\langle\psi|^{\otimes 2}\right](\mathbb{F}_A \otimes \mathbb{I}_B)\right) = \mathbf{Tr}\left(\left[\frac{\mathbb{I}_A \otimes \mathbb{I}_B + \mathbb{F}_A \otimes \mathbb{F}_B}{d_A d_B(d_A d_B + 1)}\right](\mathbb{F}_A \otimes \mathbb{I}_B)\right) \tag{26}$$

$$= \frac{1}{d_A d_B(d_A d_B + 1)}\left(d_A d_B^2 + d_A^2 d_B\right) = \frac{d_B + d_A}{d_A d_B + 1} \tag{27}$$

$$\mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H} |\psi\rangle\langle\psi|^{\otimes 2} = \frac{I \otimes I + F \otimes F}{d(d+1)}$$

$d \to d_A d_B$

Concentration inequalities

Markov's inequality states that for a non-negative random variable X and any $\varepsilon > 0$, the probability that X exceeds ε is bounded by the ratio of the expected value of X to ε:

$$\mathrm{Prob}\,(X \geq \varepsilon) \leq \frac{\mathbb{E}\,[X]}{\varepsilon}. \tag{28}$$

In a more general form, if g is a strictly increasing non-negative function, the inequality can be expressed as:

$$\mathrm{Prob}\,(X \geq \varepsilon) \leq \frac{\mathbb{E}\,[g(X)]}{g(\varepsilon)}. \tag{29}$$

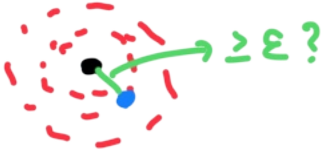
Lemma 8 (Levy's lemma)

Consider the set $\mathbb{S}^{2d-1} := \{v \in \mathbb{C}^d : \|v\|_2 = 1\}$. *Let* $f : \mathbb{S}^{2d-1} \to \mathbb{R}$ *be a function satisfying the Lipschitz condition* $|f(v) - f(w)| \leq L\|v - w\|_2$. *For all* $\varepsilon \geq 0$, *we have the probability bound:*

2^n

$$\operatorname*{Prob}_{|\phi\rangle\sim\mu_H}\left[\left|f(\phi) - \mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H}[f(\psi)]\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2d\varepsilon^2}{9\pi^3 L^2}\right). \tag{30}$$

double exponential



Examples of concentration inequalities

$$\mathbb{E}_{|\psi\rangle\sim\mu_H} \langle\psi|O|\psi\rangle = \mathbb{E}_{|\psi\rangle\sim\mu_H} \mathrm{Tr}(|\psi\rangle\langle\psi|^{\otimes 1} O) = \frac{\mathrm{Tr}(O)}{d}$$

Example 9

Let $O \in \mathrm{Herm}(\mathbb{C}^d)$ be a Hermitian operator. For all $\varepsilon > 0$, we have:

$$\underset{|\psi\rangle\sim\mu_H}{\mathrm{Prob}} \left(\left| \langle\psi| O |\psi\rangle - \frac{\mathbf{Tr}(O)}{d} \right| \geq \varepsilon \right) \leq 2\exp\left(-\frac{d\varepsilon^2}{18\pi^3 \|O\|_\infty^2} \right). \quad (31)$$

L

In particular, if O is a Pauli string $P \in \{I, X, Y, Z\}^{\otimes n} \backslash \{I^{\otimes n}\}$, we have:

$\mathrm{Tr}(P) = 0$

$$\underset{|\psi\rangle\sim\mu_H}{\mathrm{Prob}} \left(|\langle\psi| P |\psi\rangle| \geq \varepsilon \right) \leq 2\exp\left(-\frac{2^n\varepsilon^2}{18\pi^3} \right). \quad (32)$$

Proof of Example 9

To apply Levy's lemma, we consider the function $f(\psi) = \langle\psi| O |\psi\rangle$ and compute its expected value and Lipschitz constant. First, we observe that

$$\mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H}[f(\psi)] = \mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H}[\langle\psi| O |\psi\rangle] = \mathbf{Tr}[O \underline{\mathop{\mathbb{E}}_{|\psi\rangle\sim\mu_H} |\psi\rangle\langle\psi|}] = \frac{\mathbf{Tr}(O)}{\underline{d}}. \tag{33}$$

Next, we determine the Lipschitz constant. We have

$$\underline{|f(v) - f(\overset{u}{\cancel{w}})|} = \underline{|\mathbf{Tr}\left[(|u\rangle\langle u| - |v\rangle\langle v|)\, O\right]|} \leq \underline{\|O\|_\infty \||u\rangle\langle u| - |v\rangle\langle v|\|_1} \tag{34}$$

where we used the matrix Hölder inequality.

$f(v) = \langle v|O|v\rangle$

Proof of Example 9

We then observe that:

$$\| |u\rangle\langle u| - |v\rangle\langle v| \|_1 \leq 2 \|u - v\|_2 . \tag{35}$$

Hence, we have $|f(v) - f(w)| \leq 2\|O\|_\infty \|u - v\|_2$. By applying Levy's lemma with the Lipschitz constant of f being $2\|O\|_\infty$, we can conclude.

Examples of concentration inequalities

Example 10

Consider a n-qubit state $|\phi\rangle \in \mathbb{C}^d$ with $d = 2^n$. If we randomly pick a state $|\psi\rangle$ from the Haar measure, the probability that the overlap between $|\psi\rangle$ and $|\phi\rangle$ is larger than $\varepsilon > 0$ decays double exponentially with the number of qubits n:

$$\underset{|\psi\rangle \sim \mu_H}{\mathrm{Prob}} \left[|\langle\psi|\phi\rangle|^2 \geq \varepsilon \right] \leq 2\exp(-\frac{d}{2}\varepsilon). \tag{36}$$

Barren plateaus

VQA input $\to$ $\boxed{\theta}$ $\xrightarrow{\text{output}}$

NN

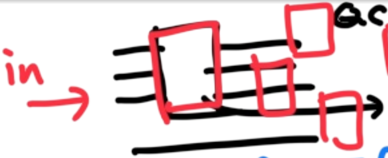
Observation 11

Let ν be a distribution defined over the set of unitaries $\{U(\theta)\}_{\theta \in \mathbb{R}^L}$. If ν forms a 2-design distribution, then the following properties hold:

$$\mathop{\mathbb{E}}_{U\sim\nu}[C(\theta)] = 0, \quad \mathop{\mathrm{Var}}_{U\sim\nu}[C(\theta)] \in O\left(\frac{\mathrm{poly}(n)}{2^n}\right). \tag{37}$$

0

$\theta \leftarrow \theta - \mu \frac{\partial C}{\partial \theta}$



$R_z(\theta) = e^{-iZ\theta}$

$C(\theta) = \mathrm{Tr}(O\, U(\theta)\, \rho_{in}\, U(\theta)^{\dagger})$

Proof of Observation 11

We have:

Pauli

$$\mathop{\mathbb{E}}_{U\sim\nu}[C(\boldsymbol{\theta})] = \mathbf{Tr}\left[\mathop{\mathbb{E}}_{U\sim\nu}\left(U(\boldsymbol{\theta})\rho_0 U^\dagger(\boldsymbol{\theta})\right) O\right] = \frac{\mathbf{Tr}(O)}{d} = 0, \tag{38}$$

where in the last step we used that O is traceless.

Proof of Observation 11

The second equality follows from $\underset{U\sim\nu}{\mathrm{Var}}[C(\boldsymbol{\theta})] = \underset{U\sim\nu}{\mathbb{E}}\left[C(\boldsymbol{\theta})^2\right] - \underset{U\sim\nu}{\mathbb{E}}[C(\boldsymbol{\theta})]^2$

$$\underset{U\sim\nu}{\mathbb{E}}\left[C(\boldsymbol{\theta})^2\right] = \mathbf{Tr}\left[\underset{U\sim\nu}{\mathbb{E}}\left(U^{\otimes 2}(\boldsymbol{\theta})\rho_0^{\otimes 2}U^{\dagger\otimes 2}(\boldsymbol{\theta})\right)O^{\otimes 2}\right] \tag{39}$$

$$= c_{\mathbb{I},\rho_0^{\otimes 2}}\mathbf{Tr}\left[\mathbb{I}O^{\otimes 2}\right] + c_{\mathbb{F},\rho_0^{\otimes 2}}\mathbf{Tr}\left[\mathbb{F}O^{\otimes 2}\right] \tag{40}$$

$$= c_{\mathbb{F},\rho_0^{\otimes 2}}\mathbf{Tr}\left[O^2\right], \tag{41}$$

$\frac{1}{2^{2n}}$

where $c_{\mathbb{F},\rho_0^{\otimes 2}} = \frac{\mathbf{Tr}\left(\mathbb{F}\rho_0^{\otimes 2}\right) - d^{-1}\mathbf{Tr}\left(\rho_0^{\otimes 2}\right)}{d^2-1} = \frac{\mathbf{Tr}\left(\rho_0^2\right) - 2^{-n}}{2^{2n}-1}$, and in the last step we used that O is traceless. The proof is concluded by noting that $c_{\mathbb{F},\rho_0^{\otimes 2}} \leq \frac{1}{2^n(2^n+1)}$ and using that $\mathbf{Tr}\left[O^2\right] \in O\left(\mathrm{poly}(n)2^n\right)$.

$$\frac{1}{2^{2n}} \cdot O(\mathrm{poly}(n) \cdot 2^n)$$
$$= O\left(\frac{\mathrm{poly}(n)}{2^n}\right)$$

Probability of finding a large point in cost function

We can then apply Chebyshev's inequality, which states that for all $\varepsilon > 0$ we have:

$$\operatorname*{Prob}_{U\sim\nu}\left(\left|C(\boldsymbol{\theta}) - \mathop{\mathbb{E}}_{U\sim\nu}[C(\boldsymbol{\theta})]\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\operatorname*{Var}_{U\sim\nu}[C(\boldsymbol{\theta})]. \tag{42}$$

This inequality provides an upper bound on the probability of encountering a point in the parameter space where the cost function deviates from its expected value by more than ε. In particular, the probability of finding a point with a cost function larger than ε decays exponentially with the number of qubits: $\operatorname*{Prob}_{U\sim\nu}(|C(\boldsymbol{\theta})| \geq \varepsilon) \in O\left(\varepsilon^{-2}\frac{\mathrm{poly}(n)}{2^n}\right)$.

Barren plateaus

$$\theta \leftarrow \theta - \underbrace{\mu}_{\text{constant}} \cdot \frac{\partial C(\theta)}{\partial \theta}$$

Similarly, with slightly more involved calculations, we can show that the exponential decay also applies to the variance of the partial derivatives of the cost function. This phenomenon, where the variance of the partial derivatives of the cost function decays exponentially with the number of qubits n, is commonly referred to as **Barren Plateaus**.

Barren plateaus

To analyze the partial derivatives of the cost function, we can express the parameterized unitary circuit $U(\boldsymbol{\theta})$ as the product of two unitary operators: $U(\boldsymbol{\theta}) = U_A U_B$, where $U_A = \prod_{l=\mu+1}^{L} e^{-i\theta_l H_l}$ and $U_B = \prod_{l=1}^{\mu} e^{-i\theta_l H_l}$. Consequently, we can write the partial derivative of the cost function as follows:

$$\partial_\mu C(\boldsymbol{\theta}) = \mathrm{Tr}\left[(\partial_\mu U(\boldsymbol{\theta}))\,\rho_0 U^\dagger(\boldsymbol{\theta}) O\right] + \mathrm{Tr}\left[U(\boldsymbol{\theta})\rho_0 \left(\partial_\mu U^\dagger(\boldsymbol{\theta})\right) O\right] \tag{43}$$

$$= -i\,\mathrm{Tr}\left[U_A H_\mu U_B \rho_0 U_B^\dagger U_A^\dagger O\right] + i\,\mathrm{Tr}\left[U_A U_B \rho_0 U_B^\dagger H_\mu U_A^\dagger O\right] \tag{44}$$

$$= i\,\mathrm{Tr}\left[U_B \rho_0 U_B^\dagger \left[H_\mu, U_A^\dagger O U_A\right]\right], \tag{45}$$

where we denoted by ∂_μ the partial derivative with respect to θ_μ, we used that $\partial_\mu U(\boldsymbol{\theta}) = -iU_A H_\mu U_B$ and the cyclicity of the trace. Using this expression for the partial derivative of the cost function, we can prove the following:

Barren plateaus

Observation 12

Let ν_A, ν_B be probability distributions defined over the sets of unitaries $\{U_A(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathbb{R}^{L-\mu}}$ and $\{U_B(\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathbb{R}^{\mu}}$, respectively. Suppose that both ν_A and ν_B are 2-designs distributions. In this case, the following properties hold:

$$\mathop{\mathbb{E}}_{\substack{U_A\sim\nu_A\\ U_B\sim\nu_B}}[\partial_\mu C(\boldsymbol{\theta})] = 0, \quad \mathop{\mathrm{Var}}_{\substack{U_A\sim\nu_A\\ U_B\sim\nu_B}}[\partial_\mu C(\boldsymbol{\theta})] \in O\left(\frac{\mathrm{poly}(n)}{2^n - 1}\right). \tag{46}$$

QCNN

B.P. (×) $\longleftrightarrow$ classical (○) simulability

Classical shadow tomography

ρ exponential $2^n = d$

$\mathrm{tr}(O\rho)$

Let $\rho \in \mathcal{S}\left(\mathbb{C}^d\right)$, and let $O_1, \ldots, O_M \in \mathrm{Herm}\left(\mathbb{C}^d\right)$. The goal is to estimate $\mathbf{Tr}(O_1\rho), \ldots, \mathbf{Tr}(O_M\rho)$ with a desired accuracy and probability of success.

We assume that the full classical description of the state ρ is unknown but it can be queried on a quantum device multiple times. When the state ρ is queried, a unitary U is sampled randomly from a probability distribution μ and it is applied to ρ. The resulting state, $U\rho U^\dagger$, is measured in the computational basis $\{|b\rangle\}_{b\in[d]}$.

$\langle b| U \rho U^\dagger |b\rangle$

The state $U^\dagger|b\rangle\langle b|U$ is referred to as a **classical snapshot**.

$U^\dagger |b\rangle\langle b| U$

Now, the expected value of the classical snapshot $\mathbb{E}\left[U^\dagger|b\rangle\langle b|U\right]$ is considered, where U is distributed according to the probability distribution μ, and b is distributed according to the Born's rule probability distribution $\langle b| U\rho U^\dagger |b\rangle$. We can define the measurement channel $\mathcal{M}$ as:

Prob Snapshot

$$\mathcal{M}(\rho) := \sum_{b=1}^{d} \mathbb{E}_{U\sim\mu}\left[\langle b| U\rho U^\dagger |b\rangle U^\dagger|b\rangle\langle b|U\right]. \tag{47}$$

Classical shadow tomography

$\left(\rho \xrightarrow{U} \{|b\rangle\}\ \text{measure}\right) = \mathcal{M}$

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

Assuming that $\mathcal{M}$ is invertible, $\hat{\rho} := \mathcal{M}^{-1}\left(U^\dagger |b\rangle\langle b| U\right)$ serves as an unbiased estimator for ρ, meaning $\mathbb{E}[\hat{\rho}] = \rho$. The matrix $\hat{\rho}$ is commonly known as the *classical shadow* of the state ρ. Consequently, $\hat{o}_i := \mathbf{Tr}(O_i\hat{\rho})$ is an unbiased estimator for $\mathbf{Tr}(O_i\rho)$:

$$\hat{o}_i := \mathbf{Tr}\left(O_i\mathcal{M}^{-1}\left(U^\dagger|b\rangle\langle b|U\right)\right) \quad \text{implies} \quad \mathbb{E}[\hat{o}_i] = \mathbf{Tr}(O_i\rho) \text{ for all } i \in [M]. \tag{48}$$

For appropriately chosen probability distributions μ, the estimator $\hat{o}_i$ can be efficiently computed classically.

Sample ∞

error $< \epsilon$

sample number N ??

Classical shadow tomography

To estimate the number N of copies of ρ (sample complexity) to achieve an additive accuracy $\varepsilon > 0$ in the estimation of $\mathbf{Tr}(O_i\rho)$ for all $i \in [M]$, with a failure probability of at most $\delta > 0$, it is important to bound the variance of the estimator:

$$\mathrm{Var}\left(\hat{o}_i\right) := \mathbb{E}\left[\hat{o}_i^2\right] - \mathbb{E}\left[\hat{o}_i\right]^2 . \tag{49}$$

If the median of means is used as the estimator to post-process the data $\hat{o}_i$ for each $i \in [M]$, then a number of copies

$$N = O\left(\frac{\log\left(2M/\delta\right)}{\varepsilon^2} \max_{i\in[m]} \left[\mathrm{Var}\left(\hat{o}_i\right)\right]\right) \tag{50}$$

is enough to estimate, for each $i \in [m]$, $\mathbf{Tr}(O_i\rho)$ up to precision ε with success probability at least $1 - \delta$.

Classical shadow tomography

Observation 13

2-design

$$\mathcal{M}(\rho) = \mathbf{Tr}_1\left((\rho\otimes I)\left(\sum_{b=1}^{d}\mathop{\mathbb{E}}_{U\sim\mu}\left[U^{\dagger\otimes 2}|b\rangle\langle b|^{\otimes 2}U^{\otimes 2}\right]\right)\right) \rightarrow \mathcal{M}^{-1} \tag{51}$$

$$\mathrm{Var}(\hat{o}_i) = \mathbf{Tr}\left((\rho\otimes\mathcal{M}^{-1}(O_i)\otimes\mathcal{M}^{-1}(O_i))\left(\sum_{b=1}^{d}\mathop{\mathbb{E}}_{U\sim\mu}\left[U^{\dagger\otimes 3}|b\rangle\langle b|^{\otimes 3}U^{\otimes 3}\right]\right)\right) - \mathbf{Tr}(O_i\rho)^2. \tag{52}$$

unitary 3-design

=

$\{CNOT, H, S\}$ Clifford group

Classical shadow tomography

Since the uniform distribution over the Clifford group is an exact 3-design, its first three moments coincide with those of the Haar measure.
Thus, we need to insert the formula for the second moment over the Haar measure to find the expression of the measurement channel $\mathcal{M}(\rho)$ and then invert it.

Observation 14

The measurement channel is:

$$\mathcal{M}(\rho) = \frac{1}{d+1}\left(\mathbf{Tr}(\rho)\, I + \rho\right). \tag{53}$$

Thus, its inverse is:

$$\mathcal{M}^{-1}(\rho) = (d+1)\rho - \mathbf{Tr}(\rho)\, I. \tag{54}$$

Classical shadow tomography

To bound the variance we need to compute a third moment over the Haar measure of the unitary group, due to the 3-design property of the Clifford group.

Observation 15

The variance is bounded by $\mathrm{Var}\left(\hat{o}_i\right) \leq 3\mathbf{Tr}\left(O_i^2\right)$.

Using the previous bound on the variance, we have that a number of copies ρ

$$N = O\left(\varepsilon^{-2}\log\left(2M/\delta\right)\max_{i\in[m]}\left[\mathbf{Tr}\left(O_i^2\right)\right]\right) \tag{55}$$

suffices to estimate, for each $i \in [m]$, $\mathbf{Tr}(O_i\rho)$ up to precision ε and with success probability at least $1-\delta$.

$O = \underbrace{Z\otimes Z\otimes}_{\text{local}} \cdots \otimes I \otimes I$

$O_1 = \underbrace{Z \otimes Z \otimes \cdots \otimes Z}_{n \text{ times}}$

$\mathrm{Tr}(O_1^2) = 2^n$

(x)

Thanks a lot!