PROOF OF BOUND OF VARIANCE

Proof: $tr(\hat{\rho}) = 1 \rightarrow Var(\hat{o})$ only depends on traceless

part: $O_0 = O - \frac{tr(O)}{2^n} \mathbb{I}$

$$\hat{o} - E(\hat{o}) = tr(O\hat{\rho}) - tr(O\rho) = tr\left(\left(O_0 + \frac{tr(O)}{2^n}\mathbb{I}\right)\hat{\rho}\right)$$

$$- tr\left(\rho\left(O_0 + \frac{tr(O)}{2^n}\mathbb{I}\right)\right) = tr(O_0\hat{\rho}) - tr(O_0\rho)$$

M^{-1} is self adjoint

M is invertible, if M is self adjoint, M^{-1} will be too.

$$tr(XM(Y)) = \sum_i \langle i| X \, E\sum_b \langle b|UYU^\dagger|b\rangle U^\dagger|b\rangle\langle b|U|i\rangle$$

$$= \sum_i E \sum_b \langle i| X U^\dagger|b\rangle\langle b|U|i\rangle \langle b|UYU^\dagger|b\rangle$$

$$= \sum_i E\sum_b \langle b|U|i\rangle\langle i|XU^\dagger|b\rangle\langle b|UYU^\dagger|b\rangle$$

$$= E\sum_b \langle b|UXU^\dagger|b\rangle\langle b|UYU^\dagger|b\rangle$$

$$tr(M(X)Y) = \sum_i \langle i| E\sum_b \langle b|UXU^\dagger|b\rangle U^\dagger|b\rangle\langle b|UY|i\rangle$$

$$= \sum_i E\sum_b \langle b|UY|i\rangle\langle i|U^\dagger|b\rangle\langle b|UXU^\dagger|b\rangle$$

$$= E\sum_b \langle b|UYU^\dagger|b\rangle\langle b|UXU^\dagger|b\rangle$$

$MM^{-1} = I$ $\qquad (M^{-1})^\dagger M^\dagger = I$

$M^\dagger = M$ $\qquad (M^{-1})^\dagger = (M^\dagger)^{-1} = M^{-1}$

$$\mathrm{Var}(\hat{o}) = E(\hat{o} - E(\hat{o}))^2 = E(\hat{o}^2) - E(\hat{o})^2$$

$$= E\big(\underbrace{(\mathrm{tr}\, O_0 \hat{\rho})^2}\big) - \mathrm{tr}(O_0 \underbrace{E(\hat{\rho})}_{\rho})^2 = -$$

$$\mathrm{tr}(O_0 \, M^{-1}(U^\dagger |\hat{b}\rangle\langle\hat{b}| U))^2 = \mathrm{tr}\left(M^{-1}(O_0)\, U^\dagger |\hat{b}\rangle\langle\hat{b}| U\right)^2$$

$$= \Big(\sum_i \langle i| M^{-1}(O_0)\, U^\dagger |\hat{b}\rangle\langle\hat{b}| U |i\rangle\Big)^2 = (\langle\hat{b}| U M^{-1}(O_0) U^\dagger |\hat{b}\rangle)^2$$

$$\sum_i |i\rangle\langle i| = \mathbb{1}$$

$$= E(\langle\hat{b}| U M^{-1}(O_0) U^\dagger |\hat{b}\rangle^2) - [\mathrm{tr}(O_0 \rho)]^2$$

?

Randomness comes from $\begin{cases} U \sim \mathcal{U} & \text{classically} \\ \text{which } |b\rangle\langle b| & \text{we measure (Quant.)} \end{cases}$

$$E(\langle\hat{b}| U M^{-1}(O_0) U^\dagger |\hat{b}\rangle^2) =$$

$$\underset{U \sim \mathcal{U}}{E} \sum_{b \in \{0,1\}^n} \overbrace{\langle b| U \rho U^\dagger |b\rangle}^{\text{Prob of getting } |b\rangle\langle b|} \langle\hat{b}| U M^{-1}(O_0) U^\dagger |\hat{b}\rangle^2$$

dependence on ρ!

We just take the max

By removing $(\mathrm{tr}(O_0\rho))^2$

$$\to \mathrm{Var}(\hat{o}) \leq \left\| O - \frac{\mathrm{tr}\, O}{2^n} \mathbb{I} \right\|^2_{\text{shadow}}$$

PROOF USING MEDIAN OF MEANS

→ Median of means.

Given $\hat{\mu}_{MOM}$ from a total of $n_T = \underbrace{K}_{\text{Part}} \cdot \underbrace{N}_{\text{elements in each part}}$

and assuming $Var(X_1) < \infty$

$$P(|\hat{\mu}_{MOM} - \mu_0| > \varepsilon) \leq e^{-2K\left(\frac{1}{2} - \frac{K}{n_T}\frac{\sigma^2}{\varepsilon^2}\right)^2}$$

In the paper, they look for $P \leq 2e^{-\frac{K}{2}}$ which is supposed to happen in the worst situation.

$$\leq e^{-2K\left(\underbrace{\frac{1}{2} - \frac{K}{n_T}\frac{\sigma^2}{\varepsilon^2}}_{\ll 1}\right)^2} \leq 2e^{-\frac{K}{2}}$$

n_T!

worst case

$$\frac{K}{n_T}\frac{\sigma^2}{\varepsilon^2} \ll 1 \rightarrow \frac{K}{K \cdot N}\frac{\sigma^2}{\varepsilon^2} \ll 1 \quad N \sim Const \cdot \frac{\sigma^2}{\varepsilon^2}$$

In the paper, they use $N = 34\frac{\sigma^2}{\varepsilon^2}$ as a safe number.

Using $Pr\left(\bigcup_{i=1}^{M} |\hat{O}_i - tr(O_i \rho)| \geq \varepsilon\right) \leq \sum_{i=1}^{M} Pr(|\hat{O}_i - tr(O_i\rho)| \geq \varepsilon)$

$$= M \underbrace{Pr(|\hat{O}_i - tr(O_i\rho)| \geq \varepsilon)}_{\text{We want this } \frac{\delta}{M}}$$

$$\underbrace{Pr(|\hat{\mu} - \mu_0| \geq \varepsilon)}_{\frac{\delta}{M}} \leq 2e^{-K/2} = \frac{\delta}{M}$$

$$\Rightarrow K = 2\log\frac{2M}{\delta}$$

We conclude that we need $n_T = N \cdot K$

with $N = \frac{34}{\varepsilon^2} \sigma^2 = \frac{34}{\varepsilon^2} \max_{1 \leq i \leq M} \left\| O_i - \frac{tr(O_i)}{2^n} I \right\|_{shadow}^2$

and

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

to learn

$$|\hat{o}_i(N,K) - tr(O_i \rho)| \leq \varepsilon \quad \forall \ 1 \leq i \leq M$$

with prob at least $1 - \delta$

In general:

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

LET'S TALK ABOUT UNITARY T-DESIGNS

Unitary t-designs

→ Classical spherical t-designs

$p_t : S(\mathbb{R}^d) \longrightarrow \mathbb{R}$

$S(\mathbb{R}^d)$

$p_t(x)$ max degree $\qquad x \in S(\mathbb{R}^d)$

We want to avg p_t over all points of the sphere.
Can we do that with a set of finite points?

YES!

$X = \{x : x \in S(\mathbb{R}^d)\}$ is a spherical t-design if

$$\frac{1}{|X|}\sum_{x\in X} p_t(x) = \int_{S(\mathbb{R}^d)} p_t(u)\,d\eta(u) \quad \text{f.a. } p_t(u)$$

Higher t imply bigger set X

E.g. in $\mathbb{R}^3$, a cube is a 3-design
a dodecahedron is a 5-design.

→ Complex projective designs

Extend the idea to $S(\mathbb{C}^d)$. The polynomials have degree at most t for the entries and complex conj.

X, a complex projective t-design, is a subset of X of $S(\mathbb{C}^d)$ s.t.

$$\frac{1}{|X|}\sum_{x\in X} p_{t,t}(x) = \int_{S(\mathbb{C}^d)} p_{t,t}(u)\,d\eta(u) \quad \text{f.a. } p_{t,t}(u)$$

→ Unitary t-design.

Our Sphere is now the unitary group $U(d)$

Polynomials act on elements $\in U(d)$.

A unitary t-design is the set of points $\in U(d)$ that suffice to compute the average of $P_{t,t}$ (max degree of t in the entries of U or its complex conj) over all the unitary transformations.

$\{U_k\}_1^K$ is unitary t-design if

$$\frac{1}{K}\sum_{k=1}^{K} P_{t,t}(U) = \int_{U(d)} \underbrace{d\mu(U)}_{\text{Haar measure of the group}} P_{t,t}(U)$$

→ Where does it appear?

Avg fidelity:

Λ noisy channel ~ V (perfect)

Let's see how it affect $|0\rangle$ state: $\begin{cases} V|0\rangle\langle 0|V^\dagger \\ \Lambda(|0\rangle\langle 0|) \end{cases}$

Fidelity for the 0 state: $\mathrm{Tr}\left(V|0\rangle\langle 0|V^\dagger \Lambda(|0\rangle\langle 0|)\right)$

$= \langle 0|V^\dagger \Lambda(|0\rangle\langle 0|)V|0\rangle$

Only $|0\rangle$ state ? → Avg over all the states

$|Rand\rangle = \underbrace{U}_{\text{random}}|0\rangle \sim U(2)$

$$\bar{F}(\Lambda, V) = \int_U d\mu(U)\, \langle 0|U^\dagger V^\dagger \Lambda(U|0\rangle\langle 0|U^\dagger)VU|0\rangle$$

This is a polynomial of degree 2 in the entries of U & $U^\dagger$ $\rightarrow$ We can use a u. 2 design.

Clifford group is a 3-design (also a 2, 1-design) (NOT 4)

Clifford group: Normalizer of the Pauli group. Sends Paulis to Pauli (up to a phase) via conjugation

$$C P C^\dagger = \pm P' \qquad \underbrace{C \in \mathcal{C}\ell(2^n)}_{\text{Clifford g. of n qubit}} \qquad \underbrace{P, P^\dagger \in \mathcal{P}(2^n)}_{\text{Pauli group on n qubits}}$$

$$P = \sigma_1 \otimes \dots \otimes \sigma_n \quad \in \mathcal{P}(1)$$

Generated by H, S, CNOT

$\rightarrow$ Gottesman-Knill theorem. Circuits with preparation and measurements in the comp. basis + Clifford gates can be efficiently simulated on a classical comp.
The key idea is to keep track of how operators change instead of the state.

We only need to evaluate the fidelity on the states that $V \in \mathcal{C}\ell(2^1)$ generates by $U|0\rangle$!

There are 24 elements in $\mathcal{C}\ell(2)$
" 11520 in $\mathcal{C}\ell(4)$

For our interests

$$\mathbb{E}_{U\sim C(2^n)} (UAU)^{\otimes k} = \int_{U(d)} (UAU^\dagger)^{\otimes k}\, d\mu(U) \qquad \forall\, 2^n \times 2^n \text{ mat } A$$

$$k = 1,2,3$$

There are explicit formulas for:

$$\mathbb{E}_{U\sim C(2^n)} U^\dagger |x\rangle\langle x| U \langle x|UAU^\dagger|x\rangle = \frac{A + tr(A) I}{(2^n+1)2^n} * = \frac{1}{2^n} D_{1/(2^n+1)}(A)$$

for $A \in H_{2^n}$

$$\mathbb{E}_{U\sim C(2^n)} U^\dagger |x\rangle\langle x| U \langle x|U B_0 U^\dagger|x\rangle\langle x|U C_0 U^\dagger|x\rangle = \frac{tr(B_0 C_0) I + B_0 C_0 + C_0 B_0}{(2^n+2)(2^n+1)2^n}$$

for $B_0, C_0 \in H_{2^n}$

* this is eq to a depolarizing ch. of n qubits with loss prob. p

$$D_p(A) = pA + (1-p)\frac{tr(A)}{2^n} I$$

THIS CAN BE INVERTED!

$$D^{-1}_{1/(2^n+1)}(A) = (2^n+1)A - tr(A) I$$

SHADOWS NORM FOR GLOBAL CLIFFORD UNITARIES

Proof:

We already know that

$$\hat{\rho} = \mathcal{M}^{-1}\left(U^\dagger |\hat{b}\rangle\langle \hat{b}| U \right) = \underline{(2^n+1) U^\dagger |\hat{b}\rangle\langle\hat{b}| U - \mathbb{1}}$$

and $\mathcal{M}^{-1}(O_0) = (2^n+1) O_0$ $\quad \underbrace{O_0}_{\text{traceless}} \in \mathbb{H}_{2^n}$

→ Let's compute the shadow norm

$$\|O_0\|_{sh.}^2 = \max_{\sigma} \Big| \mathbb{E}_{U \in \mathcal{U}} \sum_b \langle b| U \sigma U^\dagger |b\rangle \langle b| U \mathcal{M}^{-1}(O) U^\dagger |b\rangle^2 \Big|^{1/2}$$

$\sum_i |i\rangle\langle i| = \mathbb{1}$

$$= \max_{\sigma} \left[\operatorname{tr}\left(\sigma \sum_b \mathbb{E}\, \underline{U^\dagger |b\rangle\langle b| U}\ \underline{\langle b| U \mathcal{M}^{-1}(O) U^\dagger |b\rangle^2} \right] \right]^{1/2}$$

We can use the fact that $\mathcal{C}\ell$ is a unitary 3-design

$$\mathbb{E}_{U \sim \mathcal{C}\ell(2^n)} \underline{U^\dagger |x\rangle\langle x| U} \langle x| U B_0 U^\dagger |x\rangle\langle x| U C_0 U^\dagger |x\rangle =$$

$$= \frac{\operatorname{tr}(B_0 C_0) I + B_0 C_0 + C_0 B_0}{(2^n+2)(2^n+1)2^n} \quad \text{for } B_0, C_0 \text{ traceless } \in \mathbb{H}_{2^n}$$

If $B_0 = C_0 = O_0$ $\quad$ from the sum

$$= \max_{\sigma} \operatorname{tr}\left(\sigma \frac{2^n (2^n+1)^2 \left(\operatorname{tr}(O_0^2) I + 2 O_0^2 \right)}{(2^n+2)\cdot(2^n+1)2^n} \right)$$

$$= \left(\frac{2^n+1}{2^n+2} \right) \max_{\sigma} \left(\underbrace{\operatorname{tr}(\sigma)}_{1} \operatorname{tr}(O_0^2) + 2 \operatorname{tr}(\sigma O_0^2) \right)$$

Now we use $\max\limits_{\substack{\text{state}\\ \sigma}} (\sigma O_0^2) \equiv \underbrace{\|O_0^2\|_\infty}_{\text{spectr. norm}} \leq \operatorname{tr}(O_0^2)$

$$\rightarrow \operatorname{tr}(O_0^2) \leq \|O_0\|_{shadow}^2 \leq 3 \operatorname{tr}(O_0^2)$$

SCALING FULL PROCESS TOMOGRAPHY

For a $\mathcal{E}$ acting on n-qubits, we need to learn a ρ on a $2n$-qubits system.

A reduced k-qubit ρ_k can be expressed as

$$\rho_k = \frac{1}{2^{2k}} \sum_i^{4^{2k}} \alpha_i O_i^{(2k)}$$

w. $\underbrace{\alpha_i}_{} = tr\left(\rho_k O_i^{(2k)}\right) = tr\left(\rho O_i^{(2k)} \otimes \mathbb{1}_{2n-2k}\right)$

We estimate this w. classical shadows

$$\hat{\rho}_k = \frac{1}{4^k} \sum^{16^k} \hat{\alpha}_i O_i^{(2k)}$$

$$\langle \hat{\rho}_k - \rho_k, \hat{\rho}_k - \rho_k \rangle_F = \langle \frac{1}{4^k} \sum_i^{16^k} (\hat{\alpha}_i - \alpha_i) O_i^{2k}, \frac{1}{4^k} \sum_j (\hat{\alpha}_j - \alpha_j) O_j^{2k} \rangle_F$$

$$= \frac{1}{4^{2k}} \sum_{ij} (\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j) \langle O_i^{2k}, O_j^{2k} \rangle_F$$

$$= \frac{1}{4^{2k}} \cdot 4^k \sum_i^{16^k} (\hat{\alpha}_i - \alpha_i)^2 = \frac{1}{4^k} \sum_i^{16^k} (\hat{\alpha}_i - \alpha_i)^2$$

Including the normalization in the previous theorem;

$$|\hat{\alpha}_i - \alpha_i| = |\hat{\alpha}_i - tr(\rho O_i^{2k} \otimes \mathbb{1}_{2n-2k})| \leq \frac{\varepsilon}{2^k} \; \forall i$$

$\cancel{\quad}^2$

By substitution:

$$\langle \hat{\rho}_k - \rho_k, \hat{\rho}_k - \rho_k \rangle_F \leq \frac{16^k}{4^k} \frac{\varepsilon^2}{2^{2k}} = \varepsilon^2$$

$$\Rightarrow \| \hat{\rho}_k - \rho_k \|_F \leq \varepsilon$$

For Random global Clifford:

$$\| O - \frac{tr(O)}{2} \mathbb{1} \|^2_{shadow} \leq 3\, tr(O^2) = 3 \cdot 4^n \text{ for } O = O_i^{(2k)}$$

$$O_i = 2^{2n}\underbrace{\begin{pmatrix} \underbrace{\boxed{}\,2^{2k}}_{2^{2k}} & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}}_{2^{2n}} \rightarrow O_i^2 = 2^{2n}\underbrace{\begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}}_{2^{2n}}$$

$$tr(O_i^2) = 2^{2n} = 4^n$$

$$N = \frac{68}{\varepsilon^2} \log(2M/\delta) \max_i \left\| O_i - \frac{tr(O_i)}{d} I \right\|^2_{shadow}$$

→ global Clifford

$$N = \frac{68}{\varepsilon^2} 3 \cdot 4^n \log(2M/\delta) \sim O(4^n) \text{ with } n = \text{no. of qubits}$$