CHAPTER 6

Predicting properties using classical shadows

In the previous lecture, we saw how to achieve sample-optimal quantum state tomography using multi-copy measurements and tools from representation theory. While this approach is optimal for learning a full description of the state ρ, the sample complexity of $\mathcal{O}(d^2)$ for achieving a constant error is daunting for existing quantum devices with a hundred or more qubits. Furthermore, the required entangled measurements across many copies of ρ are not readily available in many quantum platforms.

In many practical scenarios, however, we do not actually need a full description of the quantum state. Instead, we are often interested in predicting a number of properties, such as the expectation values $\mathrm{tr}(O_i\rho)$ for a given list of observables $\{O_i\}_{i=1}^M$. This shifts the goal from learning the state to predicting its properties. This task, first introduced by Aaronson [**Aar18**], is called **shadow tomography** and will be the subject of the next two lectures.

Remarkably, there is an efficient method for this task that relies only on simple, single-copy measurements. The number of measurements required will depend not on the Hilbert space dimension d, but only on the number of properties M we wish to predict. Furthermore, the dependence on M will be very favorable. An initial version of the protocol for k-local Pauli observables called **quantum overlapping tomography** was developed in [**CW20**] and refined in [**EHF19**]. This was later absorbed into the more general framework of **classical shadows**, proposed in [**HKP20**], which enables the prediction of a larger class of observables with a more flexible protocol (which was partially anticipated by the work of [**Wri16**, Chapter 5.1]). Classical shadow tomography is now widely studied and used, and is the subject of this section.

1. How to Predict Properties?

Before introducing the classical shadow formalism, let us consider a straightforward approach that directly measures each observables O_i on fresh copies of ρ. Given a set of M observables $\{O_i\}_{i=1}^M$, the strategy estimates each expectation value $\mathrm{tr}(O_i\rho)$ independently by repeatedly measuring observable O_i.

For simplicity, assume each observable is normalized such that its eigenvalues lie in $[-1, 1]$. A measurement of such an observable yields an outcome λ with probability p_λ, and the expectation value is $\sum \lambda p_\lambda$. Assume the total number of copies of ρ is N. To estimate the expectation value, one can perform N/M independent measurements of O_i on N/M fresh copies of the state ρ, yielding outcomes $\{\lambda_1, \ldots, \lambda_{N/M}\}$. The empirical mean $\hat{o}_i = \frac{M}{N}\sum_{j=1}^{N/M} \lambda_j$ serves as the estimate for $\mathrm{tr}(O_i\rho)$. To estimate all M properties, we repeat this process for each observable. A pseudocode for this algorithm is given in Algorithm 2.

Algorithm 2: DIRECTMEASUREMENT(N, ρ, $\{O_i\}_{i=1}^M$)

Input: Access to N copies of state ρ, observables $\{O_i\}_{i=1}^M$
Output: Estimates $\{\hat{o}_i\}_{i=1}^M$

```
for i = 1, ..., M do
    Initialize an empty list Λ_i;
    for j = 1, ..., N/M do
        Take a fresh copy of ρ;
        Measure the observable O_i to get an outcome λ_j;
        Append λ_j to Λ_i;
    end
    Compute the empirical average ô_i = (M/N) Σ_{λ∈Λ_i} λ;
end
return {ô_i}_{i=1}^M
```

This direct approach is simple but can be inefficient, especially when M is large. Its sample complexity is given by the following lemma.

Lemma 93 (Sample Complexity of Direct Measurement)**.** *To estimate M observables $\{O_i\}_{i=1}^M$ with $\|O_i\|_\infty \leq 1$ to an additive error ϵ with total failure probability at most δ, the direct measurement strategy requires a total of*

$$N = \mathcal{O}\left(\frac{M \log(M/\delta)}{\epsilon^2}\right)$$

quantum measurements.

PROOF. Consider a single observable O_i. A single measurement yields a random outcome X_j with $\mathbb{E}[X_j] = \operatorname{tr}(O_i\rho)$ and $|X_j| \leq 1$. We take N/M samples and compute the empirical mean $\hat{o}_i = \frac{M}{N}\sum_{j=1}^{N/M} X_j$. By Hoeffding's inequality for bounded random variables, the probability of a large deviation is bounded by:

$$\Pr[|\hat{o}_i - \operatorname{tr}(O_i\rho)| > \epsilon] \leq 2\exp\left(-\frac{2(N/M)\epsilon^2}{(1-(-1))^2}\right) = 2\exp\left(-\frac{(N/M)\epsilon^2}{2}\right).$$

To ensure this failure probability is less than some δ', we must choose (N/M) such that $2e^{-(N/M)\epsilon^2/2} \leq \delta'$, which means $(N/M) \geq \frac{2}{\epsilon^2}\log(2/\delta')$. Thus, for each observable, we need $(N/M) = \mathcal{O}(\log(1/\delta')/\epsilon^2)$ measurements. To bound the total failure probability for all M observables by δ, we use a union bound. We set the failure probability for each individual observable to $\delta' = \delta/M$. The number of measurements for each observable is then:

$$(N/M) = \mathcal{O}\left(\frac{\log(M/\delta)}{\epsilon^2}\right).$$

Since we perform this procedure independently for each of the M observables, the total number of measurements is $N = \mathcal{O}(M\log(M/\delta)/\epsilon^2)$. □

The linear dependence on M makes this approach costly when many properties are of interest. The classical shadow formalism shows that an exponential improvement is possible, replacing the linear scaling in M with a logarithmic scaling.

2. Classical Shadow Formalism

The core of the classical shadow formalism is to use a tomographically complete set of randomized, single-copy measurements to construct an unbiased estimator for the unknown state ρ. This estimator, which we call a classical snapshot, is a classical data structure representing a $2^n \times 2^n$-size Hermitian matrix that can be stored and manipulated on a conventional computer.

2.1. Measurement Channel and Classical Snapshot

Definition 94 (Measurement Channel and Classical Snapshot)**.** *Let $\mathcal{U}$ be an ensemble of n-qubit unitary operators. Let ρ be an unknown n-qubit quantum state. Consider the procedure of drawing $U \sim \mathcal{U}$, applying U to the state ρ, and measuring $U\rho U^\dagger$ in the computational basis to obtain a bitstring $b \in \{0,1\}^n$. This defines a quantum channel $\mathcal{M}$:*

$$\mathcal{M}(\rho) = \mathbb{E}_{U\sim\mathcal{U}}\left[\sum_{b\in\{0,1\}^n} \mathrm{tr}(|b\rangle\langle b|U\rho U^\dagger)\cdot U^\dagger|b\rangle\langle b|U\right].$$

If $\mathcal{U}$ is tomographically complete, $\mathcal{M}$ is invertible. For a single experimental outcome $(U,\hat{b})$, the classical snapshot is defined as

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

The classical snapshot $\hat{\rho}$ can be stored on a classical computer by storing a classical description for the unitary U (e.g., a circuit description) and the n-bit string $\hat{b}$. A set of N snapshots, $S(\rho;N) = \{\hat{\rho}_1,\ldots,\hat{\rho}_N\}$, forms the classical shadow of ρ.

Lemma 95. *The classical snapshot $\hat{\rho}$ is an unbiased estimator of the state ρ.*

Proof. By linearity of expectation,

$$\mathbb{E}[\hat{\rho}] = \mathbb{E}_{U,\hat{b}}[\mathcal{M}^{-1}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U)] = \mathcal{M}^{-1}\left(\mathbb{E}_{U,\hat{b}}[U^\dagger|\hat{b}\rangle\langle\hat{b}|U]\right).$$

The expectation over the measurement outcome $\hat{b}$ for a fixed U is

$$\mathbb{E}_{\hat{b}}[U^\dagger|\hat{b}\rangle\langle\hat{b}|U] = \sum_{b\in\{0,1\}^n} \mathrm{tr}(|b\rangle\langle b|U\rho U^\dagger)\cdot U^\dagger|\hat{b}\rangle\langle\hat{b}|U.$$

Plugging this in gives

$$\mathbb{E}[\hat{\rho}] = \mathcal{M}^{-1}\left(\mathbb{E}_{U\sim\mathcal{U}}\left[\sum_{b\in\{0,1\}^n} \mathrm{tr}(|b\rangle\langle b|U\rho U^\dagger)\cdot U^\dagger|b\rangle\langle b|U\right]\right) = \mathcal{M}^{-1}(\mathcal{M}(\rho)) = \rho,$$

which concludes the proof. □

Because of the unbiased property, we can think of $\hat{\rho}$ as a classical surrogate of the unknown quantum state ρ. To predict an expectation value $\mathrm{tr}(O\rho)$, we use the single-shot estimator $\hat{o} = \mathrm{tr}(O\hat{\rho})$. By linearity of expectation value, we have

$$\mathbb{E}[\hat{o}] = \mathrm{tr}(O\mathbb{E}[\hat{\rho}]) = \mathrm{tr}(O\rho).$$

However, there will be statistical fluctuation in $\hat{o}$ that causes any single-shot estimator $\hat{o}$ to deviate away from the expectation value $\mathrm{tr}(O\rho)$. To understand the statistical fluctuation, we need to look at the variance of $\hat{o}$:

$$\mathrm{Var}[\hat{o}] = \mathbb{E}[\hat{o}^2] - \mathbb{E}[\hat{o}]^2.$$

If the variance is large, the sample complexity N required to achieve a given precision must be larger, and vice versa. To bound the variance of $\hat{o}$, we need to first establish the basic properties of the measurement channel $\mathcal{M}$.

Lemma 96 (Properties of the Measurement Channel)**.** *The channel $\mathcal{M}$ and its inverse $\mathcal{M}^{-1}$ have the following properties:*

(i) $\mathcal{M}$ *is* ***trace-preserving****, i.e.,* $\operatorname{tr}(\mathcal{M}(X)) = \operatorname{tr}(X)$ *for any operator* X*.*

(ii) $\mathcal{M}$ *and* $\mathcal{M}^{-1}$ *are* ***self-adjoint*** *with respect to the Hilbert-Schmidt inner product,* $\langle A, B\rangle_{HS} = \operatorname{tr}(A^\dagger B)$*. This means that for any operators* A *and* B*, the channel can be moved from one side to the other:*

$$\langle A, \mathcal{M}(B)\rangle_{HS} = \langle \mathcal{M}(A), B\rangle_{HS}.$$

(iii) $\mathcal{M}$ *and* $\mathcal{M}^{-1}$ *are* ***unital****, i.e.,* $\mathcal{M}(\mathrm{Id}) = \mathcal{M}^{-1}(\mathrm{Id}) = \mathrm{Id}$*.*

(iv) The classical snapshot $\hat{\rho}$ *has unit trace,* $\operatorname{tr}(\hat{\rho}) = 1$*.*

Proof. (i) We take the trace of $\mathcal{M}(X)$ and use the linearity and cyclic property of the trace:

$$\begin{aligned}
\operatorname{tr}(\mathcal{M}(X)) &= \operatorname{tr}\left(\mathbb{E}_U\left[\sum_b \operatorname{tr}(|b\rangle\langle b|UXU^\dagger)U^\dagger|b\rangle\langle b|U\right]\right) \\
&= \mathbb{E}_U\left[\sum_b \operatorname{tr}(|b\rangle\langle b|UXU^\dagger)\operatorname{tr}(U^\dagger|b\rangle\langle b|U)\right] \\
&= \mathbb{E}_U\left[\operatorname{tr}\left(\left(\sum_b |b\rangle\langle b|\right)UXU^\dagger\right)\right] = \mathbb{E}_U[\operatorname{tr}(\mathrm{Id}\cdot UXU^\dagger)] = \operatorname{tr}(X).
\end{aligned}$$

This completes the proof of (i).

(ii) We verify the self-adjoint condition for Hermitian operators A, B, where the inner product is $\operatorname{tr}(AB) = \operatorname{tr}(A^\dagger B)$.

$$\begin{aligned}
\operatorname{tr}(A\mathcal{M}(B)) &= \operatorname{tr}\left(A\cdot\mathbb{E}_U\sum_b \operatorname{tr}(|b\rangle\langle b|UBU^\dagger)U^\dagger|b\rangle\langle b|U\right) \\
&= \mathbb{E}_U\sum_b \operatorname{tr}(|b\rangle\langle b|UBU^\dagger)\operatorname{tr}(AU^\dagger|b\rangle\langle b|U) \\
&= \mathbb{E}_U\sum_b \langle b|UBU^\dagger|b\rangle\langle b|UAU^\dagger|b\rangle.
\end{aligned}$$

This final expression is symmetric in A and B, so $\operatorname{tr}(A\mathcal{M}(B)) = \operatorname{tr}(B\mathcal{M}(A)) = \operatorname{tr}(\mathcal{M}(A)B)$, proving $\mathcal{M}$ is self-adjoint. To show $\mathcal{M}^{-1}$ is self-adjoint, let $X = \mathcal{M}^{-1}(A)$ and $Y = \mathcal{M}^{-1}(B)$. We must show $\operatorname{tr}(A\mathcal{M}^{-1}(B)) = \operatorname{tr}(\mathcal{M}^{-1}(A)B)$, which is equivalent to showing $\operatorname{tr}(\mathcal{M}(X)Y) = \operatorname{tr}(X\mathcal{M}(Y))$. This is true because we have shown that $\mathcal{M}$ is self-adjoint.

(iii) This proof proceeds in three steps. First, we show that the adjoint of a trace-preserving (TP) map is unital. Second, we apply this to $\mathcal{M}$. Third, we extend the property to $\mathcal{M}^{-1}$. Consider any TP map Φ. The adjoint $\Phi^\dagger$ is defined by $\operatorname{tr}(A^\dagger\Phi(B)) = \operatorname{tr}((\Phi^\dagger(A))^\dagger B)$ for all A, B. To show $\Phi^\dagger$ is unital, we must show $\Phi^\dagger(\mathrm{Id}) = \mathrm{Id}$. We can prove this by showing that for any arbitrary matrix X, $\operatorname{tr}(X^\dagger\Phi^\dagger(\mathrm{Id})) = \operatorname{tr}(X^\dagger\mathrm{Id})$. Let $A = \mathrm{Id}$ and $B = X$ in the adjoint definition:

$$\operatorname{tr}((\Phi^\dagger(\mathrm{Id}))^\dagger X) = \operatorname{tr}(\mathrm{Id}^\dagger\Phi(X)) = \operatorname{tr}(\Phi(X)).$$

Since Φ is trace-preserving, $\mathrm{tr}(\Phi(X)) = \mathrm{tr}(X)$. Thus we have:

$$\mathrm{tr}((\Phi^\dagger(\mathrm{Id}))^\dagger X) = \mathrm{tr}(X) = \mathrm{tr}(\mathrm{Id} \cdot X).$$

Since this equality holds for all X, it implies $(\Phi^\dagger(\mathrm{Id}))^\dagger = \mathrm{Id}$, and therefore $\Phi^\dagger(\mathrm{Id}) = \mathrm{Id}$. From (i), we know $\mathcal{M}$ is trace-preserving. Therefore its adjoint, $\mathcal{M}^\dagger$, must be unital. From (ii), we know $\mathcal{M}$ is self-adjoint, so $\mathcal{M} = \mathcal{M}^\dagger$. Combining these, $\mathcal{M}$ itself must be unital, so $\mathcal{M}(\mathrm{Id}) = \mathrm{Id}$. Applying the map $\mathcal{M}^{-1}$ to both sides gives:

$$\mathcal{M}^{-1}(\mathcal{M}(\mathrm{Id})) = \mathcal{M}^{-1}(\mathrm{Id}) \implies \mathrm{Id} = \mathcal{M}^{-1}(\mathrm{Id}).$$

Thus, $\mathcal{M}^{-1}$ is unital.

(iv) The trace of the snapshot is:

$$\begin{aligned}
\mathrm{tr}(\hat{\rho}) &= \mathrm{tr}(\mathcal{M}^{-1}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U)) \\
&= \langle \mathrm{Id}, \mathcal{M}^{-1}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U)\rangle_{HS} && \text{(using } \mathrm{tr}(X) = \langle \mathrm{Id}, X\rangle_{HS}\text{)} \\
&= \langle \mathcal{M}^{-1}(\mathrm{Id}), U^\dagger|\hat{b}\rangle\langle\hat{b}|U\rangle_{HS} && \text{(by self-adjointness of } \mathcal{M}^{-1}\text{)} \\
&= \langle \mathrm{Id}, U^\dagger|\hat{b}\rangle\langle\hat{b}|U\rangle_{HS} && \text{(by unitality of } \mathcal{M}^{-1}\text{)} \\
&= \mathrm{tr}(\mathrm{Id} \cdot U^\dagger|\hat{b}\rangle\langle\hat{b}|U) = \mathrm{tr}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U) = 1.
\end{aligned}$$

This completes the proof of (iv). □

With these basic properties of the measurement channel $\mathcal{M}$ and the classical snapshot $\hat{\rho}$, we can establish the variance of the unbiased estimator $\hat{o} = \mathrm{tr}(O\hat{\rho})$.

Lemma 97 (Variance of the Single-Shot Estimator)**.** *Let O be an observable, $\hat{\rho}$ be a classical snapshot of ρ, and $\hat{o} = \mathrm{tr}(O\hat{\rho})$. The variance of the single-shot estimator $\hat{o}$ is bounded above as follows,*

$$\mathrm{Var}[\hat{o}] \leq \|O - \tfrac{\mathrm{tr}(O)}{d}\mathrm{Id}\|^2_{\mathrm{shadow}},$$

where the ***shadow norm*** *is defined by the measurement procedure:*

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

Proof. Let $O_0 = O - \frac{\mathrm{tr}(O)}{d}\mathrm{Id}$. We have

$$\begin{aligned}
\mathrm{Var}[\hat{o}] &= \mathbb{E}[\mathrm{tr}(O\hat{\rho})^2] - \mathbb{E}[\mathrm{tr}(O\hat{\rho})]^2 = \mathbb{E}[(\mathrm{tr}(O\hat{\rho}) - \mathbb{E}[\mathrm{tr}(O\hat{\rho})])^2] \\
&= \mathbb{E}\left[\left(\mathrm{tr}(O_0\hat{\rho}) + \frac{\mathrm{tr}(O)}{d}\mathrm{tr}(\hat{\rho}) - \mathbb{E}[\mathrm{tr}(O_0\hat{\rho})] - \frac{\mathrm{tr}(O)}{d}\mathbb{E}[\mathrm{tr}(\hat{\rho})]\right)^2\right].
\end{aligned}$$

By Lemma 96, $\mathrm{tr}(\hat{\rho}) = 1$, so the variance can be simplified to

$$\mathrm{Var}[\hat{o}] = \mathbb{E}\left[(\mathrm{tr}(O_0\hat{\rho}) - \mathbb{E}[\mathrm{tr}(O_0\hat{\rho})])^2\right] = \mathbb{E}\left[\mathrm{tr}(O_0\hat{\rho})^2\right] - \mathbb{E}[\mathrm{tr}(O_0\hat{\rho})]^2.$$

Algorithm 3: ShadowDataCollection(N, ρ, $\mathcal{U}$)

Input: Access to N copies of state ρ, random unitary ensemble $\mathcal{U}$
Output: A classical shadow $S = \{(U_t, \hat{b}_t)\}_{t=1}^{N_{tot}}$

```
Initialize an empty list S;
for t = 1, ..., N do
    Take a fresh copy of ρ;
    Sample a unitary U_t ~ U;
    Measure the state U_t ρ U_t^† in the computational basis to get outcome b̂_t;
    Append the pair (U_t, b̂_t) to S;
end
return S
```

The first term is

$$\begin{aligned}
\mathbb{E}[\mathrm{tr}(O_0\hat{\rho})^2] &= \mathbb{E}\left[\left(\mathrm{tr}\left(O_0\mathcal{M}^{-1}\left(U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right)\right)\right)^2\right] && \text{(by definition)} \\
&= \mathbb{E}\left[\left(\mathrm{tr}\left(\mathcal{M}^{-1}(O_0)U^\dagger|\hat{b}\rangle\langle\hat{b}|U\right)\right)^2\right] && \text{(by self-adjointness)} \\
&= \mathbb{E}_{U,\hat{b}}\left[\langle\hat{b}|U\mathcal{M}^{-1}(O_0)U^\dagger|\hat{b}\rangle^2\right] \\
&= \mathbb{E}_U\left[\sum_b \mathrm{tr}(|b\rangle\langle b|U\rho U^\dagger)\langle b|U\mathcal{M}^{-1}(O_0)U^\dagger|b\rangle^2\right].
\end{aligned}$$

Plugging this into the variance expression and dropping the second non-positive term $-(\mathrm{tr}(O_0\rho))^2$ gives the stated variance bound. □

2.2. Algorithm

The full algorithm for the classical shadow formalism involves (1) collecting a number of snapshots N and then (2) processing them classically to produce estimates for the expectation values of all M observables. The data collection phase of classical shadow is given in Algorithm 3 and the prediction phase of classical shadow is given in Algorithm 4.

2.3. Performance Guarantee

To analyze the performance of the classical shadow formalism, we begin with an analysis of the estimators used to convert many single-shot predictions into a final, high-confidence estimate. A powerful statistical tool for this is the median-of-means estimator, which enables us to obtain exponentially decaying failure probabilities from any random variable that has a bounded variance.

It is important to note that the standard mean estimator does not provide such strong guarantees. While simple to implement, its sample complexity scales poorly with the desired success probability.

Lemma 98 (Performance of the Standard Mean Estimator)**.** *Let X be a random variable with mean μ and finite variance σ^2. Let $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^N X_i$ be the empirical*

Algorithm 4: ShadowPrediction($\{O_i\}_{i=1}^M$, S, K)

Input: Observables $\{O_i\}_{i=1}^M$, a classical shadow $S = \{\hat{\rho}_i\}_{i=1}^N$ of size N organized into K groups of size N/K
Output: Estimates $\{\hat{o}_i\}_{i=1}^M$

Let $N' = N/K$;
for $i = 1, \ldots, M$ **do**
  Initialize an empty list of means Means_i;
  for $k = 1, \ldots, K$ **do**
    Let S_k be the k-th group of S consisting of N/K snapshots;
    Compute empirical mean
      $\hat{o}_i^{(k)} = \frac{1}{N} \sum_{(U_t, \hat{b}_t) \in S_k} \text{tr}(O_i \cdot \mathcal{M}^{-1}(U_t^\dagger |\hat{b}_t\rangle\langle\hat{b}_t| U_t))$;
    Append $\hat{o}_i^{(k)}$ to Means_i;
  end
  Set $\hat{o}_i = \text{median}(\text{Means}_i)$;
end
return $\{\hat{o}_i\}_{i=1}^M$

mean of N independent samples. To guarantee that $|\hat{\mu}_N - \mu| \leq \epsilon$ with a failure probability of at most δ, the number of samples required scales as:

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

Proof. The proof follows directly from Chebyshev's inequality. The variance of the empirical mean is $\text{Var}[\hat{\mu}_N] = \sigma^2/N$. Chebyshev's inequality states that for any random variable Y with finite variance, $\Pr[|Y - \mathbb{E}[Y]| \geq \epsilon] \leq \text{Var}[Y]/\epsilon^2$. Applying this to our empirical mean $\hat{\mu}_N$:

$$\Pr[|\hat{\mu}_N - \mu| \geq \epsilon] \leq \frac{\text{Var}[\hat{\mu}_N]}{\epsilon^2} = \frac{\sigma^2}{N\epsilon^2}.$$

To ensure this failure probability is at most δ, we require:

$$\frac{\sigma^2}{N\epsilon^2} \leq \delta \implies N \geq \frac{\sigma^2}{\epsilon^2 \delta}.$$

This completes the proof. The crucial point is the sample complexity's $1/\delta$ dependence, which is unfavorable for high-confidence predictions (i.e., small δ). □

The median-of-means estimator circumvents this issue and achieves a much better logarithmic dependence on $1/\delta$.

Lemma 99 (Performance of Median-of-Means)**.** *Let X be a random variable with mean μ and variance σ^2. Let $\{\hat{\mu}_k\}_{k=1}^K$ be K independent empirical means, each constructed from N' independent samples of X. If $N' \geq 4\sigma^2/\epsilon^2$, then*

$$\Pr[|\text{median}\{\hat{\mu}_k\} - \mu| \geq \epsilon] \leq 2\exp(-K/8).$$

Proof. The variance of any of the K empirical means is $\text{Var}[\hat{\mu}_k] = \sigma^2/N'$. By Chebyshev's inequality, the probability that a single empirical mean differs from the true expectation value by more than ϵ is

$$p = \Pr[|\hat{\mu}_k - \mu| > \epsilon] \leq \frac{\text{Var}[\hat{\mu}_k]}{\epsilon^2} = \frac{\sigma^2}{N'\epsilon^2}.$$

By choosing $N' \geq 4\sigma^2/\epsilon^2$, we ensure $p \leq 1/4$. The median estimate fails only if at least $K/2$ of the means are incorrect. Let Z_k be an indicator for the k-th mean being incorrect. The Z_k are i.i.d. Bernoulli variables with parameter $p \leq 1/4$. By a Hoeffding bound for the sum of Bernoulli variables,

$$\Pr\left[\sum_{k=1}^{K} Z_k \geq K/2\right] = \Pr\left[\frac{1}{K}\sum Z_k - p \geq \frac{1}{2} - p\right] \leq \exp(-2K(1/2-p)^2).$$

Since $p \leq 1/4$, we have $(1/2 - p) \geq 1/4$. Therefore, the failure probability is bounded by $\exp(-2K(1/4)^2) = \exp(-K/8)$. □

With the concentration inequalities provided above, we can obtain the following performance guarantee for classical shadow formalism.

Theorem 100 (Performance of Classical Shadow Formalism)**.** *Fix a random unitary ensemble $\mathcal{U}$, a set of M observables $\{O_i\}$, and accuracy parameters $\epsilon, \delta \in (0,1)$. Let $B = \max_i \|O_i - \frac{\mathrm{tr}(O_i)}{d}\mathrm{Id}\|^2_{\mathrm{shadow}}$. Using a total of*

$$N = \mathcal{O}\left(\frac{B}{\epsilon^2}\log\left(\frac{M}{\delta}\right)\right)$$

measurements, the median-of-means procedure with $K = \mathcal{O}(\log(M/\delta))$ outputs estimates $\{\hat{o}_i\}$ such that with probability at least $1-\delta$,

$$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \leq \epsilon \quad \textit{for all } i = 1, \ldots, M.$$

Proof. We combine the concentration inequality for the median-of-means estimator from Lemma 99 and the variance bound of the single-shot estimator $\mathrm{tr}(O_i\hat{\rho}_t)$ from Lemma 97. For each observable O_i with $i = 1, \ldots, M$, the variance of the single-shot estimator $\mathrm{tr}(O_i\hat{\rho}_t)$ is bounded as follows,

$$\mathrm{Var}[\mathrm{tr}(O_i\hat{\rho}_t)] \leq \left\|O_i - \frac{\mathrm{tr}(O_i)}{d}\mathrm{Id}\right\|^2_{\mathrm{shadow}} \leq B.$$

For each observable O_i, we set the number of snapshots per empirical mean to be $N' = \lceil 4B/\epsilon^2 \rceil$. To ensure the total failure probability over all M observables is at most δ, we use a union bound. We require the failure probability for each observable to be at most δ/M. From the lemma, we need to choose K such that $2e^{-K/8} \leq \delta/M$, which gives $K = \lceil 8\log(2M/\delta) \rceil = \mathcal{O}(\log(M/\delta))$. The total number of samples is $N = N' \cdot K = \mathcal{O}\left(\frac{B}{\epsilon^2}\log\left(\frac{M}{\delta}\right)\right)$. □

We can compare with the direct measurement approach to see that the dependence on M is now improved from $M \log M$ to just $\log M$. However, the classical shadow formalism introduces an important dependence on the shadow norm. In the next section, we will look at how these shadow norm scales with the choice of the random unitary ensemble and the family of observables.

3. Instantiations of the Random Unitary Ensemble

The abstract sample complexity derived in the previous section becomes concrete once we specify the ensemble of random unitaries $\mathcal{U}$ and compute the corresponding shadow norm. In this section, we analyze two of the most important ensembles: random global Clifford circuits and random local Pauli measurements. The proofs for their properties rely on unitary designs.

3.1. A Useful Tool: Averaging over Unitary Group

The key feature of the Clifford group and many other ensembles is that they reproduce the statistical properties of the full unitary group (endowed with the Haar measure). This is formalized by the concept of a unitary t-design.

Definition 101 (Unitary t-design)**.** *An ensemble of unitaries $\mathcal{U}$ is a **unitary t-design** if, for any polynomial P with degree at most t in the matrix entries of U and t in the entries of $U^\dagger$, the average over the ensemble is equal to the average over the full unitary group with its unique uniform (Haar) measure:*

$$\mathbb{E}_{U\sim\mathcal{U}}[P(U,U^\dagger)] = \int_{U(d)} P(U,U^\dagger)dU.$$

Equivalently, the ensemble must reproduce the first t moments of the Haar measure, which means that for all operators X:

$$\mathbb{E}_{U\sim\mathcal{U}}[U^{\otimes t}X(U^\dagger)^{\otimes t}] = \int_{U(d)} U^{\otimes t}X(U^\dagger)^{\otimes t}dU.$$

Lemma 102. *Unitary t-design is unitary t'-design for any $t' < t$.*

Proof. Let $P(U,U^\dagger)$ be a polynomial of degree t' in the entries of U and $U^\dagger$, where $t' < t$. Then P is also a polynomial of degree at most t. Since the ensemble $\mathcal{U}$ is a t-design, the defining equality $\mathbb{E}_{U\sim\mathcal{U}}[P(U,U^\dagger)] = \int_{U(d)} P(U,U^\dagger)dU$ holds. By definition, this means $\mathcal{U}$ is also a t'-design. □

The power of a t-design is that we can compute averages over its elements using known formulas for Haar integrals. A general method for this is the **Weingarten calculus**. While the full calculus is beyond the scope of this lecture, we can use some of its key results about moments of random vectors. If U is a Haar-random unitary, then for a fixed vector $|b\rangle$, the vector $|\psi\rangle = U|b\rangle$ is a random pure state uniformly distributed on the unit sphere.

Fact 103 (Moments of Haar-Random Pure States)**.** *Let $|\psi\rangle$ be a random pure state in $\mathbb{C}^d$ distributed uniformly according to the Haar measure. The first three moments of the operator $|\psi\rangle\langle\psi|$ are given by:*

$$\mathbb{E}[|\psi\rangle\langle\psi|] = \frac{\mathrm{Id}}{d}$$

$$\mathbb{E}[(|\psi\rangle\langle\psi|)^{\otimes 2}] = \frac{\mathrm{Id}\otimes\mathrm{Id} + S_2}{d(d+1)} \tag{25}$$

$$\mathbb{E}[(|\psi\rangle\langle\psi|)^{\otimes 3}] = \frac{1}{d(d+1)(d+2)}\sum_{\pi\in S_3} P_\pi \tag{26}$$

where S_3 is the symmetric group on 3 elements, S_2 is the SWAP operator on $(\mathbb{C}^d)^{\otimes 2}$, and P_π is the permutation operator on $(\mathbb{C}^d)^{\otimes 3}$ corresponding to $\pi \in S_3$.

We can now use these tools to derive the specific formulas needed to analyze the measurement channels and shadow norm.

Lemma 104 (Derivation of Key Integral Formulas)**.** *Let the average be over an ensemble $\mathcal{U}$ that forms a unitary 3-design (e.g., the Clifford group [**Web15**]). For a fixed vector $|b\rangle$, the following identity holds from the 2-design property:*

$$\mathbb{E}_{U\in\mathcal{U}}[\langle b|UAU^\dagger|b\rangle U^\dagger|b\rangle\langle b|U] = \frac{\mathrm{tr}(A)\mathrm{Id} + A}{d(d+1)}. \tag{27}$$

From the 3-design property, we also have:

$$\mathbb{E}_{U\in\mathcal{U}}[\langle b|UA_0U^\dagger|b\rangle\langle b|UB_0U^\dagger|b\rangle U^\dagger|b\rangle\langle b|U] = \frac{\mathrm{tr}(A_0B_0)\mathrm{Id} + A_0B_0 + B_0A_0}{d(d+1)(d+2)} \tag{28}$$

for any traceless operators A_0, B_0.

Proof. Let $|\psi\rangle = U|b\rangle$ be a Haar-random pure state. The expectation over $U \in \mathcal{U}$ is equivalent to the expectation over $|\psi\rangle$.

Proof of Eq. (27): Let $\Phi(A) = \mathbb{E}_{|\psi\rangle}[\langle\psi|A|\psi\rangle|\psi\rangle\langle\psi|]$. To identify the operator $\Phi(A)$, we can test it against an arbitrary operator C by taking the trace:

$$\begin{aligned}\mathrm{tr}(\Phi(A)C) &= \mathrm{tr}\left(\mathbb{E}_{|\psi\rangle}[\langle\psi|A|\psi\rangle|\psi\rangle\langle\psi|]C\right)\\ &= \mathbb{E}_{|\psi\rangle}[\langle\psi|A|\psi\rangle\langle\psi|C|\psi\rangle] \quad \text{(by linearity of trace and expectation)}\\ &= \mathbb{E}_{|\psi\rangle}[\mathrm{tr}(A|\psi\rangle\langle\psi|)\mathrm{tr}(C|\psi\rangle\langle\psi|)].\end{aligned}$$

We can express this as a trace over a larger Hilbert space $(\mathbb{C}^d)^{\otimes 2}$:

$$\begin{aligned}\mathrm{tr}(\Phi(A)C) &= \mathbb{E}_{|\psi\rangle}\left[\mathrm{tr}_{1,2}\left((A\otimes C)(|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|)\right)\right]\\ &= \mathrm{tr}_{1,2}\left((A\otimes C)\mathbb{E}[(|\psi\rangle\langle\psi|)^{\otimes 2}]\right).\end{aligned}$$

Now, we substitute the second moment formula from Eq. (25):

$$\begin{aligned}\mathrm{tr}(\Phi(A)C) &= \mathrm{tr}_{1,2}\left((A\otimes C)\frac{\mathrm{Id}\otimes\mathrm{Id} + S_2}{d(d+1)}\right)\\ &= \frac{1}{d(d+1)}\left(\mathrm{tr}(A\cdot\mathrm{Id})\mathrm{tr}(C\cdot\mathrm{Id}) + \mathrm{tr}((A\otimes C)\cdot S_2)\right).\end{aligned}$$

Using the identity $\mathrm{tr}((X\otimes Y)S_2) = \mathrm{tr}(XY)$, we get:

$$\mathrm{tr}(\Phi(A)C) = \frac{1}{d(d+1)}(\mathrm{tr}(A)\mathrm{tr}(C) + \mathrm{tr}(AC)).$$

This holds for all C. We can see that this is satisfied by $\Phi(A) = \frac{\mathrm{tr}(A)\mathrm{Id}+A}{d(d+1)}$, since

$$\mathrm{tr}\left(\left(\frac{\mathrm{tr}(A)\mathrm{Id}+A}{d(d+1)}\right)C\right) = \frac{1}{d(d+1)}(\mathrm{tr}(A)\mathrm{tr}(C) + \mathrm{tr}(AC)).$$

This concludes the proof of the first identity.

Proof of Eq. (28): Let $\Psi(A_0, B_0) = \mathbb{E}_{|\psi\rangle}[\langle\psi|A_0|\psi\rangle\langle\psi|B_0|\psi\rangle|\psi\rangle\langle\psi|]$. Again, we test it against an arbitrary operator C:

$$\begin{aligned}\mathrm{tr}(\Psi(A_0,B_0)C) &= \mathbb{E}_{|\psi\rangle}[\langle\psi|A_0|\psi\rangle\langle\psi|B_0|\psi\rangle\langle\psi|C|\psi\rangle]\\ &= \mathbb{E}_{|\psi\rangle}[\mathrm{tr}(A_0|\psi\rangle\langle\psi|)\mathrm{tr}(B_0|\psi\rangle\langle\psi|)\mathrm{tr}(C|\psi\rangle\langle\psi|)].\end{aligned}$$

Using the same trace trick on $(\mathbb{C}^d)^{\otimes 3}$:

$$\mathrm{tr}(\Psi(A_0,B_0)C) = \mathrm{tr}_{1,2,3}\left((A_0\otimes B_0\otimes C)\mathbb{E}[(|\psi\rangle\langle\psi|)^{\otimes 3}]\right).$$

Now we use the third moment formula from Eq. (26). A key identity for evaluating the trace with a permutation operator is $\mathrm{tr}_{1,\dots,t}((O_1\otimes\cdots\otimes O_t)P_\pi) = \mathrm{tr}(O_1O_{\pi(1)}\dots)$, where the trace is taken over the product of operators according to the cycle decomposition of π. For S_3, we have:

- 1 identity permutation id:

$$\mathrm{tr}(A_0)\mathrm{tr}(B_0)\mathrm{tr}(C).$$

- 3 transpositions $(12), (13), (23)$:
$$\mathrm{tr}(A_0B_0)\mathrm{tr}(C), \mathrm{tr}(A_0C)\mathrm{tr}(B_0), \mathrm{tr}(B_0C)\mathrm{tr}(A_0).$$
- 2 three-cycles $(123), (132)$:
$$\mathrm{tr}(A_0B_0C), \mathrm{tr}(A_0CB_0).$$

Summing these terms and dividing by the prefactor $d(d+1)(d+2)$ gives the full expression for $\mathrm{tr}(\Psi(A_0, B_0)C)$. Since A_0 and B_0 are traceless, all terms containing $\mathrm{tr}(A_0)$ or $\mathrm{tr}(B_0)$ vanish. We are left with:
$$\mathrm{tr}(\Psi(A_0, B_0)C) = \frac{\mathrm{tr}(A_0B_0)\mathrm{tr}(C) + \mathrm{tr}(A_0B_0C) + \mathrm{tr}(A_0CB_0)}{d(d+1)(d+2)}.$$
Using the cyclic property of the trace, $\mathrm{tr}(A_0CB_0) = \mathrm{tr}(B_0A_0C)$. This must hold for all C. We check this against the trace of the right-hand side of Eq. (28):
$$\begin{aligned}&\mathrm{tr}\left(\left(\frac{\mathrm{tr}(A_0B_0)\mathrm{Id} + A_0B_0 + B_0A_0}{d(d+1)(d+2)}\right)C\right)\\ &= \frac{\mathrm{tr}(A_0B_0)\mathrm{tr}(C) + \mathrm{tr}(A_0B_0C) + \mathrm{tr}(B_0A_0C)}{d(d+1)(d+2)}.\end{aligned}$$
The expressions match hence completes the proof. □

3.2. Random Clifford Measurements

The first ensemble we consider is the group of n-qubit Clifford circuits. We will define precisely what these are in a later lecture, but for now all that we need is that they comprise a subgroup of the unitary group which forms a unitary 3-design [**Web15**, **Zhu17**], so we can use the formulas from Lemma 104. While experimentally demanding for large systems, this ensemble has powerful theoretical properties.

Lemma 105 (Measurement Channel for Clifford Ensemble)**.** *For the ensemble of global n-qubit Clifford unitaries, $\mathcal{U} = \mathrm{Cl}(2^n)$, where $d = 2^n$: The measurement channel $\mathcal{M}$ and its inverse $\mathcal{M}^{-1}$ are given by*
$$\begin{aligned}\mathcal{M}(X) &= \frac{X + \mathrm{tr}(X)\mathrm{Id}}{d+1},\\ \mathcal{M}^{-1}(X) &= (d+1)X - \mathrm{tr}(X)\mathrm{Id}.\end{aligned}$$

Proof. We compute the channel $\mathcal{M}$ by applying the result from Eq. (27). For a state ρ with $\mathrm{tr}(\rho) = 1$:
$$\begin{aligned}\mathcal{M}(\rho) &= \mathbb{E}_{U\in\mathrm{Cl}(d)}\left[\sum_{b\in\{0,1\}^n} \mathrm{tr}(|b\rangle\langle b|U\rho U^\dagger)\cdot U^\dagger|b\rangle\langle b|U\right]\\ &= \sum_{b\in\{0,1\}^n} \mathbb{E}_{U\in\mathrm{Cl}(d)}\left[\langle b|U\rho U^\dagger|b\rangle\cdot U^\dagger|b\rangle\langle b|U\right].\end{aligned}$$
Since the expression inside the expectation is the same for any basis state $|b\rangle$ due to the average over the unitary group, we can evaluate it for a single $|b\rangle$ and multiply by d. Using Eq. (27):
$$\mathcal{M}(\rho) = d\cdot\left(\frac{\mathrm{tr}(\rho)\mathrm{Id} + \rho}{d(d+1)}\right) = \frac{\mathrm{Id} + \rho}{d+1}.$$

By linearity of the channel, for any operator X, $\mathcal{M}(X) = \frac{X+\mathrm{tr}(X)\mathrm{Id}}{d+1}$. To find the inverse, we set $Y = \mathcal{M}(X)$ and solve for X:

$$Y = \frac{X + \mathrm{tr}(X)\mathrm{Id}}{d+1} \implies (d+1)Y = X + \mathrm{tr}(X)\mathrm{Id}.$$

Taking the trace of both sides gives $(d+1)\mathrm{tr}(Y) = \mathrm{tr}(X) + \mathrm{tr}(X)\mathrm{tr}(\mathrm{Id}) = \mathrm{tr}(X)(1+d)$. Thus, $\mathrm{tr}(X) = \mathrm{tr}(Y)$. Substituting this back gives:

$$(d+1)Y = X + \mathrm{tr}(Y)\mathrm{Id} \implies X = (d+1)Y - \mathrm{tr}(Y)\mathrm{Id}.$$

So, $\mathcal{M}^{-1}(Y) = (d+1)Y - \mathrm{tr}(Y)\mathrm{Id}$. □

Proposition 106 (Clifford Shadows)**.** *For the random Clifford ensemble:*

(i) The classical snapshot is $\hat{\rho} = (d+1)U^\dagger|\hat{b}\rangle\langle\hat{b}|U - \mathrm{Id}$.

(ii) The shadow norm is bounded by $\|O_0\|^2_{\mathrm{shadow}} \leq 3\mathrm{tr}(O_0^2)$ *for any traceless operator* O_0.

Proof. (i) The snapshot is $\hat{\rho} = \mathcal{M}^{-1}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U)$. Since $\mathrm{tr}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U) = 1$, applying the inverse channel formula gives

$$\hat{\rho} = (d+1)U^\dagger|\hat{b}\rangle\langle\hat{b}|U - \mathrm{tr}(U^\dagger|\hat{b}\rangle\langle\hat{b}|U)\mathrm{Id} = (d+1)U^\dagger|\hat{b}\rangle\langle\hat{b}|U - \mathrm{Id}.$$

(ii) For a traceless operator O_0, the inverse map is simply $\mathcal{M}^{-1}(O_0) = (d+1)O_0$. We now compute the shadow norm:

$$\begin{aligned}\|O_0\|^2_{\mathrm{shadow}} &= \max_\sigma \mathbb{E}_U \sum_b \langle b|U\sigma U^\dagger|b\rangle \left(\langle b|U(d+1)O_0U^\dagger|b\rangle\right)^2 \\ &= (d+1)^2 \max_\sigma \mathrm{tr}\left(\sigma \sum_b \mathbb{E}_U[U^\dagger|b\rangle\langle b|U\langle b|UO_0U^\dagger|b\rangle^2]\right).\end{aligned}$$

Using the 3-design formula from Eq. (28) with $A_0 = B_0 = O_0$:

$$\sum_b \mathbb{E}_U[\dots] = \sum_b \frac{\mathrm{tr}(O_0^2)\mathrm{Id} + O_0^2 + O_0^2}{d(d+1)(d+2)} = d\frac{\mathrm{tr}(O_0^2)\mathrm{Id} + 2O_0^2}{d(d+1)(d+2)} = \frac{\mathrm{tr}(O_0^2)\mathrm{Id} + 2O_0^2}{(d+1)(d+2)}.$$

Plugging this back into the norm expression:

$$\begin{aligned}\|O_0\|^2_{\mathrm{shadow}} &= (d+1)^2 \max_\sigma \mathrm{tr}\left(\sigma \frac{\mathrm{tr}(O_0^2)\mathrm{Id} + 2O_0^2}{(d+1)(d+2)}\right) \\ &= \frac{d+1}{d+2} \max_\sigma \left(\mathrm{tr}(O_0^2)\mathrm{tr}(\sigma) + 2\mathrm{tr}(\sigma O_0^2)\right) \\ &= \frac{d+1}{d+2}\left(\mathrm{tr}(O_0^2) + 2\max_\sigma \mathrm{tr}(\sigma O_0^2)\right).\end{aligned}$$

Since $\max_\sigma \mathrm{tr}(\sigma O_0^2)$ is the largest eigenvalue of the Hermitian operator O_0^2, denoted $\|O_0^2\|_\infty$, and since $\|O_0^2\|_\infty = \|O_0\|^2_\infty \leq \sum_i \lambda_i(O_0)^2 = \mathrm{tr}(O_0^2)$, we have

$$\|O_0\|^2_{\mathrm{shadow}} \leq \frac{d+1}{d+2}(\mathrm{tr}(O_0^2) + 2\mathrm{tr}(O_0^2)) = 3\mathrm{tr}(O_0^2)\frac{d+1}{d+2} < 3\mathrm{tr}(O_0^2),$$

which establishes the shadow norm bound. □

The key strength of Clifford shadows is the dependence on $\mathrm{tr}(O^2)$. For example, to estimate the fidelity with an n-qubit pure state $|\psi\rangle$, $F = \mathrm{tr}(|\psi\rangle\langle\psi|\rho)$, we use the observable $O = |\psi\rangle\langle\psi|$. Here, $\mathrm{tr}(O^2) = 1$. Hence the required number of measurements to estimate fidelities with any M pure states $|\psi_1\rangle, \dots, |\psi_M\rangle$ is only $N = \mathcal{O}(\log(M/\delta)/\epsilon^2)$. This is independent of the system size n.

3.3. Random Pauli Measurements

The second ensemble we consider is when the random unitary corresponds to a tensor product of n single-qubit Clifford unitary. This ensemble is highly practical, as it only involves single-qubit rotations. Furthermore, the measurement protocol obtained from this ensemble is equivalent to measuring each qubit in a random Pauli basis (X, Y, or Z).

Lemma 107 (Measurement Channel for Pauli Ensemble)**.** *For the ensemble of local random unitaries, $\mathcal{U} = \mathrm{Cl}(2)^{\otimes n}$: The measurement channel $\mathcal{M}$ and its inverse $\mathcal{M}^{-1}$ factorize into single-qubit channels:*

$$\mathcal{M}(X) = \mathcal{M}_1(X_1) \otimes \cdots \otimes \mathcal{M}_1(X_n),$$
$$\mathcal{M}^{-1}(X) = \mathcal{M}_1^{-1}(X_1) \otimes \cdots \otimes \mathcal{M}_1^{-1}(X_n),$$

where $\mathcal{M}_1(Y) = (\mathrm{tr}(Y)\mathrm{Id} + Y)/3$ is the single-qubit depolarizing channel, and $\mathcal{M}_1^{-1}(Y) = 3Y - \mathrm{tr}(Y)\mathrm{Id}$.

Proof. The ensemble is a product distribution giving rise to $U = U_1 \otimes \cdots \otimes U_n$, and the measurement basis is a product basis $|b\rangle = |b_1\rangle \otimes \cdots \otimes |b_n\rangle$. For a product input $X = X_1 \otimes \cdots \otimes X_n$, the channel action is

$$\begin{aligned}\mathcal{M}(X) &= \mathbb{E}_U \sum_{b_1,\ldots,b_n} \mathrm{tr}\left(\bigotimes_{j=1}^n |b_j\rangle\langle b_j| \bigotimes_{k=1}^n U_k X_k U_k^\dagger\right) \bigotimes_{l=1}^n U_l^\dagger |b_l\rangle\langle b_l| U_l \\ &= \bigotimes_{j=1}^n \left(\mathbb{E}_{U_j} \sum_{b_j} \mathrm{tr}(|b_j\rangle\langle b_j| U_j X_j U_j^\dagger) U_j^\dagger |b_j\rangle\langle b_j| U_j\right) = \bigotimes_{j=1}^n \mathcal{M}_1(X_j).\end{aligned}$$

The form of the single-qubit channel $\mathcal{M}_1$ follows from the Clifford case with $d = 2$. The inverse also factorizes accordingly. □

Proposition 108 (Pauli Shadows)**.** *For the random Pauli measurement ensemble:*

(i) The snapshot is a tensor product: $\hat{\rho} = \bigotimes_{j=1}^n (3U_j^\dagger |\hat{b}_j\rangle\langle\hat{b}_j| U_j - \mathrm{Id})$.

(ii) For a Pauli operator $O = P_1 \otimes \cdots \otimes P_n$, where $P_i \in \{I, X, Y, Z\}$ and only k P_i's are not identity, the shadow norm is exactly $\|O\|_{\mathrm{shadow}}^2 = 3^k$.

Proof. (i) This follows directly from the factorized inverse channel derived in the preceding lemma, applied to the product state $U^\dagger |\hat{b}\rangle\langle\hat{b}| U = \bigotimes_j U_j^\dagger |\hat{b}_j\rangle\langle\hat{b}_j| U_j$. For each qubit j, we have $\mathrm{tr}(U_j^\dagger |\hat{b}_j\rangle\langle\hat{b}_j| U_j) = 1$, so the single-qubit inverse map $\mathcal{M}_1^{-1}(Y) = 3Y - \mathrm{tr}(Y) \cdot \mathrm{Id}$ yields the desired form.

(ii) Let the observable be $O = \bigotimes_{j=1}^n P_j$. First, we compute the action of the inverse channel on O. Since the channel factorizes, so does its inverse:

$$\mathcal{M}^{-1}(O) = \bigotimes_{j=1}^n \mathcal{M}_1^{-1}(P_j).$$

For each qubit, if $P_j \in \{X, Y, Z\}$, it is traceless, so $\mathcal{M}_1^{-1}(P_j) = 3P_j$. If $P_j = \mathrm{Id}$, it has trace 2, so $\mathcal{M}_1^{-1}(\mathrm{Id}) = 3\mathrm{Id} - \mathrm{tr}(\mathrm{Id}) \cdot \mathrm{Id} = 3\mathrm{Id} - 2\mathrm{Id} = \mathrm{Id}$. Therefore,

$$\mathcal{M}^{-1}(O) = \left(\bigotimes_{j:P_j \neq \mathrm{Id}} 3P_j\right) \otimes \left(\bigotimes_{j:P_j = \mathrm{Id}} \mathrm{Id}_j\right) = 3^k O.$$

The shadow norm is the maximum of $\mathrm{tr}(\sigma L)$ over any state σ for the operator

$$L = \mathbb{E}_U \sum_b (U^\dagger|b\rangle\langle b|U) \left(\langle b|U\mathcal{M}^{-1}(O)U^\dagger|b\rangle\right)^2.$$

Substituting our result for $\mathcal{M}^{-1}(O)$:

$$L = (3^k)^2 \mathbb{E}_U \sum_b (U^\dagger|b\rangle\langle b|U) \left(\langle b|UOU^\dagger|b\rangle\right)^2.$$

Because the ensemble, basis, and observable are all tensor products ($U = \bigotimes_j U_j$, $|b\rangle = \bigotimes_j |b_j\rangle$, $O = \bigotimes_j P_j$), the operator L itself factorizes into a tensor product of single-qubit operators, $L = \bigotimes_{j=1}^n L_j$:

$$L = (3^k)^2 \bigotimes_{j=1}^n \left(\mathbb{E}_{U_j} \sum_{b_j} U_j^\dagger|b_j\rangle\langle b_j|U_j \left(\langle b_j|U_j P_j U_j^\dagger|b_j\rangle\right)^2 \right) = 9^k \bigotimes_{j=1}^n L_j.$$

We now evaluate the single-qubit operator L_j for the two cases:

If $P_j = \mathrm{Id}$: The squared term is $(\langle b_j|U_j \cdot \mathrm{Id} \cdot U_j^\dagger|b_j\rangle)^2 = 1^2 = 1$. Then

$$L_j = \mathbb{E}_{U_j} \sum_{b_j} U_j^\dagger|b_j\rangle\langle b_j|U_j = \mathbb{E}_{U_j} \left[U_j^\dagger \left(\sum_{b_j} |b_j\rangle\langle b_j| \right) U_j \right] = \mathrm{Id}.$$

If $P_j \in \{X, Y, Z\}$: P_j is traceless. So the operator L_j is precisely the sum over the basis states of the operator in Eq. (28) with $d = 2$ and $A_0 = B_0 = P_j$.

$$\begin{aligned} L_j &= \sum_{b_j} \mathbb{E}_{U_j} [\langle b_j|U_j P_j U_j^\dagger|b_j\rangle^2 U_j^\dagger|b_j\rangle\langle b_j|U_j] \\ &= d \cdot \frac{\mathrm{tr}(P_j^2)\mathrm{Id} + 2P_j^2}{d(d+1)(d+2)} \quad (\text{where } d = 2) \\ &= 2 \cdot \frac{2\mathrm{Id} + 2\mathrm{Id}}{2(3)(4)} = \frac{4\mathrm{Id}}{12} = \frac{1}{3}\mathrm{Id}. \end{aligned}$$

We have k operators of the form $\frac{1}{3}\mathrm{Id}$ and $n-k$ operators of the form Id. Assembling the full operator L:

$$L = 9^k \left(\bigotimes_{j:P_j \neq \mathrm{Id}} \frac{1}{3}\mathrm{Id} \right) \otimes \left(\bigotimes_{j:P_j = \mathrm{Id}} \mathrm{Id} \right) = 9^k \cdot \left(\frac{1}{3}\right)^k \cdot \mathrm{Id}^{\otimes n} = 3^k \mathrm{Id}.$$

The shadow norm is simply 3^k:

$$\|O\|_{\text{shadow}}^2 = \max_\sigma \mathrm{tr}(\sigma L) = \max_\sigma \mathrm{tr}(\sigma \cdot 3^k \mathrm{Id}) = 3^k \max_\sigma \mathrm{tr}(\sigma) = 3^k.$$

This completes the proof. □

Pauli shadows are ideal for problems where the observables we are interested in predicting can be decomposed to a few low-weight Pauli operators. Examples include two-point correlation function, energy, and energy variance.