density matrices, and let $\Pi_{\mathcal{D}}$ be the Euclidean (Frobenius) projection onto $\mathcal{D}$. Since $\mathcal{D}$ is closed and convex, $\Pi_{\mathcal{D}}$ is nonexpansive:

$$\|\Pi_{\mathcal{D}}(A) - \Pi_{\mathcal{D}}(B)\|_F \leq \|A - B\|_F \quad \textit{for all } A, B.$$

In particular, because $\rho \in \mathcal{D}$,

$$\|\rho - \Pi_{\mathcal{D}}(\widehat{\rho})\|_F \leq \|\rho - \widehat{\rho}\|_F.$$

Thus the Frobenius bound from Corollary 50 continues to hold (and can only improve) after projecting $\widehat{\rho}$ onto $\mathcal{D}$. Operationally, if $\widehat{\rho} = U\mathrm{diag}(\lambda)U^\dagger$ is an eigendecomposition, then $\Pi_{\mathcal{D}}(\widehat{\rho})$ is obtained by projecting the eigenvalue vector λ onto the probability simplex $\{\mu \geq 0 : \sum_i \mu_i = 1\}$ (via the usual simplex projection) and setting $\widetilde{\rho} = U\mathrm{diag}(\mu)U^\dagger$.

Before moving on to fancier versions of quantum state tomography, we comment here on how practical it is to perform it. Consider a system of 8 qubits, corresponding to $d = 2^8 = 256$; therefore a ρ is described by $65,535$ real numbers. Such a quantum state tomography (with a slightly different method than the one presented above) was performed in [**HHR+05**] with about 10 hours of data acquisition. Note that each individual qubit increases the number of parameters and data acquisition time exponentially. As such, full quantum state tomography is often totally impractical even for modest system sizes.

Nonetheless, we study full quantum state tomography since it is a fundamental problem in quantum learning theory, and allows us to build tools for more pragmatic, 'partial' forms of state tomography which are highly practical and often used.

2. Learning a state in the operator norm

In many applications we do not need to reconstruct ρ entrywise; rather, we want to *predict expectation values* of observables with respect to ρ. If $\widehat{\rho}$ is an estimate, the prediction error for an observable O is

$$\big|\langle O\rangle_\rho - \langle O\rangle_{\widehat{\rho}}\big| = \big|\mathrm{tr}\big(O(\rho - \widehat{\rho})\big)\big|.$$

Different matrix norms control this quantity for different classes of O. We next introduce the operator norm and explain when it is the right notion of accuracy.

Definition 52 (Operator (spectral) norm)**.** *The* ***operator norm*** *of a matrix $A \in \mathbb{C}^{d\times d}$ is*

$$\|A\|_{\mathrm{op}} := \sup_{\|v\|_2=1} \|Av\|_2 = \sigma_{\max}(A)\,,$$

which is the largest singular value of A. If A is Hermitian, then $\|A\|_{\mathrm{op}} = \max_k |\lambda_k(A)|$, the largest eigenvalue magnitude. We will sometimes also write $\|\cdot\|_\infty$ for $\|\cdot\|_{\mathrm{op}}$.

Two elementary inequalities will be useful: the Hilbert–Schmidt Cauchy–Schwarz bound

$$|\mathrm{tr}(X^\dagger Y)| \leq \|X\|_{\mathrm{F}}\|Y\|_{\mathrm{F}}\,,$$

and the trace–operator Hölder bound (duality of $\|\cdot\|_1$ and $\|\cdot\|_\infty$)

$$|\mathrm{tr}(XY)| \leq \|X\|_1\|Y\|_\infty\,.$$

Using these inequalities we have the following:

Proposition 53 (Expectation bounds via operator norm)**.** *For any Hermitian observable O and states $\rho, \widehat{\rho}$,*

$$\big|\mathrm{tr}\big(O(\rho - \widehat{\rho})\big)\big| \leq \|O\|_1 \, \|\rho - \widehat{\rho}\|_\infty \,, \tag{22}$$

$$\big|\mathrm{tr}\big(O(\rho - \widehat{\rho})\big)\big| \leq \|O\|_{\mathrm{F}} \, \|\rho - \widehat{\rho}\|_{\mathrm{F}} \,. \tag{23}$$

In particular, if O is a projector P of rank r, then $\|P\|_1 = r$ and

$$\big|\mathrm{tr}\big(P(\rho - \widehat{\rho})\big)\big| \leq r \, \|\rho - \widehat{\rho}\|_\infty \,.$$

Thus an operator–norm guarantee on the state directly controls probabilities of low-rank projectors (e.g. rank-1 tests are controlled with constant factor).

Equation (22) is especially useful when the observables of interest have small trace norm (low rank or few outcomes with bounded weights). Conversely, when we only know $\|O\|_\infty \leq 1$ (a very common normalization), (23) gives

$$\big|\mathrm{tr}\big(O(\rho - \widehat{\rho})\big)\big| \leq \|O\|_{\mathrm{F}} \|\rho - \widehat{\rho}\|_{\mathrm{F}} \leq \sqrt{\mathrm{rank}(O)} \, \|\rho - \widehat{\rho}\|_{\mathrm{F}} \leq \sqrt{d} \, \|\rho - \widehat{\rho}\|_{\mathrm{F}} \,,$$

which shows a $\sqrt{d}$ amplification in worst case (e.g. full-rank observables). Hence a Frobenius guarantee $\|\rho - \widehat{\rho}\|_{\mathrm{F}} \leq \varepsilon_{\mathrm{F}}$ only implies

$$\big|\langle O \rangle_\rho - \langle O \rangle_{\widehat{\rho}}\big| \lesssim \sqrt{d} \, \varepsilon_{\mathrm{F}} \quad \text{for } \|O\|_\infty \leq 1,$$

and to achieve a constant error in such expectation values one needs $\varepsilon_{\mathrm{F}} = O(1/\sqrt{d})$. This is precisely why a Frobenius bound from entrywise tomography may not translate to a useful operator-norm control for large d.

Remark 54 (Relating Frobenius and operator norms)**.** *For any matrix X,*

$$\|X\|_\infty \leq \|X\|_{\mathrm{F}} \leq \sqrt{d} \, \|X\|_\infty.$$

Under our entrywise scheme, a uniform per-entry accuracy $\max_{i,j} |X_{ij}| \leq \varepsilon_{\max}$ yields $\|X\|_{\mathrm{F}} \leq d \, \varepsilon_{\max}$ (and therefore $\|X\|_\infty \leq d \, \varepsilon_{\max}$). Consequently, to guarantee an operator-norm target $\|\rho - \widehat{\rho}\,\|_\infty \leq \varepsilon$ via entrywise control, one must take $\varepsilon_{\max} = \varepsilon/d$, which (through the Hoeffding scaling $1/\varepsilon^2_{\max}$) inflates the copy complexity by a factor d^2 relative to the entrywise case, i.e. to $O\big(\frac{d^3}{\varepsilon^2} \log(d^2/\delta)\big)$, as explained above. By contrast, operator-norm accuracy immediately controls all rank-r projector expectations within $r \, \varepsilon$ by (22).

The above considerations motivate us to consider quantum state tomography that is better-suited to the operator norm, so that we can get a better query complexity bound in that setting. There are many approaches to do this, but here we follow the proof strategy of [**CHL+23**] which is based in part on [**GKKT20**]. We will show that there exists an estimator $\widehat{\rho}$ such that with probability at least $1 - \delta$ we have $\|\rho - \hat{\rho}\|_\infty \leq \varepsilon$ with at most $O(\frac{d + \log(1/\delta)}{\varepsilon^2})$ measurements. This is much better than our naïve approach above, by a factor of around $\sim d^2 \log(d)$.

To begin, we need to consider the uniform POVM on the sphere in $\mathbb{C}^d$, which we develop below.

2.1. The uniform POVM on the sphere and its Naimark dilation

We briefly recall the measurement formalism. A POVM is a finite (or measurable) collection of positive semidefinite operators $\{M_z\}$ summing to the identity; upon measuring ρ, outcome z is observed with probability $\mathrm{tr}(\rho M_z)$ (after which the state is discarded). See Definition 2.1 for our conventions. Throughout our discussion here, all POVMs will be rank 1.

Let $\mathbb{S}^{2d-1} \subset \mathbb{C}^d$ denote the unit sphere with normalized Haar measure $\mathrm{d}v$ (so $\int_{\mathbb{S}^{2d-1}} \mathrm{d}v = 1$). The *uniform POVM* is the continuous-outcome POVM with operator–valued density

$$M(\mathrm{d}v) := d\,|v\rangle\langle v|\,\mathrm{d}v\,,$$

so that for any measurable $B \subseteq \mathbb{S}^{2d-1}$ we have $M(B) = \int_B d\,|v\rangle\langle v|\,\mathrm{d}v$. It is well-defined because $\int_{\mathbb{S}^{2d-1}} |v\rangle\langle v|\,\mathrm{d}v = \alpha\,\mathbb{1}$ by unitary invariance, and taking traces gives $1 = \int \mathrm{tr}(|v\rangle\langle v|)\,\mathrm{d}v = \mathrm{tr}(\alpha I) = \alpha\,d$, hence $\alpha = 1/d$, i.e. $\int_{\mathbb{S}} d\,|v\rangle\langle v|\,\mathrm{d}v = \mathbb{1}$. When ρ is measured with this POVM, the outcome $v \in \mathbb{S}^{2d-1}$ has density $p(v)\,\mathrm{d}v = \mathrm{tr}\big(\rho\,M(\mathrm{d}v)\big) = d\,\langle v|\rho|v\rangle\,\mathrm{d}v$.

Recall that Naimark's theorem says any POVM can be realized as a projective measurement (PVM) on a larger Hilbert space, followed by discarding the ancillas. Concretely, consider the following example.

Example 1 (discrete case): Suppose we approximate the uniform POVM by a finite frame $\{w_k, |v_k\rangle\}_{k=1}^m$ with weights $w_k > 0$ obeying $\sum_k w_k = d$ and $\sum_k w_k |v_k\rangle\langle v_k| = \mathbb{1}$. Define POVM elements $M_k = w_k\,|v_k\rangle\langle v_k|$. Let the "pointer" ancilla belong to the Hilbert space $\mathcal{K} \simeq \mathbb{C}^m$ with basis $\{|k\rangle\}_{k=1}^m$, and define the isometry

$$V : \mathcal{H} \longrightarrow \mathcal{H} \otimes \mathcal{K}, \quad V|\psi\rangle = \sum_{k=1}^m |v_k\rangle \otimes \left(\sqrt{w_k}\,\langle v_k|\psi\rangle\right)|k\rangle\,.$$

If we then measure the ancilla in the computational basis with projectors $\Pi_k = I \otimes |k\rangle\langle k|$, this induces the POVM

$$V^\dagger \Pi_k V = w_k\,|v_k\rangle\langle v_k| = M_k$$

on our original system, where $\Pr[k] = \mathrm{tr}(\rho M_k)$. We have thus realized our discrete POVM via a PVM on $\mathcal{H} \otimes \mathcal{K}$.

We can generalize the example above to the setting of our desired continuous uniform POVM. We replace the finite-dimensional pointer Hilbert space by the infinite-dimensional pointer Hilbert space $\mathcal{K} \simeq L^2(\mathbb{S}^{2d-1}, \mathrm{d}v)$ with basis $\{|v\rangle : v \in \mathbb{S}^{2d-1}\}$ and define the isometry $V : \mathcal{H} \to \mathcal{H} \otimes \mathcal{K}$ by

$$V|\psi\rangle = \int_{\mathbb{S}^{2d-1}} \sqrt{d}\,\langle v|\psi\rangle |v\rangle_{\mathcal{H}} \otimes |v\rangle_{\mathcal{K}}\,\mathrm{d}v \in \mathcal{H} \otimes \mathcal{K}\,,$$

where here we have put $\mathcal{H}$ and $\mathcal{K}$ subscripts on the $|v\rangle$'s for clarity. Let $\Pi(B)$ be the PVM on $\mathcal{K}$ given by multiplication by the indicator of $B \subseteq \mathbb{S}^{2d-1}$, i.e. $[\Pi(B)\phi](v) = \mathbf{1}_B(v)\,\phi(v)$. Then

$$V^\dagger \Pi(B) V = \int_B d\,|v\rangle\langle v|\,\mathrm{d}v = M(B),$$

because for any $|\psi\rangle$, we have $\langle\psi|V^\dagger\Pi(B)V|\psi\rangle = \int_B d\,|\langle v|\psi\rangle|^2\,\mathrm{d}v = \langle\psi|M(B)|\psi\rangle$. Thus the uniform POVM is induced on $\mathcal{H}$ by projected measurement on the extended space.

In implementations one replaces the continuous POVM by a finite approximation (e.g. randomly sampled Haar vectors or a spherical 2-design) and uses a corresponding discrete version as we gave in the example above. For convenience, we will stick with the continuous version in our proofs below.

2.2. Learning a density matrix with a continuous POVM

Consider, as above, the POVM given by the continuous operator density $M(\mathrm{d}v) := d\,|v\rangle\langle v|\,\mathrm{d}v$. Suppose we are given a device that prepares copies of a unknown density matrix ρ, and that we measure each ρ we are given with the POVM. Let $|v_i\rangle$ be the outcome of the ith measurement where each $v_i \in \mathbb{S}^{2d-1} \subset \mathbb{C}^d$. Then for our improved quantum state tomography procedure, we will use the estimator $H_N(\rho) = H_N(\rho, v_1, ..., v_N)$ given by

$$H_N(\rho) := \frac{1}{N}\sum_{i=1}^{N}\left((d+1)|v_i\rangle\langle v_i| - \mathbb{1}\right)\,.$$

Then we have the following result.

Theorem 55 (Quantum state tomography for the operator norm)**.** *For accuracy $\varepsilon \in (0,1)$ and confidence $\delta \in (0,1)$. Then with probability at least $1-\delta$, we obtain $\widehat{\rho} := H_N(\rho)$ such that*

$$\|\rho - \widehat{\rho}\,\|_\infty \leq C\,\max\left\{\frac{d+\log(1/\delta)}{N}\,,\,\sqrt{\frac{d+\log(1/\delta)}{N}}\right\},$$

for some universal constant C. Letting $N = O(\frac{d+\log(1/\delta)}{\varepsilon^2})$, we in particular find that with probability at least $1-\delta$ we obtain $\widehat{\rho} := H_N(\rho)$ such that

$$\|\rho - \widehat{\rho}\,\|_\infty \leq \varepsilon\,.$$

To prove this theorem, we require several lemmas.

Lemma 56 (Second moment of the uniform POVM)**.** *Let $\mathbb{S}^{2d-1} \subset \mathbb{C}^d$ be the unit sphere equipped with the normalized Haar measure $\mathrm{d}v$ (so $\int_{\mathbb{S}^{2d-1}} \mathrm{d}v = 1$). Then*

$$\int_{\mathbb{S}^{2d-1}} |v\rangle\langle v| \otimes |v\rangle\langle v|\;\mathrm{d}v = \frac{1}{d(d+1)}\big(\mathbb{1} + \mathsf{SWAP}\big),$$

where SWAP is the swap operator on $\mathcal{H}\otimes\mathcal{H}$.

Proof. We first record the one-fold identity. Set

$$B := \int_{\mathbb{S}^{2d-1}} |v\rangle\langle v|\;\mathrm{d}v.$$

For any unitary U on $\mathcal{H}$, the change of variables $v \mapsto Uv$ gives

$$UBU^\dagger = \int |Uv\rangle\langle Uv|\;\mathrm{d}v = B.$$

Thus B commutes with every unitary and hence $B = \alpha\,\mathbb{1}$ for some scalar α. Taking traces,

$$1 = \mathrm{tr}(B) = \alpha\,\mathrm{tr}(\mathbb{1}) = \alpha\,d,$$

so $\alpha = 1/d$ and therefore

$$\int_{\mathbb{S}^{2d-1}} |v\rangle\langle v|\;\mathrm{d}v = \frac{\mathbb{1}}{d}\,.$$

Now set

$$A := \int_{\mathbb{S}^{2d-1}} |v\rangle\langle v| \otimes |v\rangle\langle v| \; \mathrm{d}v \,.$$

We will show $A = \alpha\,\mathbb{1} + \beta\ \mathsf{SWAP}$ and then determine α, β by two trace identities. For any unitary U,

$$(U \otimes U)A(U^\dagger \otimes U^\dagger) = A,$$

and, because the integrand is symmetric under swapping tensor factors,

$$\mathsf{SWAP}\ A\ \mathsf{SWAP} = A\,.$$

Working in the usual computational basis $\{|i\rangle\}_{i=1}^{d}$, we write

$$A_{ij,k\ell} := \langle i|\langle j|A|k\rangle|\ell\rangle.$$

Taking U to be a diagonal phase matrix $D(\theta) = \mathrm{diag}(e^{\mathrm{i}\theta_1}, \dots, e^{\mathrm{i}\theta_d})$ and comparing matrix elements yields

$$A_{ij,kl} = e^{\mathrm{i}(\theta_i+\theta_j-\theta_k-\theta_\ell)}\, A_{ij,k\ell} \quad \text{for all } \theta_i, \theta_j, \theta_k, \theta_\ell \in \mathbb{R}^d\,.$$

Varying the phases independently forces $A_{ij,k\ell} = 0$ unless the sets $\{i, j\}$ and $\{k, l\}$ coincide. Hence the only potentially nonzero entries are

$$A_{ij,ij} \quad \text{and} \quad A_{ij,ji}\,.$$

Invariance under permutation matrices then implies these coefficients depend only on whether $i = j$ or $i \neq j$. Thus there exist scalars a, b, c such that

$$A = \sum_i a\,|ii\rangle\langle ii| + \sum_{i\neq j} b\,|ij\rangle\langle ij| + \sum_{i\neq j} c\,|ij\rangle\langle ji|.$$

Commuting with SWAP further forces, on each two-dimensional block $\mathrm{span}\{|ij\rangle, |ji\rangle\}$ with $i \neq j$, the form

$$\begin{pmatrix} b & c \\ c & b \end{pmatrix},$$

which is a linear combination of the identity and the flip on that block. Noting that

$$\mathbb{1} = \sum_{i,j} |ij\rangle\langle ij|\,, \quad \mathsf{SWAP} = \sum_{i,j} |ij\rangle\langle ji|\,,$$

we may rewrite

$$A = \alpha\,\mathbb{1} + \beta\ \mathsf{SWAP}$$

where $\alpha = b$, $\beta = c$, and $a = \alpha + \beta$.

Next we set out to determine α, β. First,

$$\mathrm{tr}(A) = \int \mathrm{tr}\big(|v\rangle\langle v| \otimes |v\rangle\langle v|\big)\ \mathrm{d}v = \int 1\ \mathrm{d}v = 1.$$

On the other hand,

$$\mathrm{tr}(A) = \alpha\,\mathrm{tr}(\mathbb{1}) + \beta\,\mathrm{tr}(\mathsf{SWAP}) = \alpha\, d^2 + \beta\, d\,.$$

Second, using $\mathrm{tr}\big(\mathsf{SWAP}(X \otimes Y)\big) = \mathrm{tr}(XY)$,

$$\mathrm{tr}(\mathsf{SWAP}\,A) = \int \mathrm{tr}\big((|v\rangle\langle v| \otimes |v\rangle\langle v|)\,\mathsf{SWAP}\,\big)\ \mathrm{d}v = \int \mathrm{tr}\big(|v\rangle\langle v| \cdot |v\rangle\langle v|\big)\ \mathrm{d}v = \int 1\ \mathrm{d}v = 1\,,$$

while

$$\operatorname{tr}(\mathsf{SWAP}\,A) = \alpha\operatorname{tr}(\mathsf{SWAP}) + \beta\operatorname{tr}(\mathbb{1}) = \alpha\, d + \beta\, d^2.$$

The linear system

$$d^2\,\alpha + d\,\beta = 1, \quad d\,\alpha + d^2\,\beta = 1$$

has the unique solution

$$\alpha = \beta = \frac{1}{d(d+1)}\,,$$

and therefore

$$A = \frac{1}{d(d+1)}\big(\mathbb{1} + \mathsf{SWAP}\,\big)\,.$$

This is the desired identity. □

We also have the similar result below.

Lemma 57 (Third moment of the uniform POVM)**.** *Letting $\mathbb{S}^{2d-1} \subset \mathbb{C}^d$ as before and considering the normalized Haar measure, we have the identity*

$$\Pi_3 := \int_{\mathbb{S}^{2d-1}} |v\rangle\langle v| \otimes |v\rangle\langle v| \otimes |v\rangle\langle v|\,\mathrm{d}v = \frac{1}{d(d+1)(d+2)}\sum_{\pi\in S_3}\mathrm{Perm}(\pi)\,,$$

where S_3 is the set of permutations on three items.

The proof proceeds in a similar, albeit more tedious way, than the one above. Later, we will demonstrate a more high-powered way to bound all of the moments.

Next we require the notion of an ε-net:

Definition 58 (ε-net on pure states)**.** *Let $(\mathcal{X}, d)$ be a metric space. A subset $\mathcal{N} \subseteq \mathcal{X}$ is an ε-**net** if for every $x \in \mathcal{X}$ there exists $y \in \mathcal{N}$ with $d(x, y) \leq \varepsilon$.*

For our purposes, $\mathcal{X}$ will be the unit sphere of pure states in a d-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$, and we take

$$d_{\mathrm{F}}\big(|u\rangle, |u'\rangle\big) := \big\|\, |u\rangle\langle u| - |u'\rangle\langle u'|\,\big\|_{\mathrm{F}}.$$

It is often convenient to work with the projective Euclidean ("chordal") distance

$$d_{\mathrm{E}}\big(|u\rangle, |u'\rangle\big) := \min_{\phi\in\mathbb{R}}\big\|\,|u\rangle - e^{i\phi}|u'\rangle\,\big\|_2,$$

or with the Fubini–Study geodesic distance; these metrics are equivalent up to universal constants. In particular, for unit vectors

$$d_{\mathrm{E}}\big(|u\rangle, |u'\rangle\big) \leq d_{\mathrm{F}}\big(|u\rangle, |u'\rangle\big) \leq \sqrt{2}\, d_{\mathrm{E}}\big(|u\rangle, |u'\rangle\big),$$

since $d_{\mathrm{F}}^2 = 2\left(1 - |\langle u|u'\rangle|^2\right)$ and $d_{\mathrm{E}}^2 = 2(1 - |\langle u|u'\rangle|)$.

Lemma 59 (Volumetric ε-net bound)**.** *Let $0 < \varepsilon \leq 1$. There exists an ε-net $\mathcal{N}$ for the unit sphere of pure states in $\mathcal{H} \simeq \mathbb{C}^d$ (with respect to the projective Euclidean metric) of cardinality at most*

$$|\mathcal{N}| \leq \left(\frac{3}{\varepsilon}\right)^{2d}.$$

By the metric equivalence in Definition 58, the same bound holds for d_{F} up to a universal rescaling of ε.

PROOF SKETCH. Identify $\mathbb{C}^d \simeq \mathbb{R}^{2d}$ and view the pure-state sphere as $\mathbb{S}^{2d-1}$. Let $\mathcal{M} \subset \mathbb{S}^{2d-1}$ be a maximal ε-separated set in the ambient Euclidean metric. Then the closed Euclidean balls $\{B(x, \varepsilon/2) : x \in \mathcal{M}\}$ are disjoint and all lie inside the radius-$(1+\varepsilon/2)$ ball in $\mathbb{R}^{2d}$. Comparing volumes yields

$$|\mathcal{M}| \operatorname{vol}\big(B_{2d}(\varepsilon/2)\big) \leq \operatorname{vol}\big(B_{2d}(1+\varepsilon/2)\big),$$

so

$$|\mathcal{M}| \leq \left(\tfrac{1+\varepsilon/2}{\varepsilon/2}\right)^{2d} = \left(\tfrac{2+\varepsilon}{\varepsilon}\right)^{2d} \leq \left(\tfrac{3}{\varepsilon}\right)^{2d}$$

since $\varepsilon \leq 1$. Maximality of $\mathcal{M}$ implies it is an ε-net for the Euclidean metric on $\mathbb{S}^{2d-1}$; since $d_{\mathrm{E}} \leq \|\cdot\|_2$ on the sphere, the same set is an ε-net for d_{E}. Finally, the equivalence between d_{E} and d_{F} transfers the bound to the Frobenius–projector metric (at the cost of a universal constant in ε), completing the proof. □

Finally, we need one more result, which is a highly useful inequality.

Lemma 60 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_n$ be independent, mean-zero real random variables. Assume $|X_i| \leq b$ almost surely and set $\sigma^2 := \sum_{i=1}^n \mathbb{E}[X_i^2]$. Then for all $t > 0$,*

$$\Pr\left[\left|\sum_{i=1}^n X_i\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2(\sigma^2 + bt/3)}\right).$$

Equivalently, for the empirical mean $\overline{X} = \frac{1}{n}\sum_i X_i$,

$$\Pr\big[\,|\overline{X}| \geq \eta\,\big] \leq 2\exp\left(-\frac{n\,\eta^2}{2(\overline{\sigma}^2 + b\,\eta/3)}\right), \quad \overline{\sigma}^2 := \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i^2].$$

Let us unpack the Bernstein inequality. A mean-zero random variable X is *sub-Gaussian* with proxy variance ν^2 if its moment generating function obeys $\mathbb{E}[\exp(\lambda X)] \leq \exp(\lambda^2\nu^2/2)$ for all $\lambda \in \mathbb{R}$. This implies Gaussian-type concentration $\Pr(|X| \geq t) \leq 2\exp(-t^2/(2\nu^2))$. Bounded variables are sub-Gaussian (with $\nu \lesssim b$ when $|X| \leq b$), and sums of independent sub-Gaussians remain sub-Gaussian with variance proxy adding in quadrature. A closely related class is *sub-exponential*: X is sub-exponential (with parameters (α, β)) if $\mathbb{E}[\exp(\lambda X)] \leq \exp(\frac{\alpha^2\lambda^2}{2})$ for $|\lambda| \leq 1/\beta$. Bernstein's bound quantitatively captures the sum of independent sub-exponential variables: the tail looks Gaussian $\sim \exp(-ct^2)$ for moderate deviations and transitions to exponential $\sim \exp(-ct)$ beyond a scale set by the individual "heaviness", here given by b.

With the above lemmas at hand, we are now prepared to give the proof of Theorem 55:

PROOF OF THEOREM 55. We measure each copy of ρ with the uniform POVM $M(\mathrm{d}v) = d\,|v\rangle\langle v|\,\mathrm{d}v$ on the unit sphere $\mathbb{S}^{2d-1} \subset \mathbb{C}^d$. Let $v_1, \ldots, v_N$ be the outcomes, and we recall the estimator

$$H_N(\rho) := \frac{1}{N}\sum_{i=1}^N \Big((d+1)\,|v_i\rangle\langle v_i| - \mathbb{1}\Big).$$

We will show that, with probability at least $1-\delta$,

$$\|H_N(\rho) - \rho\|_\infty \leq C\,\max\left\{\frac{d + \log(1/\delta)}{N},\ \sqrt{\frac{d + \log(1/\delta)}{N}}\right\}.$$

Let $p_\rho(v)\,\mathrm{d}v$ denote the outcome density of the uniform POVM on ρ. By definition,

$$p_\rho(v)\,\mathrm{d}v = \mathrm{tr}\big(\rho\, M(\mathrm{d}v)\big) = d\,\langle v|\rho|v\rangle\,\mathrm{d}v.$$

Using the second–moment identity of the uniform POVM from Lemma 56 together with $\int |v\rangle\langle v|\,\mathrm{d}v = \mathbb{1}/d$ (which we also derived in the proof of the same lemma), we compute

$$\begin{aligned}\mathbb{E}_{v\sim p_\rho}\big[\,|v\rangle\langle v|\,\big] &= d\int \langle v|\rho|v\rangle\,|v\rangle\langle v|\,\mathrm{d}v = d\,\mathrm{tr}_2\Big[(\mathbb{1}\otimes\rho)\int |v\rangle\langle v|\otimes|v\rangle\langle v|\,\mathrm{d}v\Big]\\ &= d\,\mathrm{tr}_2\Big[(\mathbb{1}\otimes\rho)\,\frac{\mathbb{1}+\mathsf{SWAP}}{d(d+1)}\Big] = \frac{1}{d+1}\,\mathrm{tr}_2\big[\mathbb{1}\otimes\rho+\mathsf{SWAP}(\mathbb{1}\otimes\rho)\big]\\ &= \frac{1}{d+1}\,\big(\mathbb{1}+\rho\big).\end{aligned}$$

Therefore $\mathbb{E}\big[\,(d+1)\,|v\rangle\langle v|-\mathbb{1}\,\big]=\rho$, giving us $\mathbb{E}\big[H_N(\rho)\big]=\rho$. This establishes that our estimator $H_N(\rho)$ is unbiased.

Fix now a unit vector $|u\rangle$. Define the centered scalar random variables

$$Y_i := \langle u|\big((d+1)\,|v_i\rangle\langle v_i| - \mathbb{1} - \rho\big)|u\rangle = (d+1)\,|\langle u|v_i\rangle|^2 - 1 - \langle u|\rho|u\rangle,$$

so that $\mathbb{E}[Y_i]=0$ and

$$\langle u|\big(H_N(\rho)-\rho\big)|u\rangle = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

Our goal is to control $\sup_{\|u\|=1}\Big|\frac{1}{N}\sum_i Y_i\Big| = \|H_N(\rho)-\rho\|_\infty$. Now for any v, $|\langle u|v\rangle|^2\in[0,1]$, and $\langle u|\rho|u\rangle\in[0,1]$. Therefore $-2\le Y_i\le d$, and so we can take $|Y_i|\le b:=d+1$. Let $Z:=|\langle u|v\rangle|^2\in[0,1]$. Using Lemma 56 we have

$$\mathbb{E}_{v\sim p_\rho}[Z] = \langle u|\,\mathbb{E}_{v\sim p_\rho}[\,|v\rangle\langle v|\,]\,|u\rangle = \frac{1+\langle u|\rho|u\rangle}{d+1} \le \frac{2}{d+1}\,.$$

We now use Lemma 57 to bound the second moment of Z:

$$\begin{aligned}\mathbb{E}_{v\sim p_\rho}[Z^2] &= d\int \langle v|\rho|v\rangle\,|\langle u|v\rangle|^4\,\mathrm{d}v = d\,\mathrm{tr}\Big[\big(\rho\otimes|u\rangle\langle u|\otimes|u\rangle\langle u|\big)\int |v\rangle\langle v|^{\otimes 3}\,\mathrm{d}v\Big]\\ &= d\,\mathrm{tr}\Big[\big(\rho\otimes|u\rangle\langle u|\otimes|u\rangle\langle u|\big)\frac{1}{d(d+1)(d+2)}\sum_{\pi\in S_3}\mathrm{Perm}(\pi)\Big]\\ &\le \frac{d}{d(d+1)(d+2)}\sum_{\pi\in S_3}\mathrm{tr}\Big[\big(\rho\otimes|u\rangle\langle u|\otimes|u\rangle\langle u|\big)\mathrm{Perm}(\pi)\Big]\\ &\le \frac{6}{(d+1)(d+2)}.\end{aligned}$$

The last inequality uses that each trace term is at most 1 (e.g., for the identity permutation it equals $\mathrm{tr}(\rho)\mathrm{tr}(|u\rangle\langle u|)^2=1$), while for any other $\pi\in S_3$ it reduces to either $\langle u|\rho|u\rangle$ or 1. Thus we have the variance

$$\begin{aligned}\mathrm{Var}\big((d+1)Z-1\big) &= (d+1)^2\Big(\mathbb{E}[Z^2]-(\mathbb{E}Z)^2\Big)\\ &\le (d+1)^2\cdot\frac{6}{(d+1)(d+2)} \le 6.\end{aligned}$$

Recalling $Y_i = (d+1)Z - 1 - \langle u|\rho|u\rangle$, we have $\mathbb{E}[Y_i] = 0$, $\mathrm{Var}(Y_i) \le 6$, and the range $|Y_i| \le b := d+1$ (since $Z \in [0,1]$ and $\langle u|\rho|u\rangle \in [0,1]$). For the empirical mean $\overline{Y} := \frac{1}{N}\sum_{i=1}^N Y_i$, Bernstein's inequality (Lemma 60) gives, for all $\eta > 0$,

$$\Pr\big[\,|\overline{Y}| \ge \eta\,\big] \le 2\exp\!\left(-\frac{N\,\eta^2}{2(\overline{\sigma}^2 + b\,\eta/3)}\right) \le 2\exp\!\Big(-\,c\,N\,\min\{\eta/d, \eta^2\}\Big),$$

for a universal constant $c > 0$, where we used $\overline{\sigma}^2 \le 6$ and $b = d+1$.

Let $\mathcal{N}$ be a 1/4-net of unit vectors in projective Euclidean distance with $|\mathcal{N}| \le C_0^d$ (Lemma 59). By a union bound,

$$\Pr\left[\sup_{|u\rangle\in\mathcal{N}} \big|\langle u|\big(H_N(\rho)-\rho\big)|u\rangle\big| \;\ge\; \eta\right] \le 2\,C_0^d\,\exp\!\Big(-\,c\,N\,\min\{\eta^2,\; \eta/d\}\Big).$$

Choosing

$$\eta = C\,\max\left\{\frac{d+\log(1/\delta)}{N}, \sqrt{\frac{d+\log(1/\delta)}{N}}\right\}$$

with a sufficiently large universal C_1 makes the right-hand side at most $\delta/2$.

We now pass from the net to all unit vectors. A standard covering argument shows that if $\mathcal{N}$ is a $\frac{1}{4}$–net and X is Hermitian, then

$$\|X\|_\infty \le \frac{1}{1-2\cdot(1/4)}\sup_{|u\rangle\in\mathcal{N}}|\langle u|X|u\rangle| \le 2\sup_{|u\rangle\in\mathcal{N}}|\langle u|X|u\rangle|.$$

Indeed, if $|w\rangle$ is a maximizer of $|\langle w|X|w\rangle|$ and $|u\rangle \in \mathcal{N}$ with $\||w\rangle - |u\rangle\|_2 \le 1/4$, then

$$|\langle w|X|w\rangle| \le |\langle u|X|u\rangle| + 2\,\|X\|_\infty\,\||w\rangle - |u\rangle\|_2 \le |\langle u|X|u\rangle| + \tfrac{1}{2}\|X\|_\infty,$$

which implies $\|X\|_\infty \le 2\sup_{u\in\mathcal{N}}|\langle u|X|u\rangle|$.

Applying this with $X = H_N(\rho) - \rho$, we conclude that, with probability at least $1-\delta$, that

$$\|H_N(\rho)-\rho\|_\infty \le 2\,\eta \le C\,\max\left\{\frac{d+\log(1/\delta)}{N}, \sqrt{\frac{d+\log(1/\delta)}{N}}\right\}.$$

This is the result we claimed, for a suitable universal constant C. □