CHAPTER 17

Lower Bound for Tomography with Single-Copy Measurements

In this lecture, we return to the original question of state tomography and show a separation between single-copy and entangled measurements. Recall that to learn a d-dimensional state ρ to trace distance ϵ, with entangled measurements it suffices to measure $O(d^2/\epsilon^2)$ copies, whereas with single-copy measurements, the protocol we gave in Chapter 4 required $O(d^3/\epsilon^2)$ copies. We will show that in fact this gap is inherent, a result due to [**CHL+23**]:

Theorem 242. *For tomography of d-dimensional states to trace distance ϵ, $\Omega(d^3/\epsilon^2)$ copies are* necessary *for any learning protocol that only uses single-copy measurements.*

The twist in this lecture will be that we cannot simply formulate a suitable distinguishing problem and prove a lower bound via Le Cam's method. The reason is that such distinguishing problems are generally useful for proving lower bounds for *testing* problems where the goal is tell apart whether the state comes from one of two ensembles. In tomography however, the goal is not just to output a single bit about whether the state comes from one ensemble or the other, but instead to output a full description of its density matrix. For such tasks, we need to appeal to a different strategy.

The technique we use in this lecture is somewhat non-standard within the literature on learning lower bounds and is inherently *Bayesian* in flavor. We imagine that the unknown state to be learned is sampled from some *prior* over possible states, and our goal is to show that if one only uses single-copy measurements, then regardless of the learning protocol if one only uses $o(d^3/\epsilon^2)$ copies, then the *posterior* upon observing the outcomes of the various measurements performed does not place very much mass around the true unknown state ρ and it is therefore impossible to deduce meaningful information about ρ from the measurement outcomes.

1. Bayesian Boilerplate

Let μ be some *prior* over density matrices, which we will specify in the next section. Let $\mathcal{T}$ be the learning tree associated to a a learning protocol that performs T single-copy measurements. Given a node u within $\mathcal{T}$, denote by $\mathcal{T}_\rho(u)$ the probability of reaching that node given copy access to ρ. If the POVMs encountered along the path from root to u are $\{M^{u_1}_{s_1}\}_{s_1}, \ldots, \{M^{u_t}_{s_t}\}_{s_t}$ for rank-1 POVM elements

$$M^u_s = w^u_s d \cdot |\psi^u_s\rangle \langle \psi^u_s| \ ,$$

then

$$\mathcal{T}_\rho(u) = d^t \prod_{i=1}^{t} w_{s_i}^{u_i} \langle \psi_{s_i}^{u_i} | \rho | \psi_{s_i}^{u_i} \rangle \,, \tag{82}$$

so by Bayes' rule, conditioned on observing the outcomes along the path from the root to u, the *posterior distribution* on ρ drawn from the prior μ, which we denote by $\nu_u(\rho)$, is given by

$$\nu_u(\rho) \propto \mu(\rho) \prod_{i=1}^{t} w_{s_i}^{u_i} \langle \psi_{s_i}^{u_i} | \rho | \psi_{s_i}^{u_i} \rangle \,.$$

The main technical result of this lecture will be the following result showing that with $o(d^3/\epsilon^2)$ single-copy measurements, the posterior will with high probability place very little mass in a neighborhood around the true state.

Theorem 243 (Posterior anticoncentration)**.** *Let $d = \omega(1)$, let $\epsilon > 0$ be at most some sufficiently small absolute constant, and suppose $T = o(d^3/\epsilon^2)$. There is a prior μ over density matrices, and a subset $S_{\text{good}} \subseteq \text{supp}(\mu)$ for which $\Pr_{\rho\sim\mu}[\rho \in S_{\text{good}}] = 1 - o(1)$, such that for any $\rho_0 \in S_{\text{good}}$, with probability $1 - o(1)$ over the distribution over leaves ℓ given by $\mathcal{T}_{\rho_0}(\ell)$, the posterior anti-concentrates around ρ_0 in the sense that*

$$\nu_\ell(\{\rho : \|\rho - \rho_0\|_{\mathsf{tr}} \le \epsilon\}) = o(1) \,.$$

We will define μ and S_{good} in the next section. Here, we deduce Theorem 242 from Theorem 243:

Proof of Theorem 242. Let $\mathcal{S}$ be the event that $\rho \sim \mu$ lies in S_{good} and $\ell \sim \mathcal{T}_\rho$ satisfies the conditions in Theorem 243. Let $\mathcal{A}$ denote the (possibly randomized) classical post-processing that maps a given leaf ℓ of the learning tree to an estimate of ρ. We will abuse notation and use $\mathcal{A}$ to also denote the internal randomness used by $\mathcal{A}$. We will show that if $T = o(d^3/\epsilon^2)$, then

$$\Pr \mathcal{A}, \rho \sim \mu, \ell \in \mathcal{T}_\rho[\|\mathcal{A}(\ell) - \rho\|_{\mathsf{tr}} \le \epsilon] = o(1) \,.$$

Instead of sampling $\rho \sim \mu$ and then sampling $\ell \sim \mathcal{T}_\rho$, we can equivalently first sample $\rho_0 \sim \mu$, then $\ell \sim \mathcal{T}_{\rho_0}$ and then sample ρ from the posterior ν_ℓ. We can then rewrite the left-hand side of the above as

$$\begin{aligned} &\mathbb{E}_{\mathcal{A},\rho_0\sim\mu,\ell\sim\mathcal{T}_{\rho_0}} \Pr_{\rho\sim\nu_\ell}[\|\mathcal{A}(\ell) - \rho\|_{\mathsf{tr}} \le \epsilon] \\ &\qquad \le \mathbb{E}_{\mathcal{A},\rho_0\sim\mu,\ell\sim\mathcal{T}_{\rho_0}} \Pr_{\rho\sim\nu_\ell}[\|\mathcal{A}(\ell) - \rho\|_{\mathsf{tr}} \le \epsilon \text{ and } (\rho,\ell) \in \mathcal{S}] + o(1) \,. \end{aligned} \tag{83}$$

For any choice of internal randomness for $\mathcal{A}$ and any leaf ℓ, let $\rho_\ell^{\mathcal{A}}$ denote an arbitrary state for which $(\rho_\ell^{\mathcal{A}}, \ell) \in \mathcal{S}$ and $\|\mathcal{A}(\ell) - \rho_\ell^{\mathcal{A}}\|_{\mathsf{tr}} \le \epsilon$, if such a state exists. Denote by $\mathcal{E}$ the event that such a state exists. Then under $\mathcal{E}$, for any state ρ for which $\|\mathcal{A}(\ell) - \rho\|_{\mathsf{tr}} \le \epsilon$, we have $\|\rho_\ell^{\mathcal{A}} - \rho\|_{\mathsf{tr}} \le 2\epsilon$. If $\mathcal{E}$ does not occur for some choice of internal randomness for $\mathcal{A}$ and some ℓ, note that the corresponding inner expectation in Eq. (83) is zero. We can thus upper bound Eq. (83) by

$$\mathbb{E}_{\mathcal{A},\ell|\mathcal{E}} \Pr_{\rho\sim\nu_\ell}[\|\rho_\ell^{\mathcal{A}} - \rho\|_{\mathsf{tr}} \le 2\epsilon \text{ and } (\rho,\ell) \in \mathcal{S}] \le \mathbb{E}_{\mathcal{A},\ell|\mathcal{E}} \Pr_{\rho\sim\nu_\ell}[\|\rho_\ell^{\mathcal{A}} - \rho\|_{\mathsf{tr}} \le 2\epsilon] = o(1) \,,$$

where in the last step we used that for any $(\rho', \ell) \in \mathcal{S}$, the posterior ν_ℓ places $o(1)$ mass on $B_{\mathsf{tr}}(\rho', \epsilon)$ if the number of measurements is $o(d^3/\epsilon^2)$. □

2. Construction of Hard Distribution

In state tomography, the "hardest" distributions are those that are close to maximally mixed. As such, the hard prior μ we construct will consist of perturbations of the maximally mixed state of the form

$$\rho = \frac{1}{d}(\mathrm{Id} + \sigma G)\,, \qquad \sigma = 1/100\,, \tag{84}$$

where G is a sample from an appropriately chosen distribution over perturbations.

Intuitively, we will take G to be a *Gaussian* perturbation, with the slight twist that it must be adjusted to have zero trace, and furthermore we want to "clip" the distribution so that $\|G\|_{\mathsf{op}}$ is not too large.

Definition 244 (Trace-centered GOE ensemble)**.** *The* Gaussian orthogonal ensemble (GOE) *is the distribution over $d \times d$ symmetric matrices where every entry along the diagonal is an independent sample from $\mathcal{N}(0, 2/d)$, every entry strictly above the diagonal is an independent sample from $\mathcal{N}(0, 1/d)$. We denote such a matrix by $G \sim \mathrm{GOE}(d)$.*

The trace-centered GOE *is the distribution given by sampling $G \sim \mathrm{GOE}(d)$ and outputting $G' = G - \mathrm{tr}(G) \cdot \mathrm{Id}/d$. We denote this by $G' \sim \mathrm{GOE}^*(d)$.*

Finally, the clipped, trace-centered GOE *is $\mathrm{GOE}^*(d)$ conditioned on the matrix having operator norm at most 4.*[1] *We denote this distribution by $\mathrm{GOE}'(d)$.*

In Eq. (84), we will take $G \sim \mathrm{GOE}'$. By design, ρ is a valid density matrix, and concretely the density μ for this distribution is given by

$$\mu(\rho) \propto \exp\Big(-\frac{d^3}{4\sigma^2}\|\rho - \mathrm{Id}/d\|_F^2\Big) \cdot \mathbb{1}\{\rho \in S_{\mathrm{supp}}\}\,, \quad S_{\mathrm{supp}} = \Big\{\rho \in U : \|\rho - \mathrm{Id}/d\|_{\mathsf{op}} \le \frac{4\sigma}{d}\Big\}\,,$$

where U is the linear space of trace-1 matrices.

We will further define a set of "good" states in its interior of the support

$$S_{\mathrm{good}} = \Big\{\rho \in U : \|\rho - \mathrm{Id}/d\|_{\mathsf{op}} \le \frac{3\sigma}{d}\Big\}\,,$$

corresponding to the event that $G \sim \mathrm{GOE}'(d)$ satisfies $\|G\|_{\mathsf{op}} \le 3$. By standard bounds from random matrix theory, these good states comprise an overwhelming majority of the mass of μ, satisfying the condition in Theorem 243 on S_{good}.

Fact 245. $\mu(S_{\mathrm{good}}) \ge 1 - e^{-\Omega(d)}$.

One useful consequence of working with states in a neighborhood around the maximally mixed state is that the prior looks relatively "flat" in the sense that the likelihood of seeing any state is not too different from that of seeing any other state:

Lemma 246 (Flatness of prior)**.** *For all ρ, ρ' in the support of μ,*

$$|\log(\mu(\rho)/\mu(\rho'))| \le 4d^2\,.$$

Proof. Note that

$$\log(\mu(\rho)/\mu(\rho')) = \frac{d^3}{4\sigma^2}(\|\rho - \mathrm{Id}/d\|_F^2 - \|\rho' - \mathrm{Id}/d\|_F^2)\,.$$

and each of these squared norms is a number in between 0 and $16\sigma^2/d$ as ρ, ρ' are in an operator norm ball around Id/d of radius $4\sigma/d$. □

[1] The constant is somewhat arbitrary and just needs to be sufficiently large.

3. Anticoncentration of Posterior

3.1. Outline of Argument

Intuitively, as one performs more experiments, the mass of the posterior around the true state ρ_0 will get larger and larger until, in the infinite measurement limit, it is concentrated as a spike around ρ_0. In order to show anti-concentration of the posterior, we would thus like to show that if not enough measurements are performed, then the mass of the posterior in an ϵ-neighborhood around the true state ρ_0 is not elevated significantly beyond the mass of the rest of the support.

Henceforth, let $B_{\mathsf{tr}}(\rho_0, \epsilon)$ denote the trace norm ball of radius ϵ around ρ_0. To bound $\nu_\ell(B_{\mathsf{tr}}(\rho_0, \epsilon))$, we will compare it to the posterior mass on a much larger ball around ρ_0, that is, we will try to upper bound the ratio

$$\frac{\nu_\ell(B_{\mathsf{tr}}(\rho_0, \epsilon))}{\nu_\ell(B_{\mathsf{tr}}(\rho_0, C\epsilon))}.$$

While this seems like a rather daunting quantity to bound, there are two tricks we will exploit: (1) the flatness of the prior of the prior from Lemma 246, and (2) "recentering" the posterior around ρ_0 to exploit the fact that the sets $B_{\mathsf{tr}}(\rho_0, \epsilon)$ and $B_{\mathsf{tr}}(\rho_0, C\epsilon)$ are centered around ρ_0. For (2), by Bayes' rule, we have

$$\frac{\nu_\ell(\rho)}{\nu_\ell(\rho_0)} = \frac{\mathcal{T}_\rho(\ell)}{\mathcal{T}_{\rho_0}(\ell)} \cdot \frac{\mu(\rho)}{\mu(\rho_0)},$$

so for

$$\mathcal{L}_\rho(\ell) \triangleq \frac{\mathcal{T}_\rho(\ell)}{\mathcal{T}_{\rho_0}(\ell)},$$

we can rewrite the posterior density as

$$\nu_\ell(\rho) \propto \mathcal{L}_\rho(\ell) \cdot \mu(\rho).$$

We can thus rewrite the posterior ratio as

$$\frac{\nu_\ell(B_{\mathsf{tr}}(\rho_0, \epsilon))}{\nu_\ell(B_{\mathsf{tr}}(\rho_0, C\epsilon))} = \frac{\int_{B_{\mathrm{tr}}(\rho_0, \epsilon)} \mathcal{L}_\rho(\ell) \cdot \mu(\rho)\, \mathrm{d}\rho}{\int_{B_{\mathrm{tr}}(\rho_0, C\epsilon)} \mathcal{L}_\rho(\ell) \cdot \mu(\rho)\, \mathrm{d}\rho},$$

where $\mathrm{d}\rho$ denotes the Lebesgue measure over symmetric matrices of trace 1. Although μ is some complicated distribution over density matrices, flatness from Lemma 246 allows us to replace all appearances of μ above with the *uniform measure* over S_{supp} at the cost of an $\exp(\Theta(d^2))$ factor, so

$$\frac{\nu_\ell(B_{\mathsf{tr}}(\rho_0, \epsilon))}{\nu_\ell(B_{\mathsf{tr}}(\rho_0, C\epsilon))} \le \exp(4d^2) \cdot \frac{\int_{B_{\mathrm{tr}}(\rho_0, \epsilon)} \mathcal{L}_\rho(\ell) \cdot \mathbb{1}[\rho \in S_{\mathrm{supp}}]\, \mathrm{d}\rho}{\int_{B_{\mathrm{tr}}(\rho_0, C\epsilon)} \mathcal{L}_\rho(\ell) \cdot \mathbb{1}[\rho \in S_{\mathrm{supp}}]\, \mathrm{d}\rho}. \tag{85}$$

The exponential factor might seem like a crippling loss, but it will turn out that the ratio to its right is much smaller, intuitively because as long as the likelihood ratio $\mathcal{L}_\rho(\ell)$ is somewhat diffuse across S_{supp}, then the ratio on the right-hand side is roughly the ratio in volumes between $B_{\mathsf{tr}}(\rho_0, \epsilon)$ and $B_{\mathsf{tr}}(\rho_0, C\epsilon)$. The latter volume ratio is C^{-d^2} and significantly smaller than $\exp(4d^2)$ when C is sufficiently large.[2]

Below, we make this reasoning formal by giving precise bounds on the numerator and denominator on the right-hand side of Eq. (85).

---

[2]Note that this is where we need ϵ to be bounded by a sufficiently small absolute constant, so that $B_{\mathsf{tr}}(\rho_0, C\epsilon)$ does not spill over from the space of valid density matrices.

3.2. Numerator Bound

The numerator is the easier of the two quantities to bound, because we just need an upper bound.

Likelihood ratios always integrate to 1, that is,

$$\mathbb{E}_{\ell\sim\mathcal{T}_{\rho_0}}[\mathcal{L}_\rho(\ell)] = 1\,,$$

so

$$\begin{aligned}\mathbb{E}_{\ell\sim\mathcal{T}_{\rho_0}}\int_{B_{\mathrm{tr}}(\rho_0,\epsilon)}\mathcal{L}_\rho(\ell)\cdot\mathbb{1}[\rho\in S_{\mathrm{supp}}]\,\mathrm{d}\rho &= \int_{B_{\mathrm{tr}}(\rho_0,\epsilon)}\mathbb{1}[\rho\in S_{\mathrm{supp}}]\,\mathrm{d}\rho\\ &= \mathrm{Vol}(B_{\mathsf{tr}}(\rho_0,\epsilon))\\ &= \epsilon^{(d+2)(d-1)/2}\mathrm{Vol}(B_{\mathsf{tr}}(0,1))\,,\end{aligned}$$

where Vol denotes volume under the Lebesgue measure over the linear space of trace-1 symmetric matrices, and in the last step we used that this linear space has dimension $\binom{d+1}{2}-1=(d+2)(d-1)/2$. By Markov's inequality, we immediately deduce the following:

Lemma 247. *With probability at least $1-e^{-d^2}$ over $\ell\sim\mathcal{T}_{\rho_0}$, the numerator in Eq. (85) is upper bounded by*

$$\int_{B_{\mathrm{tr}}(\rho_0,\epsilon)}\mathcal{L}_\rho(\ell)\cdot\mathbb{1}[\rho\in S_{\mathrm{supp}}]\,\mathrm{d}\rho \le e^{d^2}\epsilon^{(d+2)(d-1)/2}\mathrm{Vol}(B_{\mathsf{tr}}(0,1))\,.$$

3.3. Denominator Bound

The denominator bound is trickier because we want a *lower bound*, so it is not enough to simply compute its expectation with respect to $\ell\sim\mathcal{T}_{\rho_0}$. Instead, we have to argue that with high probability over ℓ, $\mathcal{L}_\rho(\ell)$ is not too small.

Likelihood ratio calculation and a thought experiment. Let's write out $\mathcal{L}_\rho(\ell)$ explicitly (interestingly, this will be the only place in the proof where we actually use the form of $\mathcal{T}_\rho$). By Eq. (82),

$$\begin{aligned}\log\mathcal{L}_\rho(\ell) &= \sum_{i=1}^T \log\Big(\frac{\langle\psi^{u_i}_{s_i}|\,\rho\,|\psi^{u_i}_{s_i}\rangle}{\langle\psi^{u_i}_{s_i}|\,\rho_0\,|\psi^{u_i}_{s_i}\rangle}\Big)\\ &= \sum_{i=1}^T \log\Big(1+\frac{\langle\psi^{u_i}_{s_i}|\,(\rho-\rho_0)\,|\psi^{u_i}_{s_i}\rangle}{\langle\psi^{u_i}_{s_i}|\,\rho_0\,|\psi^{u_i}_{s_i}\rangle}\Big)\,. \end{aligned}\tag{86}$$

For $\rho,\rho_0\in S_{\mathrm{supp}}$, note that for any unit vector $|\psi\rangle$ we have

$$-0.1\le -\frac{4\sigma}{1+4\sigma}\le\frac{\langle\psi|\,(\rho-\rho_0)\,|\psi\rangle}{\langle\psi|\,\rho_0\,|\psi\rangle}\le\frac{4\sigma}{1-4\sigma}\le 0.1\,,$$

so using the elementary inequality $\log(1+a)\ge a-\frac{2}{3}a^2$ for $|a|\le 0.1$, we can lower bound Eq. (86) to get

$$\begin{aligned}\log\mathcal{L}_\rho(\ell) &\ge \sum_{i=1}^T \frac{\langle\psi^{u_i}_{s_i}|\,(\rho-\rho_0)\,|\psi^{u_i}_{s_i}\rangle}{\langle\psi^{u_i}_{s_i}|\,\rho_0\,|\psi^{u_i}_{s_i}\rangle}-\frac{2}{3}\frac{\langle\psi^{u_i}_{s_i}|\,(\rho-\rho_0)\,|\psi^{u_i}_{s_i}\rangle^2}{\langle\psi^{u_i}_{s_i}|\,\rho_0\,|\psi^{u_i}_{s_i}\rangle^2}\\ &\ge \sum_{i=1}^T \frac{\langle\psi^{u_i}_{s_i}|\,(\rho-\rho_0)\,|\psi^{u_i}_{s_i}\rangle}{\langle\psi^{u_i}_{s_i}|\,\rho_0\,|\psi^{u_i}_{s_i}\rangle}-d^2\,\langle\psi^{u_i}_{s_i}|\,(\rho-\rho_0)\,|\psi^{u_i}_{s_i}\rangle^2\,,\end{aligned}$$

where in the last step we used that $\frac{2}{3}\|\rho_0\|_{\mathsf{op}}^{-2} \le \frac{2}{3}d^2(1-4\sigma)^{-1} \le d^2$.

Let's now play the following thought experiment. Suppose that instead of being sampled from μ, which is centered around the maximally mixed state, ρ were sampled from a distribution $\mathcal{D}$ over $S_{\rm supp}$ which is *symmetric* around ρ_0, then note that $\mathbb{E}_{\mathcal{D}}[\rho-\rho_0] = 0$, so by Jensen's inequality,

$$\log \mathbb{E}_{\rho\sim\mathcal{D}}\mathcal{L}_\rho(\ell) \ge \mathbb{E}_{\rho\sim\mathcal{D}} \log \mathcal{L}_\rho(\ell) = -d^2 \sum_{i=1}^T \mathbb{E}_{\rho\sim\mathcal{D}} \langle \psi^{u_i}_{s_i}| (\rho-\rho_0) |\psi^{u_i}_{s_i}\rangle^2 \,.$$

Let's push our luck further; in addition to $\mathcal{D}$ being symmetric around ρ_0, let's suppose it is even *rotation-invariant* around ρ_0, that is, the distribution over $\rho-\rho_0$ is unchanged upon conjugating by a random unitary matrix. Then for any unit vector $|\psi\rangle$, by Haar integration and the fact that $\mathrm{tr}(\rho-\rho_0) = 0$, we get

$$d^2\mathbb{E}_{\rho\sim\mathcal{D}} \langle\psi| (\rho-\rho_0) |\psi\rangle^2 = \frac{d^2}{d(d+1)}\mathbb{E}_{\rho\sim\mathcal{D}}\|\rho-\rho_0\|_F^2 \le \mathbb{E}_{\rho\sim\mathcal{D}}\|\rho-\rho_0\|_F^2 \,.$$

As all of the states in $S_{\rm supp}$ have operator norm at most $O(1/d)$, the right-hand side scales as $O(1/d)$. Putting everything together, this would imply that $\mathbb{E}_{\rho\sim\mathcal{D}}\mathcal{L}_\rho(\ell) \ge e^{-O(T/d)}$, and as long as $T \ll d^3$, this would be dominated by $e^{O(d^2)}$ factors from other parts of our analysis. This is the origin of the d^3 dependence in our lower bound. Below, we will make all of this concrete by showing how to get a distribution $\mathcal{D}$ which is symmetric and rotation-invariant around ρ_0.

Subsampling. Consider the following neighborhood of density matrices centered around ρ_0:

$$N^*(\rho_0) \triangleq \{\rho \in B_{\mathsf{tr}}(\rho_0, C\epsilon) : \|\rho-\rho_0\|_{\mathsf{op}} \le C\epsilon/d\} \,.$$

First observe that this neighborhood is still confined to the support of μ:

Proposition 248. *If $\rho_0 \in S_{\rm good}$ and $\epsilon \le \sigma/C$, then $N^*(\rho_0) \in S_{\rm supp}$.*

Proof. We have

$$\|\rho - \mathrm{Id}/d\|_{\mathsf{op}} \le \|\rho-\rho_0\|_{\mathsf{op}} + 3\sigma/d \le 4\sigma/d$$

where the last step holds provided that $C\epsilon \le \sigma$. □

We can thus lower bound the denominator in Eq. (85) by

$$\begin{aligned}
&\int_{B_{\mathsf{tr}}(\rho_0,C\epsilon)} \mathcal{L}_\rho(\ell)\cdot \mathbb{1}[\rho\in S_{\rm supp}]\,\mathrm{d}\rho \\
&\quad\ge \int_{N^*(\rho_0)} \mathcal{L}_\rho(\ell)\,\mathrm{d}\rho \\
&\quad= \mathbb{E}_{\rho\in N^*(\rho_0)}\mathcal{L}_\rho(\ell)\cdot \mathrm{Vol}(N^*(\rho_0)) \\
&\quad= \mathbb{E}_{\rho\in N^*(\rho_0)}\mathcal{L}_\rho(\ell)\cdot (C\epsilon)^{(d+2)(d-1)/2}\cdot \mathrm{Vol}(B_{\mathsf{tr}}(0,1)\cap B_{\mathsf{op}}(0,1/d)),,
\end{aligned}$$

where in the first step we used Proposition 248 to subsample the measure over $B_{\mathsf{tr}}(\rho_0, C\epsilon)$ to only include mass from $N^*(\rho_0)$, and in the third step we again used translation-invariance of the Lebesgue measure and the fact that the linear space of trace-1 symmetric matrices has dimension $(d+2)(d-1)/2$.

We can thus take the distribution $\mathcal{D}$ from the above thought experiment to be the uniform distribution over $N^*(\rho_0)$. This is clearly symmetric and rotation-invariant around ρ_0. Furthermore,

$$\mathbb{E}_{\rho\in N^*(\rho_0)}\|\rho-\rho_0\|_F^2 \le \mathbb{E}_{\rho\in N^*(\rho_0)}\|\rho-\rho_0\|_{\mathsf{tr}}\cdot\|\rho-\rho_0\|_{\mathsf{op}} \le \frac{C^2\epsilon^2}{d}\,,$$

by Holder's inequality and the bounds on the norms of $\rho-\rho_0$ for $\rho\in N^*(\rho_0)$. From the reasoning in the thought experiment, we conclude that

$$\mathcal{L}_\rho(\ell) \ge \exp(-TC^2\epsilon^2/d)\,,$$

so the denominator in Eq. (85) can be lower bounded as follows:

Lemma 249. *For any leaf ℓ, the denominator in Eq. (85) is lower bounded by*

$$\int_{B_{\mathsf{tr}}(\rho_0,C\epsilon)} \mathcal{L}_\rho(\ell)\cdot \mathbb{1}[\rho\in S_{\mathrm{supp}}]\,\mathrm{d}\rho \ge e^{-TC^2\epsilon^2/d}(C\epsilon)^{(d+2)(d-1)/2}\cdot \mathrm{Vol}(B_{\mathsf{tr}}(0,1)\cap B_{\mathsf{op}}(0,1/d))\,.$$

3.4. Putting Everything Together

Proof of Theorem 243. Combining Lemmas 247 and 249 with Eq. (85), we get

$$\frac{\nu_\ell(B_{\mathsf{tr}}(\rho_0,\epsilon))}{\nu_\ell(B_{\mathsf{tr}}(\rho_0,C\epsilon))} \le \exp\Big(5d^2+\frac{TC^2\epsilon^2}{d}-\frac{(d+2)(d-1)}{2}\log C\Big)\cdot\frac{\mathrm{Vol}(B_{\mathsf{tr}}(0,1))}{\mathrm{Vol}(B_{\mathsf{tr}}(0,1)\cap B_{\mathsf{op}}(0,1/d))}\,.$$

If $T\le\frac{d^3}{C^2\epsilon^2}$, then if C is a sufficiently large constant, the $\log C$ term coming from the volume ratio between $B(\rho_0,C\epsilon)$ and $B(\rho_0,\epsilon)$ will dominate the exponent, yielding a prefactor of $\exp(-\Omega(d^2))$ for as large of a constant factor as we would like. It remains to show that

$$\frac{\mathrm{Vol}(B_{\mathsf{tr}}(0,1))}{\mathrm{Vol}(B_{\mathsf{tr}}(0,1)\cap B_{\mathsf{op}}(0,1/d))} \le \exp(O(d^2)) \tag{87}$$

at which point we would be done as it would imply that

$$\nu_\ell(B_{\mathsf{tr}}) \le \frac{\nu_\ell(B_{\mathsf{tr}}(\rho_0,\epsilon))}{\nu_\ell(B_{\mathsf{tr}}(\rho_0,C\epsilon))} \le \exp(-\Omega(d^2)) = o(1)\,.$$

The bound in Eq. (87) can be shown using some basic facts in random matrix theory. For the interested reader, the details can be found in Section 4 of [**CHL+23**]. □