CHAPTER 11

Agnostic Tomography

In the last few lectures we have seen several examples of interesting classes of quantum states which can be learned given access to copies thereof. But there is an elephant in the room. It may be too much to hope for to be given *exact* copies of some state from some nice structured family $\mathcal{F}$, either because the copies have incurred some noise or because the ansatz captured by $\mathcal{F}$ is insufficiently expressive and ultimately just an approximation of reality. At the same time, the algorithms we have covered make use of fine-grained properties of the classes of states in question, and there is a real worry that they may be overtuning to modeling assumptions.

In this lecture, we consider a new solution concept, **agnostic tomography**, that seeks to address some of these issues. The premise is that even if we do not have exact access to copies of a state from $\mathcal{F}$, we can still try to find the *best approximation* to the unknown state by a state from $\mathcal{F}$.

Definition 186 (Agnostic tomography)**.** *Let $\mathcal{F}$ be a class of quantum states admitting efficient classical descriptions.* Agnostic tomography *is the following task: given copies of any unknown state ρ (not necessarily from $\mathcal{F}$), and given parameters $0 < \epsilon, \delta < 1$, output the classical description of a state $\hat{\sigma} \in \mathcal{F}$ for which*

$$F(\hat{\sigma}, \rho) \geq \tau_{\mathcal{F}}(\rho) - \epsilon \quad \textit{where} \quad \tau_{\mathcal{F}}(\rho) \triangleq \max_{\sigma \in \mathcal{F}} F(\sigma, \rho)$$

with probability at least $1 - \delta$.

This is a strict generalization of what we have been doing in previous lectures, which is the special case where $\max_{\sigma \in \mathcal{F}} F(\sigma, \rho) = 1$.

Remark 187. *In this lecture, we will assume that $\tau_{\mathcal{F}}(\rho)$ is exactly known to us. While this might seem overly strong, it can be remedied by a simple "doubling argument" in which we repeatedly guess the value of $\tau_{\mathcal{F}}(\rho)$ at different scales at re-run the algorithm with each of these guesses. The analysis below can be made robust to only knowing a constant-factor approximation of $\tau_{\mathcal{F}}(\rho)$, and we will not belabor such details here.*

1. Sample Complexity

We first note that agnostic tomography is primarily interesting from a *computational* perspective, rather than a *statistical* perspective. To see this, we observe below that sample-efficient shadow tomography already implies a simple, albeit computationally inefficient, algorithm for agnostic tomography.

Definition 188 (Covering number)**.** *Given $\epsilon > 0$ and a set S^* of vectors in $\mathbb{C}^d$, an **ϵ-net** is a discrete set of vectors S such that for any $v^* \in S^*$, there exists $v \in T$ such that $\|v^* - v\| \leq \epsilon$.*

The ***covering number*** $\mathcal{N}_\epsilon(S^*)$ *is the size of the smallest* ϵ*-net for* S^*.

The rule of thumb is that for a family of states $\mathcal{F}$ which is described by p parameters, the covering number scales like $(1/\epsilon)^p$. The following shows that agnostic tomography is very *sample-efficient* for such states.

Proposition 189. *Suppose that* $\mathcal{F}$ *is a family of pure states for which* $\mathcal{N}_\epsilon(\mathcal{F}) \leq M$. *Then there is a (computationally inefficient) algorithm for agnostic tomography for* $\mathcal{F}$ *which only requires* $O((\log M + \log 1/\delta)/\epsilon^2)$ *copies.*

Proof. Let S be an ϵ-net for $\mathcal{F}$ of size at most M. By Lemma 190 below, we can estimate $\langle\phi|\,\rho\,|\phi\rangle$ for all $|\phi\rangle \in S$ to error ϵ with probability $1-\delta$. Suppose this event happens. For any $|\phi^*\rangle \in \mathcal{F}$, there exists $|\phi\rangle \in S$ for which $\|\,|\phi^*\rangle - |\phi\rangle\,\| \leq \epsilon$. We then have

$$\langle\phi^*|\,\rho\,|\phi^*\rangle = \langle\phi|\,\rho\,|\phi\rangle + \langle(\phi^* - \phi)|\,\rho\,|\phi\rangle + \langle\phi^*|\,\rho\,|(\phi^* - \rho)\rangle$$

But

$$|\,\langle(\phi^* - \phi)|\,\rho\,|\phi\rangle\,| \leq \|\phi^* - \phi\| \leq \epsilon\,,$$

so by triangle inequality, $|\,\langle\phi^*|\,\rho\,|\phi^*\rangle - \langle\phi|\,\rho\,|\phi\rangle\,| \leq 2\epsilon$. So if $\langle\phi^*|\,\rho\,|\phi^*\rangle = \tau_{\mathcal{F}}(\rho)$, there is some $|\phi\rangle \in S$ for which $\langle\phi|\,\rho\,|\phi\rangle \geq \tau - 2\epsilon$. By selecting the $|\phi\rangle \in S$ maximizing our estimated value of $\langle\phi|\,\rho\,|\phi\rangle$, we thus obtain $|\phi\rangle \in S$ for which $\langle\phi|\,\rho\,|\phi\rangle \geq \tau_{\mathcal{F}}(\rho) - 4\epsilon$. The proposition follows by replacing ϵ with 4ϵ in the above. □

The above used the following result which is implicit from the lecture on classical shadows.

Lemma 190. *Given a collection of pure states* $\{|\xi_1\rangle, \ldots, |\xi_M\rangle\}$ *with efficient classical descriptions, and given copies of unknown state* ρ, *there is a polynomial-time algorithm for estimating the fidelities* $\{\langle\xi_i|\,\rho\,|\xi_i\rangle\}_i$ *each to within additive error* ϵ *with probability* $1-\delta$, *using* $O((\log M + \log 1/\delta)/\epsilon^2)$ *copies of* ρ.

The sample-efficient algorithm described above gives a nice connection between shadow tomography and agnostic tomography, but unfortunately the algorithm is computationally inefficient: it requires brute-force searching over the ϵ-net S, whose size can be exponentially large in the number of parameters defining $\mathcal{F}$. The goal for the rest of the lecture is to develop a surprisingly simple and general framework for *computationally efficient* agnostic tomography.

2. Stabilizer States

Our algorithm and analysis will adhere to the following template.

- Perform many independent runs of the following recursive procedure:
  - I) **Attempt to learn "non-agnostically"**: In this step we naively run an algorithm that would work if ρ were very close to $\mathcal{F}$
  - II) **If attempt fails, "bootstrap" and return to Step I)**: As we will show, the failure of Step I) will ensure that it is easy to find a measurement for which the post-measurement state ρ' satisfies $\tau_{\mathcal{F}}(\rho') \geq C\tau_{\mathcal{F}}(\rho)$ for an absolute constant $C > 1$. The upshot is that we can now recurse on this post-measurement state, and because the fidelity increases by a constant factor in every recursive step, there can be at most $O(\log 1/\tau)$ recursive steps in total.

- **Hypothesis selection**: Using classical shadows, return the output state among these runs which has highest fidelity with ρ.

This turns out to yield agnostic tomography algorithms for many classes of states. Here we record one such application to stabilizer states, recently obtained by [CGYZ25]. In this setting, the maximum fidelity $\tau_{\mathcal{F}}(\rho)$ has a special name: the **stabilizer fidelity**. This quantity is of particular interest in the study of *quantum resource theories* which aim to quantify the extent to which quantum systems are not classically simulable. Indeed, the smaller the stabilizer fidelity, the further away ρ is from any stabilizer state and thus the harder it is to simulate the amplitudes of ρ using a classical computer.

Theorem 191. *For the class $\mathcal{F}$ of stabilizer states, there is an agnostic tomography algorithm which runs in $f(\tau)\mathrm{poly}(n, 1/\epsilon)$ time and $nf(\tau) + O((1 + \log^2(1/\tau))/\epsilon^2)$ samples for $f(\tau) \triangleq (1/\tau)^{\Theta(\log 1/\tau)}$, where $\tau = \tau_{\mathcal{F}}(\rho)$ is the stabilizer fidelity of the unknown state ρ. The algorithm only performs single-and two-copy measurements.*

Let $|\psi\rangle \in \mathcal{F}$ be a stabilizer state for which $\langle\psi|\,\rho\,|\psi\rangle = \tau_{\mathcal{F}}(\rho)$, and for convenience denote this fidelity by τ.

Recall from the previous lecture that to (non-agnostically) learn stabilizer states, it sufficed to find generators for the unsigned stabilizer group. This is also the case in the agnostic setting. Indeed, if we can find generators for $\mathrm{Weyl}(|\psi\rangle)$ up to phase, then by measuring in their joint eigenbasis we obtain $|\psi\rangle$ as the post-measurement state with probability at least τ, because ρ has fidelity τ with some joint eigenstate of $\mathrm{Weyl}(|\psi\rangle)$. As we saw in the proof of the main theorem from the previous lecture, from this measurement outcome we can also read off the Clifford circuit preparing $|\psi\rangle$.

Next, we describe how to find the generators of $\mathrm{Weyl}(|\psi\rangle)$.

2.1. Collecting High-Correlation Paulis

We will employ Bell difference sampling. We seek to produce a collection of Paulis from $\{I, X, Y, Z\}^{\otimes n}$ with high correlation with ρ – in the non-agnostic case where $\tau = 1$, these would comprise the entire unsigned stabilizer group.

Definition 192. *Given $0 < \epsilon < 1$ and state ρ, a collection of Paulis $F = \{P_1, \dots, P_m\} \subset \{I, X, Y, Z\}^{\otimes n}$ is an ϵ-*high-correlation family *if*

$$\Pr_{\vec{s}\sim\mathcal{B}_\rho}[\mathrm{tr}(P_{\vec{s}}\rho)^2 \geq 0.7 \text{ and } P_{\vec{s}} \notin F] \leq \epsilon\,.$$

*We call a basis of such a family an ϵ-*high-correlation basis*.*

As usual, we will often conflate strings and Paulis and refer to the strings associated to $P_1, \dots, P_m$ as an ϵ-high-correlation family.

Intuitively, a high-correlation family is meant to capture the "bulk" of the Paulis P which one might encounter through Bell difference sampling *for which* $\mathrm{tr}(P\rho)^2$ *is large*. Below, we will use the notation $\mathrm{span}(F)$ to denote the subgroup generated by the Paulis in F.

More precisely, we will use the following algorithm in place of GreedyLocalOpt:

Algorithm 8: CollectPaulis(ρ, ϵ, δ)

```
  Input: Copies of $\rho$, error parameters $0 < \epsilon, \delta < 1$
  Output: $\epsilon$-high-correlation basis of $n$ commuting Paulis
1 Run Bell difference sampling $m = \Theta((n + \log 1/\delta)/\epsilon)$ times to obtain
    strings $\vec{s}_1, \dots, \vec{s}_m$
2 Compute an estimate $\hat{E}_i$ for $\mathrm{tr}(P_{\vec{s}_j}\rho)^2$ for all $i$ to additive error 0.1 using
    Bell sampling on $\Theta(\log(m/\delta))$ more copies of $\rho$
3 Let $F'$ consist of $\vec{s}_i$ for which $\hat{E}_i > 0.6$
4 Compute a basis $F$ for $\mathrm{span}(F')$. Abort if not all strings in $F$ are
    commuting
5 If $|F| < n$, arbitarily pad it to a basis of $n$ commuting Paulis
6 return $F$
```

The following establishes that CollectPaulis produces a high-correlation family with high probability. The intuition is simple: by taking enough Bell difference samples, we cover the bulk of the distribution, and retaining the samples $\vec{s}$ for which $\mathrm{tr}(P_{\vec{s}}\rho)^2$ is estimated to be large will suffice to yield a high-correlation family.

Lemma 193. *With probability at least* $1 - \delta$, CollectPaulis(ρ, ϵ, δ) *outputs an* ϵ*-high-correlation basis of* n *commuting Paulis, using* $O((n + \log 1/\delta)/\epsilon)$ *copies and* $\mathrm{poly}(n, 1/\epsilon, \log 1/\delta)$ *time.*

Proof. Denote by $T \triangleq \{\vec{s} \in \mathbb{F}_2^{2n} : \mathrm{tr}(P_{\vec{s}}\rho)^2 > 0.7\}$ the set of all Paulis with high correlation with ρ, and let S_{high} consist of all Bell difference samples from Line 1 that lie in T. Provided that the estimation in Line 2 succeeds, which happens with probability at least $1 - \delta/3$, S_{high} is contained in F', and by the uncertainty principle (Lemma 194), the algorithm does not abort in Line 4. So if F is the output of the algorithm,

$$\Pr_{\vec{s} \sim \mathcal{B}_\rho}[\vec{s} \in T \text{ and } \vec{s} \notin \mathrm{span}(F)] \le \Pr_{\vec{s} \sim \mathcal{B}_\rho}[\vec{s} \in T \text{ and } \vec{s} \notin \mathrm{span}(S_{\mathsf{high}})],$$

so it suffices to show that the latter quantity is at most ϵ.

Let $p \triangleq \Pr_{\vec{s} \sim \mathcal{B}_\rho}[\vec{s} \in T]$, and let D_{high} denote $\mathcal{B}_\rho$ conditioned on landing in T. If $p \le \epsilon$, we are already done. Otherwise if $p > \epsilon$, by Chernoff bound $|S_{\mathsf{high}}| \ge pm/2$ with probability at least $1 - \delta/3$, because $m \gtrsim \log(1/\delta)/\epsilon$. Every element of S_{high} is an independent sample from D_h, so by Lemma 195 and our choice of m, $\Pr_{\vec{s} \sim D_h}[y \notin \mathrm{span}(S_{\mathsf{high}})] \le \epsilon/p$. The bound on $\Pr_{\vec{s} \sim \mathcal{B}_\rho}[y \notin \mathrm{span}(S_{\mathsf{high}})]$ follows by Bayes' rule. □

The above proof uses the following two helper lemmas:

Lemma 194 (Uncertainty principle)**.** *If distinct* $P, Q \in \{I, X, Y, Z\}^{\otimes n}$ *satisfy* $\mathrm{tr}(P\rho)^2 + \mathrm{tr}(Q\rho)^2 > 1$, *then they must commute.*

Proof. Consider observable $O = \mathrm{tr}(P\rho)P + \mathrm{tr}(Q\rho)Q$. The variance of the observable is $\mathrm{tr}(O^2\rho) - \mathrm{tr}(O\rho)^2$, which is always nonnegative. A direct computation shows that if P, Q anticommute, then $\mathrm{tr}(O^2\rho) = \mathrm{tr}(O\rho) = \mathrm{tr}(P\rho)^2 + \mathrm{tr}(Q\rho)^2$, so the fact that $\mathrm{tr}(O^2\rho) \ge \mathrm{tr}(O\rho)^2$ implies that $\mathrm{tr}(P\rho)^2 + \mathrm{tr}(Q\rho)^2 \le 1$. □

We also need the following classical fact, which simply says that for any distribution over the hypercube, with enough samples their linear sample will occupy most of the mass of the cube.

Lemma 195. *For any distribution D over $\mathcal{F}_2^d$ and i.i.d. samples $x_1, \ldots, x_m \sim D$, if $m \geq 2(\log 1/\delta + d)/\epsilon$, then with probability at least $1 - \delta$ over the samples,*

$$\Pr_{y \sim D}[y \notin \mathrm{span}(x_1, \ldots, x_m)] \leq \epsilon\,.$$

Proof. Let $V_i \triangleq \mathrm{span}(x_1, \ldots, x_i)$, and let $D(V_i)$ denote the probability mass on V_i. Define the indicator variable $I_i = \mathbf{1}[D(V_{i-1}) \geq 1 - \epsilon \text{ or } x_i \notin V_i]$.

Note that for any $x_1, \ldots, x_{i-1}$, $\mathbb{E}[I_i \mid x_1, \ldots, x_{i-1}] \geq \epsilon$. Indeed, either $D(V_{i-1}) \geq 1 - \epsilon$, in which case $I_i = 1$, or $D(V_{i-1}) < 1 - \epsilon$, in which case there is at least an ϵ chance that $x_i \notin V_{i-1}$. Let $J_1, \ldots, J_m$ denote independent Bernoulli random variables with parameter ϵ. Then $\Pr[\sum_{i=1}^m I_i \geq d] \leq \Pr[\sum_{i=1}^m J_i \geq d]$, and by our choice of m and standard binomial tail bounds, this is at most ϵ as claimed.

Provided $\sum_{i=1}^m I_i \geq d$, note that we must have $D(V_m) \geq 1 - \epsilon$. Otherwise, we must have $D(V_i) < 1 - \epsilon$ for all $i = 1, \ldots, m$, meaning $x_i \notin V_{i-1}$. But the dimension of V_i cannot increase by more than d times. □

2.2. Bootstrapping with Bell Difference Sampling

Finally, we need to instantiate the "bootstrapping" step where, provided we do not successfully produce a basis for $\mathrm{Stab}(|\psi\rangle)$ using CollectPaulis, we guess a measurement which will bring ρ closer to $\mathcal{F}$. For this, we simply run Bell difference sampling one more time to obtain a Pauli P and measure with $\{\frac{I+P}{2}, \frac{I-P}{2}\}$.

The utility of this rests upon the following key structural result whose proof we defer to the next section. In a nutshell, it ensures that B_ρ is not too concentrated on any particular proper subspace of the isotropic subspace corresponding to $\mathrm{Stab}(|\psi\rangle)$.

Theorem 196 (Anti-concentration theorem)**.** *Let $W \subsetneq V^*$ be any proper subspace. If ρ has stabilizer fidelity τ, then*

$$\Pr_{\vec{s} \sim \mathcal{B}_\rho}[\vec{s} \in V^* \backslash W] \gtrsim \tau^4\,.$$

Here is the upshot. Because we are assuming CollectPaulis failed to collect a generating set for $\mathrm{Stab}(|\psi\rangle)$, the high-correlation Paulis in $\mathrm{Stab}(|\psi\rangle)$ are mostly concentrated around a *proper subspace* of $\mathrm{Stab}(|\psi\rangle)$. Anti-concentration then ensures that the next Bell difference sample has a non-negligible chance of yielding an element of $\mathrm{Stab}(|\psi\rangle)$ which lives outside of this proper subspace, and thus has low correlation with ρ. We make this formal below:

Lemma 197. *Let F be an ϵ-high-correlation family for $\epsilon = c\tau^4$, where $c > 0$ is a sufficiently small absolute constant. If $\mathrm{span}(F) \neq \mathrm{Weyl}(|\psi\rangle)$, then*

$$\Pr_{\vec{s} \sim \mathcal{B}_\rho, b \in \{\pm 1\}}[\mathrm{tr}(P_{\vec{s}}\rho)^2 \leq 0.7 \text{ and } P_{\vec{s}}\,|\psi\rangle = b\,|\psi\rangle] \gtrsim \tau^4\,.$$

Proof. Define three events on $\vec{s} \sim \mathcal{B}_\rho$:

$$\mathcal{E}_{\mathsf{lowcorr}} \triangleq \mathbf{1}[\mathrm{tr}(P_{\vec{s}}\rho)^2 \leq 0.7]\,, \quad \mathcal{E}_{\mathsf{stab}} \triangleq \mathbf{1}[P_{\vec{s}} \in \mathrm{Weyl}(|\psi\rangle)]\,, \quad \mathcal{E}_{\mathsf{out}} \triangleq \mathbf{1}[P_{\vec{s}} \notin F]\,.$$

Then

$$\begin{aligned} \Pr[\mathcal{E}_{\mathsf{lowcorr}} \cap \mathcal{E}_{\mathsf{stab}}] &\geq \Pr[\mathcal{E}_{\mathsf{lowcorr}} \cap \mathcal{E}_{\mathsf{stab}} \cap \mathcal{E}_{\mathsf{out}}] \\ &= \Pr[\mathcal{E}_{\mathsf{out}} \cap \mathcal{E}_{\mathsf{stab}}] - \Pr[\mathcal{E}_{\mathsf{out}} \cap \mathcal{E}_{\mathsf{stab}} \cap \mathcal{E}^c_{\mathsf{lowcorr}}] \\ &\geq \Pr[\mathcal{E}_{\mathsf{out}} \cap \mathcal{E}_{\mathsf{stab}}] - \Pr[\mathcal{E}_{\mathsf{out}} \cap \mathcal{E}^c_{\mathsf{lowcorr}}]\,. \end{aligned}$$

We can lower bound the first term on the right-hand side via the anti-concentration theorem (Theorem 196), and we can upper bound the second term by the assumption that F is a high-correlation family. Formally we get a bound of

$$\geq \Omega(\tau^4) - \epsilon \gtrsim \tau^4 \,.$$

If this event happens, then with probability 1/2 over $b \in \{\pm 1\}$, we have $P_{\vec{s}} |\psi\rangle = b |\psi\rangle$, concluding the proof. □

At this point we are in a position to repeat the argument that was used in the previous section verbatim. From Lemma 197 we obtain an operator $\frac{I + bP_{\vec{s}}}{2}$ which simultaneously stabilizes the ground truth stabilizer state $|\psi\rangle$ while also having low correlation with the given state ρ. By measuring with the POVM $\{\frac{I+bP_{\vec{s}}}{2}, \frac{I-bP_{\vec{s}}}{2}\}$ and post-selecting on the former outcome, the resulting post-measurement state ρ' satisfies that

$$F(\rho', |\psi\rangle) \geq \Big(\frac{1+\sqrt{0.7}}{2}\Big)^{-1} F(\rho, |\psi\rangle) \geq 1.09 F(\rho, |\psi\rangle) \,,$$

i.e., the stabilizer fidelity has increased by a constant factor. This ensures that the number of recursive rounds is upper bounded by $O(\log 1/\tau)$. Furthermore, in each round in order for the bootstrapping to succeed, we rely on Bell difference sampling to produce a low-correlation element of $\mathsf{Stab}(|\psi\rangle)$ with probability $\Omega(\tau^4)$, so in total the procedure has an $\tau^{O(\log 1/\tau)}$ chance of success. As before, we then need to repeat the procedure $(1/\tau)^{O(\log 1/\tau)}$ times, hence the $f(\tau)$ prefactor in the complexity claimed in Theorem 191. The $\log^2(1/\tau)/\epsilon^2$ term in the complexity comes from the fact that after running the algorithm many times, the number of copies needed to perform hypothesis selection scales with ϵ^2 times log in the number of outputs, and $\log f(\tau) = O(\log^2(1/\tau))$.

2.3. Proof of Anti-Concentration Theorem*

In this section, which can be skipped upon first reading, we prove Theorem 196. Throughout, denote the latter subspace by $V^* \subset \mathbb{F}_2^{2n}$. The proof in the case where ρ is pure is a little simpler, so we provide the proof in this special case and defer the general result to [**CGYZ25**, Theorem 5.5]. The argument below is due to [**GIKL24**].

Let $\rho = |\zeta\rangle\langle\zeta|$. Denote the probability mass function for $\mathcal{B}_\rho$ by $p_\zeta(\cdot)$, and recall that $\mathcal{B}_\rho$ is the convolution of the characteristic distribution which has probability mass function

$$p_\zeta(\vec{s}) \triangleq \frac{1}{2^n} \langle\zeta| P_{\vec{s}} |\zeta\rangle^2 \,.$$

Therefore,

$$q_\zeta(\vec{s}) = \sum_{\vec{t} \in \mathbb{F}_2^{2n}} p_\zeta(\vec{t}) p_\zeta(\vec{s} \oplus \vec{t}) \,.$$

2.3.1. *More symplectic Fourier analysis tools.*

Below we collect some useful definitions and algebraic identities that will be useful in the proof of Theorem 196.

Definition 198 (Symplectic complement)**.** *Given a subspace $W \subseteq \mathbb{F}_2^{2n}$, its* ***symplectic complement*** *is the subspace, denote $W^\perp$, of all $\vec{s} \in \mathbb{F}_2^{2n}$ for which $[\vec{s}, \vec{t}] = 0$ for all $\vec{t} \in W$.*

The symplectic complement has several useful properties reminiscent of the usual orthogonal complement in Euclidean geometry:

(A) $W^\perp$ *is a subspace*
(B) $(W^\perp)^\perp = W$
(C) $\dim(W) + \dim(W^\perp) = n$
(D) If $U \subseteq W$*, then* $W^\perp \subseteq U^\perp$*.*

Definition 199 (Symplectic Fourier transform)**.** *Given a function $f : \mathbb{F}_2^{2n} \to \mathbb{R}$, its* ***symplectic Fourier transform*** *$\hat{f} : \mathbb{F}_2^{2n} \to \mathbb{R}$ is defined by*

$$\hat{f}(\vec{\omega}) = \frac{1}{4^n} \sum_{\vec{s} \in \mathbb{F}_2^{2n}} (-1)^{[\vec{\omega}, \vec{s}]} f(\vec{s})\,.$$

Like the standard Fourier transform, this operation is self-dual up to a constant, so we have the following Fourier inversion identity:

$$f(\vec{s}) = \sum_{\vec{\omega} \in \mathbb{F}_2^{2n}} (-1)^{[\vec{\omega}, \vec{s}]} \hat{f}(\vec{\omega})\,.$$

Lemma 200 (Invariance of characteristic distribution, [GNW21])**.** *For any pure state $|\zeta\rangle$, $p_\zeta(\vec{s}) = 2^n \widehat{p_\zeta}(\vec{s})$ for all $\vec{s} \in \mathbb{F}_2^{2n}$.*

Proof. We have

$$\begin{aligned}
\widehat{p_\zeta}(\vec{s}) &= \frac{1}{2^{2n}} \sum_{\vec{\omega}} (-1)^{[\vec{\omega}, \vec{s}]} \langle\zeta| P_{\vec{\omega}} |\zeta\rangle \langle\zeta| P_{\vec{\omega}} |\zeta\rangle \\
&= \frac{1}{2^{2n}} \sum_{\vec{\omega}} \langle\zeta| P_{\vec{s}} P_{\vec{\omega}} P_{\vec{s}} |\zeta\rangle \langle\zeta| P_{\vec{\omega}} |\zeta\rangle \\
&= \frac{1}{2^n} \langle\zeta| P_{\vec{s}} \Big(\langle\zeta| P_{\vec{s}} |\zeta\rangle \cdot \mathrm{Id} \Big) |\zeta\rangle \\
&= \frac{1}{2^n} \langle\zeta| P_{\vec{s}} |\zeta\rangle^2 = p_\zeta(\vec{s})\,,
\end{aligned}$$

where in the penultimate step we used that $\sum_{\vec{\omega}} P_{\vec{\omega}} M P_{\vec{\omega}} = 2^n \mathrm{tr}(M) \cdot \mathrm{Id}$. □

Lemma 201. *For any subspace $W \subseteq \mathbb{F}_2^{2n}$,*

$$\sum_{\vec{s} \in W} p_\zeta(\vec{s}) = \frac{|W|}{2^n} \sum_{\vec{s} \in W^\perp} p_\zeta(\vec{s})\,.$$

Proof. We have

$$\begin{aligned}\sum_{\vec{s}\in W} p_\zeta(\vec{s}) &= \sum_{\vec{s}\in W}\sum_{\vec{\omega}}(-1)^{[\vec{\omega},\vec{s}]}\widehat{p_\zeta}(\vec{\omega})\\ &= \frac{1}{2^n}\sum_{\vec{s}\in W}\sum_{\vec{\omega}}(-1)^{[\vec{\omega},\vec{s}]}p_\zeta(\vec{\omega})\\ &= \frac{1}{2^n}\sum_{\vec{\omega}}p_\zeta(\vec{\omega})\sum_{\vec{s}\in W}(-1)^{[\vec{\omega},\vec{s}]}\\ &= \frac{|W|}{2^n}\sum_{\vec{\omega}:\vec{\omega}\in W^\perp}p_\zeta(\vec{\omega})\,,\end{aligned}$$

where in the first step we used Fourier inversion, and in the last step we used the fact that if $\vec{\omega}$ commutes with every element of W, then $\sum_{\vec{s}\in W}(-1)^{[\vec{\omega},\vec{s}]} = |W|$, whereas if $\vec{\omega}$ doesn't commute with some element of W, then it commutes with exactly half of the elements of W. □

2.3.2. *Proof of anti-concentration*

As a first step, we show that the characteristic distribution places non-negligible mass on the correct subspace:

Lemma 202. *For any subspace $W \subseteq V^*$, $\sum_{\vec{s}\in W} p_\zeta(\vec{s}) \geq \frac{|W|}{2^n}\tau^2$.*

Proof. If C is the Clifford circuit preparing the stabilizer state $|\psi\rangle$ for which $\tau = |\langle\psi|\zeta\rangle|^2$, we can apply $C^\dagger$ to ζ and assume without loss of generality that $|\psi\rangle = |0^n\rangle$ and thus that $\mathrm{Stab}(|\psi\rangle) = \{I, Z\}^{\otimes n}$.

We first prove the claimed bound for $W = V^*$. We have

$$\begin{aligned}\sum_{\vec{s}\in V^*} p_\zeta(\vec{s}) &= \frac{1}{2^n}\sum_{P\in\{I,Z\}^{\otimes n}}\langle\zeta|\,P\,|\zeta\rangle^2\\ &\geq \frac{1}{4^n}\Big(\sum_{P\in\{I,Z\}^{\otimes n}}\langle\zeta|\,P\,|\zeta\rangle\Big)^2\\ &= |\langle\zeta|0^n\rangle|^4 = \tau^2\,.\end{aligned} \tag{63}$$

Next, for general W, we use Lemma 201 to get

$$\sum_{\vec{s}\in W} p_\zeta(\vec{s}) = \frac{|W|}{2^n}\sum_{\vec{s}\in W^\perp} p_\zeta(\vec{s}) \geq \frac{|W|}{2^n}\sum_{\vec{s}\in V^*} p_\zeta(\vec{s})\,,$$

and the claim then follows from Eq. (63). □

Lemma 203. *For any subspace $W \subseteq V^*$ of dimension $n-1$, $\sum_{\vec{s}\in V^*\backslash W}\langle\zeta|\,P_{\vec{s}}\,|\zeta\rangle \gtrsim 2^n\tau$.*

Proof. As in the previous Lemma, we can assume without loss of generality that $|\psi\rangle = |0^n\rangle$ so that $\mathrm{Stab}(|\psi\rangle) = \{I, Z\}^{\otimes n}$. In this case, $V^* = \{0\}^n \times \mathbb{F}_2^n$. Without loss of generality, we can assume that $W = \{0\}^n \times \{0\} \times \mathbb{F}_2^{n-1}$. In this case,

$$\begin{aligned}\sum_{\vec{s}\in V^*\backslash W} \langle\zeta|\, P_{\vec{s}}\, |\zeta\rangle &= \sum_{P\in\{I,Z\}^{\otimes(n-1)}} \langle\zeta|\, Z\otimes P\, |\zeta\rangle \\ &= 2^{n-1}\, \langle\zeta|\, Z \otimes |0^{n-1}\rangle\, \langle 0^{n-1}|\zeta\rangle \\ &= 2^{n-1}(|a_0|^2 - |a_1|^2)\,,\end{aligned}$$

where a_0 and a_1 are the amplitudes of $|\zeta\rangle$ on $|0^n\rangle$ and $|10^{n-1}\rangle$ respectively. Note that $\tau = \frac{1}{2^n}|\,\langle 0^n|\zeta\rangle\,|^2$, and because $|0^n\rangle$ is the stabilizer state closest to $|\psi\rangle$, $|a_1|^2 \le |a_0|^2$. It thus suffices to show that $|a_1|^2$ is bounded away from $|a_0|^2$ by some constant factor < 1. The intuition is that $|a_1|^2$ cannot be too close to $|a_0|^2$ because at the extreme, if they were equal, then $|\psi\rangle$ would be more aligned with one of $|\pm 0^{n-1}\rangle$ or $|\pm i0^{n-1}\rangle$ than either of $|0^n\rangle$ or $|10^{n-1}\rangle$.

We can prove this formally as follows. Consider the amplitudes of $|\psi\rangle$ at $|\pm 0^{n-1}\rangle$ and $|\pm i0^{n-1}\rangle$, which are given by

$$\left\{\frac{1}{2}|a_0 + ba_1|^2\right\}_{b\in\{\pm1,\pm i\}} = \left\{\frac{1}{2}(|a_0|^2 + |a_1|^2 + 2\mathrm{Re}(a_0\cdot\overline{ba_1}))\right\}_{b\in\{\pm1,\pm i\}}.$$

Without loss of generality suppose a_0 is real and nonnegative and $a_1 = |a_1|e^{i\theta}$ for $0 \le \theta\pi/2$, in which case $\max_b \mathrm{Re}(a_0\cdot\overline{ba_1}) = a_0|a_1|C_\theta$ for $C_\theta \triangleq \max(\cos\theta, \sin\theta)$.

For a given θ, the constraint

$$|a_0|^2 \ge \frac{1}{2}(|a_0|^2 + |a_1|^2 + a_0|a_1|C_\theta)^2$$

is weakest when $\theta = \pi/4$, which can be solved to yield $|a_1| \le \sqrt{2-\sqrt{3}}|a_0|$. □

We are finally ready to prove the anti-concentration theorem for pure states:

Proof of Theorem 196. We can assume without loss of generality that $\dim(W) = n-1$. By the definition of $\mathcal{B}_\rho$, we have

$$\begin{aligned}\Pr_{\vec{s}\sim\mathcal{B}_\rho}[\vec{s}\in V^*\backslash W] &= \sum_{\vec{s}\in V^*\backslash W}\ \sum_{\vec{t}\in\mathbb{F}_2^{2n}} p_\zeta(\vec{t})p_\zeta(\vec{s}\oplus\vec{t}) \\ &\ge \sum_{\vec{t}\in W} p_\zeta(\vec{t}) \sum_{\vec{s}\in V^*\backslash W} p_\zeta(\vec{s}\oplus\vec{t}) \\ &= \Big(\sum_{\vec{t}\in W} p_\zeta(\vec{t})\Big)\cdot\Big(\sum_{\vec{s}\in V^*\backslash W} p_\zeta(\vec{s}\oplus\vec{t})\Big) \\ &\ge \frac{\tau^2}{2}\cdot\frac{1}{2^{n-1}}\Big(\frac{1}{2^{n/2}}\sum_{\vec{s}\in V^*\backslash W}|\,\langle\zeta|\,P_{\vec{s}\oplus\vec{t}}|\zeta\rangle\,|\Big)^2 \\ &\gtrsim \tau^4\,,\end{aligned}$$

where in the third step we used the fact that the affine subspace $V^*\backslash W$ is invariant under translation by W, that is, for every $\vec{s}\in W$, $\{\vec{s}\oplus\vec{t} : \vec{s}\in V^*\backslash W\} = V^*\backslash W$, in the fourth step we used Lemma 202 and the fact that $|W| = 2^{n-1}$, and in the last step we used Lemma 203. □