CHAPTER 14

Learning Trees

Suppose you are an experimental physicist (and if you already are, then it will not be difficult to imagine). In your lab you may have a condensed matter system, an array of atoms, a vat of chemicals, or some other type of quantum system. Your experimental system of interest is partially uncharacterized, or rather has features at least partially unknown to you, and your goal is to learn those features through the process of experiment. Accordingly, you prepare your quantum system and perhaps let it evolve, and then make interventions or measurements that decohere the system. You perhaps make a series of measurements, each possibly contingent on the outcome of the previous one, and then do follow-up experiments. Your data is classical, and you analyze the results and write up a research paper (which is necessarily written in terms of classical data) that is immortalized by Science or Nature, or better yet immortalized by science or nature. This narrative serves to emphasize that experiments take the form of a *learning problem*, where the experimental system serves as a type of oracle.

In the future, experimentalists may have a quantum computer in their lab, which they can coherently couple to their experimental system of interest. This will enable them to directly manipulate the quantum information inherent in the physical system, and perform quantum computation on the corresponding quantum data. Would such a **quantum-enhanced experiment** allow them to access aspects of the natural world which are otherwise inaccessible with conventional experiments? Remarkably, the answer is yes. That is, quantum-enhanced experiments can, in principle, allow us access to features of the natural world which would be exponentially difficult to obtain with conventional experiments.

The technical foundations for this endeavor were primarily developed in the three papers [**BCL20**, **HKP21**, **ACQ22**], following which some of the authors joined forces to write [**CCHL22**] which consolidated and substantively generalized the previous work. At a conceptual level, the papers [**HKP21**, **ACQ22**] (and in particular [**ACQ22**]) built an abstract theory of (quantum) experiments, and emphasized as well as formalized how experiments take the form of learning theory problems.

We note that the kinds of problems which arise in quantum learning for experiments are ones with quantum input (i.e. the experimental state, its dynamics, etc.) and classical output (the 'findings' of the experiment). This is in contrast to more standard quantum algorithms with classical input and classical output. In some sense, the former may be more natural for quantum computers than the latter.

The subject of quantum learning theory for quantum experiments has two main parts. Given an experimental task, we would like to show both that (i) there is a quantum-enhanced experimental protocol which renders the task easy, and (ii) for any possible conventional experimental protocol, the task is very hard,

either in terms of query complexity or computational complexity. Achieving (i) is comparatively easier, since it only requires demonstrating a single efficient quantum protocol. On the other hand, (ii) is trickier, since it requires ruling out that *any* conventional experimental protocol is efficient. As such, much of our effort will be devoted to building tools for (ii), corresponding to lower bounds on efficiency.

1. Property testing and purity testing

We begin by studying a class of experimental tasks called **property testing**, and study the special case of **purity testing** in particular. We will take $\mathcal{H} \simeq \mathbb{C}^d$ where $d = 2^n$ to be the Hilbert space of n qubits, as usual. One formulation of property testing problem for quantum states is as follows.

Definition 216 (Property testing for quantum states)**.** *Suppose there are k probability distributions over quantum states, $\mu_1, \dots, \mu_k$, corresponding to k properties. An index $j \in [k]$ is chosen (unknown to the experimenter), and a state $\rho \leftarrow \mu_j$ is sampled. Given access to copies of ρ, the task is to determine the set of indices*

$$S := \{i \in [k] : \rho \text{ lies in the support of } \mu_i\}.$$

Note that there may be multiple such indices i if the supports of the μ_i overlap.

As a concrete example, we consider the following purity testing problem.

Definition 217 (Purity testing)**.** *The purity testing problem is the special case of property testing for quantum states with two distributions: μ_1 supported only on the maximally mixed state $\mathbb{1}/d$, and μ_2 given by the Haar measure over pure states.*

In other words, we are given copies of an unknown state ρ with the promise that either $\rho = \mathbb{1}/d$ (the "mixed" case, $\rho \leftarrow \mu_1$) or $\rho = |\psi\rangle\langle\psi|$ for some pure state $|\psi\rangle$ drawn from the Haar measure (the "pure" case, $\rho \leftarrow \mu_2$), and we must decide which is the case.

The quantum algorithm for the purity testing problem is very simple: its basic primitive uses only two copies of the state ρ together with the *SWAP test*.

Theorem 218. *There is a quantum-enhanced experiment which solves the purity testing problem with success probability at least $2/3$ using $O(1)$ copies of ρ. Each run of the experiment uses only two copies of ρ and a measurement of the swap operator SWAP on $\mathcal{H}^{\otimes 2}$.*

Proof. Let SWAP be the unitary operator on $\mathcal{H}^{\otimes 2}$ defined by

$$\mathsf{SWAP}\left(|\phi\rangle \otimes |\psi\rangle\right) = |\psi\rangle \otimes |\phi\rangle \quad \text{for all } |\phi\rangle, |\psi\rangle \in \mathcal{H}.$$

The eigenvalues of SWAP are ± 1, with corresponding projectors

$$\Pi_{\text{sym}} = \frac{\mathbb{1} + \mathsf{SWAP}}{2}, \qquad \Pi_{\text{asym}} = \frac{\mathbb{1} - \mathsf{SWAP}}{2},$$

onto the symmetric and antisymmetric subspaces of $\mathcal{H}^{\otimes 2}$, respectively.

The SWAP test on two copies of a state ρ is simply the projective measurement $\{\Pi_{\text{sym}}, \Pi_{\text{asym}}\}$ on $\rho^{\otimes 2}$. (Equivalently, it can be implemented by the standard circuit with an ancilla qubit, a controlled-SWAP, and two Hadamards on the ancilla.)

For any state ρ we have

$$\begin{aligned}\Pr[\text{antisymmetric outcome}] &= \operatorname{tr}\big(\Pi_{\text{asym}}\,\rho^{\otimes 2}\big) = \frac{1}{2}\operatorname{tr}\big((\mathbb{1} - \mathsf{SWAP})\,\rho^{\otimes 2}\big) \\ &= \frac{1}{2}\Big(\operatorname{tr}(\rho^{\otimes 2}) - \operatorname{tr}\big(\mathsf{SWAP}\,\rho^{\otimes 2}\big)\Big).\end{aligned}$$

Using $\operatorname{tr}(\rho^{\otimes 2}) = \operatorname{tr}(\rho)^2 = 1$ and the identity

$$\operatorname{tr}\big(\mathsf{SWAP}(A \otimes B)\big) = \operatorname{tr}(AB) \quad \text{for all operators } A, B,$$

we obtain

$$\operatorname{tr}\big(\mathsf{SWAP}\,\rho^{\otimes 2}\big) = \operatorname{tr}(\rho^2),$$

and hence

$$\Pr[\text{antisymmetric outcome}] = \frac{1 - \operatorname{tr}(\rho^2)}{2}\,.$$

The quantity $\operatorname{tr}(\rho^2)$ is the **purity** of ρ.

We now evaluate this probability in our two promised cases. If $\rho = |\psi\rangle\langle\psi|$ is pure, then $\operatorname{tr}(\rho^2) = 1$, so

$$\Pr[\text{antisymmetric outcome} \mid \rho \text{ pure}] = \frac{1-1}{2} = 0.$$

Thus $\rho^{\otimes 2}$ always lies in the symmetric subspace, and the SWAP test never produces the antisymmetric outcome.

If instead $\rho = \mathbb{1}/d$, then

$$\operatorname{tr}(\rho^2) = \operatorname{tr}\left(\frac{\mathbb{1}}{d^2}\right) = \frac{d}{d^2} = \frac{1}{d},$$

and therefore

$$\Pr[\text{antisymmetric outcome} \mid \rho = \mathbb{1}/d] = \frac{1 - 1/d}{2} = \frac{d-1}{2d}.$$

In particular, since $d \geq 2$, we have the uniform lower bound

$$\Pr[\text{antisymmetric outcome} \mid \rho = \mathbb{1}/d] \geq \frac{1}{4}.$$

Now consider the following procedure. On each run, we take two fresh copies of ρ, perform the SWAP test, and record whether the antisymmetric outcome occurred. After T independent runs, we output `mixed` if we ever saw the antisymmetric outcome, and `pure` otherwise.

If ρ is pure, the antisymmetric outcome never occurs, so the procedure always outputs "pure'; the error probability in this case is 0 for every T. On the other hand if $\rho = \mathbb{1}/d$, each run independently produces the antisymmetric outcome with probability at least $1/4$. Hence the probability that we *never* see the antisymmetric outcome in T runs is at most

$$\Pr[\text{error} \mid \rho = \mathbb{1}/d] = \Pr[\text{no antisymmetric outcome in } T \text{ runs}] \leq \left(1 - \frac{1}{4}\right)^T = \left(\frac{3}{4}\right)^T.$$

Taking, for example, $T = 4$ we obtain

$$\Pr[\text{error} \mid \rho = \mathbb{1}/d] \leq \left(\frac{3}{4}\right)^4 = \frac{81}{256} < \frac{1}{3}.$$

Thus for $T = 4$ runs the success probability is at least $2/3$ in both of the promised cases, and the total number of copies of ρ used is $2T = 8 = O(1)$. This completes the proof. □

Next we will embark on a mathematical journey to establish that for any conventional experiment, solving the purity testing problem requires a number of copies of ρ which is *exponential in* n to solve. For this we need to define what we mean by a conventional experiment, and develop some mathematical technology. We will do so below.

2. Conventional experiments and their learning trees

What can 'conventional' experiments do? At a high level, they can prepare a single experimental sample, measure it, and then re-prepare the state and perform subsequent measurements (possibly chosen to be contingent on the previous measurement outcomes). First we recall the most general type of measurement one can make on a system. Given a state ρ, we can measure it with a POVM which is a set of operators $\{F_i\}_i$ satisfying $F_i \succeq 0$ and $\sum_i F_i = \mathbb{1}$ where the probability of measuring the ith outcome is $\text{Prob}[i] = \text{tr}(F_i\rho)$.

A general POVM can have the F_i's be any rank. However, we can always *refine* a POVM so that the F_i's are rank-1. Specifically, consider a POVM $\{F_i\}_i$. Since each F_i is positive semi-definite, we can decompose it as

$$F_i = d\sum_j a_{ij}|\phi_{ij}\rangle\langle\phi_{ij}|$$

where the $a_{ij} \geq 0$. The factor of d is useful since with this convention $\sum_{i,j} a_{ij} = 1$. The refined POVM is then $\{d\, a_{ij}|\phi_{ij}\rangle\langle\phi_{ij}|\}_{i,j}$, and it captures the same information as the original POVM since

$$\sum_j \text{Prob}[(i,j)] = \text{tr}(F_i\rho)\,.$$

As such, without loss of generality, we will consider only rank-1 POVMs henceforth. We will write such POVMs as $\{d\, a_i|\phi_i\rangle\langle\phi_i|\}_i$, namely with only a single index i.

With the above notation at hand, a conventional experiment for measuring copies of a state ρ proceeds as follows:

Step 1: Obtain a copy of ρ, measure it using a rank-1 POVM $\{d\, a_i|\phi_i\rangle\langle\phi_i|\}_i$, and measure the outcome $i = q$ which is stored in a classical memory.

Step 2: Obtain a copy of ρ, measure it using a rank-1 POVM $\{d\, a_{q,i}|\phi_{q,i}\rangle\langle\phi_{q,i}|\}_i$ which may be contingent on the previous measurement outcome, and measure the outcome $i = r$ which is stored in a classical memory.

Step 3: Obtain a copy of ρ, measure it using a rank-1 POVM $\{d\, a_{q,r,i}|\phi_{q,r,i}\rangle\langle\phi_{q,r,i}|\}_i$ which may be contingent on the previous measurement outcomes, and measure the outcome $i = s$ which is stored in a classical memory.

$$\vdots$$

And so on, say a total of T times. We note that calling such a protocol a 'conventional experiment' is perhaps overly generous, since the POVMs are unrestricted

and could be highly complex (i.e. in a way that conventional experiments cannot capture). Therefore, when we prove that all possible experiments of the above kind require e.g. exponentially many measurements to solve the purity testing problem, this is a rather strong result that allows for the possibility of highly complex measurements, so long as they are single-copy measurements. To this end, we sometimes refer to this (strong) model of conventional experiments as a **single-copy access model**.

The above type of data can be organized into a so-called **learning tree**, which we define below.

Definition 219 (Learning tree for quantum states)**.** *Fix an unknown state ρ on $\mathcal{H} \simeq \mathbb{C}^d$. A* learning tree *for ρ with T single-copy measurements is a rooted tree $\mathcal{T}$ of depth T, whose nodes represent possible classical memory states of a conventional experiment, and which satisfies:*

- *Each node v of $\mathcal{T}$ is associated with a probability $p^\rho_{\mathcal{T}}(v)$.*
- *For the root r of the tree, $p^\rho_{\mathcal{T}}(r) = 1$.*
- *For every non-leaf node u, we fix a rank-1 POVM on $\mathcal{H}$ of the form*

$$\left\{d\, a_v\, |\phi_v\rangle\langle\phi_v|\right\}_{v\in \mathrm{child}(u)}, \quad a_v \geq 0, \qquad \sum_{v\in\mathrm{child}(u)} d\, a_v\, |\phi_v\rangle\langle\phi_v| = \mathbb{1},$$

  where child(u) *denotes the set of children of u. Measuring a fresh copy of ρ with this POVM produces an outcome corresponding to some child $v \in$ child(u).*
- *If v is a child of u, then the probability of moving from u to v when the underlying state is ρ is*

$$p^\rho_{\mathcal{T}}(v) = p^\rho_{\mathcal{T}}(u) d\, a_v\, \langle\phi_v|\rho|\phi_v\rangle.$$

- *Every root-to-leaf path has length T (equivalently, the experiment performs exactly T single-copy measurements). For a leaf node ℓ, the quantity $p^\rho_{\mathcal{T}}(\ell)$ is the probability that after T measurements the classical memory is in state ℓ.*

With this definition, let $v_0, v_1, ..., v_T$ be a root-to-leaf path through the tree. We will let $r = v_0$ denote the root, and $\ell = v_T$ denote the leaf. A key feature of trees is that a leaf *defines* a unique root-to-leaf path through the tree, and so a choice of ℓ specifies the entire sequence $v_0, v_1, ..., v_T = \ell$. If we run the experiment corresponding to a particular learning tree $\mathcal{T}$, then the probability that we traverse a particular root-to-leaf path is given by

$$p^\rho_{\mathcal{T}}(\ell) = \prod_{t=1}^{T} d\, a_{v_t} \langle\phi_{v_t}|\rho|\phi_{v_t}\rangle\,.$$

Now basic idea is as follows. Suppose we have two states ρ and σ such that $p^\rho_{\mathcal{T}} \approx p^\sigma_{\mathcal{T}}$ for some appropriate sense of '$\approx$'. This would mean that we could not distinguish ρ from σ in any conventional experiment with T total (possibly adaptive) measurements. We will show that the correct notion of '$\approx$' is captured by the **total variation distance**, which we define below. After exploring some properties of this distance, we will explain why it is the 'right' distance for our purposes.

Definition 220 (Total variation distance)**.** *If p, q are two probability distributions, then the total variation distance between them is*

$$d_{\mathrm{TV}}(p,q) = \frac{1}{2}\sum_i |p_i - q_i| \,.$$

We will frequently use the equivalent formula proved in the following lemma.

Lemma 221. *Let p, q be probability distributions, and define $A := \{i : p_i \geq q_i\}$ so that accordingly $A^c := \{i : p_i < q_i\}$. Then*

$$d_{\mathrm{TV}}(p,q) = \sum_{i\in A}(p_i - q_i) = p_A - q_A \,,$$

where $p_A := \sum_{i\in A} p_i$ and $q_A := \sum_{i\in A} q_i$.

Proof. We have

$$0 = \left(\sum_i p_i\right) - \left(\sum_i q_i\right) = \sum_i (p_i - q_i) = \sum_{i\in A}(p_i - q_i) + \sum_{i\in A^c}(p_i - q_i) \,.$$

Rearranging the above, we find

$$\sum_{i\in A^c}(q_i - p_i) = \sum_{i\in A}(p_i - q_i) \,.$$

On the other hand,

$$\begin{aligned}\sum_i |p_i - q_i| &= \sum_{i\in A}(p_i - q_i) + \sum_{i\in A^c}(q_i - p_i) \\ &= \sum_{i\in A}(p_i - q_i) + \sum_{i\in A}(p_i - q_i) = 2\sum_{i\in A}(p_i - q_i) \,,\end{aligned}$$

and therefore

$$d_{\mathrm{TV}}(p,q) = \frac{1}{2}\sum_i |p_i - q_i| = \sum_{i\in A}(p_i - q_i) = p_A - q_A \,.$$

This proves the claim. □

We give one additional, equivalent formulation of the total variation distance, which we will also make use of:

Lemma 222. *Let $p_S := \sum_{i\in S} p_i$ and similarly $q_S := \sum_{i\in S} q_i$. Then*

$$d_{\mathrm{TV}}(p,q) = \sup_S |p_S - q_S| \,.$$

Proof. Let A be as in Lemma 221. For the particular choice $S = A$ we have

$$|p_S - q_S| = |p_A - q_A| = d_{\mathrm{TV}}(p,q) \,,$$

so that

$$\sup_S |p_S - q_S| \geq d_{\mathrm{TV}}(p,q).$$

Conversely, let $S \subseteq \{i\}$ be arbitrary. Decompose

$$\begin{aligned}p_S - q_S = \sum_{i\in S}(p_i - q_i) &= \sum_{i\in S\cap A}(p_i - q_i) - \sum_{i\in S\cap A^c}(q_i - p_i) \\ &\leq \sum_{i\in S\cap A}(p_i - q_i) \leq \sum_{i\in A}(p_i - q_i) = d_{\mathrm{TV}}(p,q),\end{aligned}$$

where the last equality uses Lemma 221. Interchanging the roles of p and q yields

$$q_S - p_S \leq d_{\mathrm{TV}}(p, q),$$

and hence

$$|p_S - q_S| \leq d_{\mathrm{TV}}(p, q)$$

for every S. Taking the supremum over S gives

$$\sup_S |p_S - q_S| \leq d_{\mathrm{TV}}(p, q),$$

which, together with the reverse inequality, proves the lemma. □

Now let us show how the total variation distance helps us calibrate the difficulty of property testing problems for quantum states. We have the following beautifully simple lemmas:

Lemma 223 (Le Cam's two-point method; see e.g. [Yu97])**.** *Let p, q be two probability distributions on a finite set Ω. Suppose we are given a single sample $X \in \Omega$ which is drawn from p with probability $1/2$ and from q with probability $1/2$. For any (possibly randomized) decision rule $\mathcal{A} : \Omega \to \{0, 1\}$ that outputs a guess for which distribution was used, the success probability satisfies*

$$\Pr[\mathcal{A}\text{correct}] \leq \frac{1}{2} + \frac{1}{2} d_{\mathrm{TV}}(p, q) = \frac{1}{2} + \frac{1}{4} \sum_{i \in \Omega} |p_i - q_i| .$$

Moreover, this bound is tight: there exists a decision rule that achieves equality.

Proof. Fix a decision rule $\mathcal{A}$ and let $S \subseteq \Omega$ be the set of outcomes on which $\mathcal{A}$ guesses that the sample came from p:

$$S := \{i \in \Omega : \mathcal{A}(i) = 0\}.$$

(Thus on S^c the rule guesses that the sample came from q.) When the true distribution is p, the rule is correct with probability p_S; when the true distribution is q, it is correct with probability $q_{S^c} = 1 - q_S$. Since each case occurs with prior probability $1/2$, the overall success probability is

$$\begin{aligned} \Pr[\mathcal{A} \text{ correct}] &= \frac{1}{2} p_S + \frac{1}{2} q_{S^c} = \frac{1}{2} p_S + \frac{1}{2} (1 - q_S) \\ &= \frac{1}{2} + \frac{1}{2}(p_S - q_S). \end{aligned}$$

Using the previous lemma we have $|p_S - q_S| \leq d_{\mathrm{TV}}(p, q)$, and hence

$$\Pr[\mathcal{A} \text{ correct}] \leq \frac{1}{2} + \frac{1}{2} d_{\mathrm{TV}}(p, q).$$

To see that this bound is tight, choose a subset $S^\star \subseteq \Omega$ that attains the supremum in the characterization

$$d_{\mathrm{TV}}(p, q) = \sup_S |p_S - q_S|.$$

Define $\mathcal{A}$ so that it guesses p on $S^\star$ and q on $(S^\star)^c$. For this rule,

$$\Pr[\mathcal{A} \text{ correct}] = \frac{1}{2} + \frac{1}{2}(p_{S^\star} - q_{S^\star}) = \frac{1}{2} + \frac{1}{2} d_{\mathrm{TV}}(p, q),$$

as claimed. □

Lemma 224. *Fix a learning tree $\mathcal{T}$ and a property testing instance with probability distributions $\mu_1, \dots, \mu_k$ over quantum states on $\mathcal{H}$. For each $j \in [k]$ let*

$$p_j(\ell) := \mathbb{E}_{\rho\sim\mu_j}\left[p^\rho_{\mathcal{T}}(\ell)\right]$$

denote the induced probability distribution over leaves ℓ of $\mathcal{T}$ when the unknown state is sampled from μ_j. Fix two indices $i, j \in [k]$, and suppose we are promised that the unknown index is either i or j, each with prior probability $1/2$. Then any conventional experiment described by $\mathcal{T}$ that attempts to decide whether the index is i or j has success probability at most

$$\frac{1}{2} + \frac{1}{2} d_{\mathrm{TV}}(p_i, p_j) = \frac{1}{2} + \frac{1}{4}\sum_\ell \big|p_i(\ell) - p_j(\ell)\big|.$$

In particular, if $d_{\mathrm{TV}}(p_i, p_j) \le \delta$ for some $\delta \in [0, 1]$, then no such experiment can distinguish between the two hypotheses with success probability greater than $1/2 + \delta/2$.

Proof. Consider the following experiment. First an index $b \in \{i, j\}$ is chosen uniformly at random. Conditioned on b, a state ρ is sampled from μ_b, and the learning tree $\mathcal{T}$ is executed, producing a leaf ℓ. By definition of $p_b(\ell)$, the distribution of ℓ conditioned on b is exactly p_b.

Any decision rule that, upon observing ℓ, outputs a guess for whether $b = i$ or $b = j$ is therefore just a binary hypothesis test between the classical distributions p_i and p_j. Applying Le Cam's two-point method with $p = p_i$ and $q = p_j$ yields

$$\Pr[\text{correct guess}] \le \frac{1}{2} + \frac{1}{2} d_{\mathrm{TV}}(p_i, p_j),$$

which is precisely the claimed bound. □

The above lemmas tell us that to get a lower bound on the number of measurements T we require to solve a property testing problem, then we need to upper bound

$$d_{\mathrm{TV}}\Big(\mathbb{E}_{\rho\sim\mu_i} p^\rho_{\mathcal{T}}(\ell),\, \mathbb{E}_{\rho\sim\mu_j} p^\rho_{\mathcal{T}}(\ell)\Big).$$

Such a bound may seem tricky, but we can make our lives easier with the following useful lemma:

Lemma 225. *Suppose $p_i > 0$ for all i, and that the **likelihood ratio** $\frac{q_i}{p_i}$ satisfies $\frac{q_i}{p_i} \ge 1 - c$ for all i and for some constant $c \in [0, 1]$. Then $d_{\mathrm{TV}}(p, q) \le c$.*

Proof. Let $A = \{i : p_i \ge q_i\}$ as in Lemma 221. Using that lemma together with $\frac{q_i}{p_i} \ge 1 - c$, we have

$$d_{\mathrm{TV}}(p, q) = \sum_{i\in A}(p_i - q_i) = \sum_{i\in A} p_i\left(1 - \frac{q_i}{p_i}\right) \le c\sum_{i\in A} p_i \le c\sum_i p_i = c\,.$$

□

Our desired corollary of the above lemma is:

Corollary 226. *Let $\mathcal{T}$ be a learning tree of depth T. Consider a many-versus-one distinguishing task between a fixed "null" state σ and an alternative ensemble μ over states on $\mathcal{H}$. For each leaf ℓ of $\mathcal{T}$ define*

$$p_{\text{null}}(\ell) := p_{\mathcal{T}}^{\sigma}(\ell),$$
$$p_{\text{alt}}(\ell) := \mathbb{E}_{\rho\sim\mu}\big[p_{\mathcal{T}}^{\rho}(\ell)\big].$$

Assume $p_{\text{null}}(\ell) > 0$ for all leaves ℓ. If there exists $\delta \in [0,1]$ such that

$$\frac{p_{\text{alt}}(\ell)}{p_{\text{null}}(\ell)} \geq 1-\delta \quad \textit{for all leaves } \ell,$$

then

$$d_{\text{TV}}(p_{\text{null}}, p_{\text{alt}}) \leq \delta.$$

Consequently, by Le Cam's two-point method, any conventional experiment described by $\mathcal{T}$ that tries to distinguish the null hypothesis $\rho = \sigma$ from the alternative hypothesis $\rho \sim \mu$ has success probability at most

$$\frac{1}{2} + \frac{\delta}{2}.$$

Proof. Apply the likelihood-ratio lemma with

$$p_i = p_{\text{null}}(\ell), \quad q_i = p_{\text{alt}}(\ell),$$

viewing the index i simply as labeling the leaves ℓ. Our hypothesis exactly states that

$$\frac{q_i}{p_i} = \frac{p_{\text{alt}}(\ell)}{p_{\text{null}}(\ell)} \geq 1-\delta \quad \text{for all } i,$$

with $p_i > 0$ by assumption. The lemma then gives

$$d_{\text{TV}}(p_{\text{null}}, p_{\text{alt}}) \leq \delta.$$

Substituting this bound into Le Cam's inequality from above yields

$$\Pr[\text{correct}] \leq \frac{1}{2} + \frac{1}{2} d_{\text{TV}}(p_{\text{null}}, p_{\text{alt}}) \leq \frac{1}{2} + \frac{\delta}{2},$$

which completes the proof. □

This corollary puts us in a good position to get an exponential lower bound for the purity testing problem, which we pursue in the next section below.

3. Exponential lower bounds for purity testing with conventional experiments

We now leverage the learning tree formalism to prove an exponential lower bound on purity testing with conventional experiments. Throughout this section we recall that $\mathcal{H} \simeq \mathbb{C}^d$ with $d = 2^n$.

Theorem 227 (Purity testing lower bound for conventional experiments)**.** *Consider the purity testing problem with null hypothesis $\rho = \mathbb{1}/d$ and alternative hypothesis $\rho = |\psi\rangle\langle\psi|$ for a Haar-random pure state $|\psi\rangle$ on $\mathcal{H}$. Let $\mathcal{T}$ be any learning tree of depth T describing a conventional experiment in the single-copy access model.*

Let p_{null} (respectively p_{alt}) denote the induced distribution over leaves ℓ of $\mathcal{T}$ when the unknown state is drawn from the null (respectively alternative) hypothesis. Then

$$d_{\mathrm{TV}}(p_{\mathrm{null}}, p_{\mathrm{alt}}) \le \frac{T(T-1)}{2d}\,.$$

Consequently, any such experiment that uses T single-copy measurements has success probability at most

$$\Pr[\text{correct}] \le \frac{1}{2} + \frac{T(T-1)}{4d}\,.$$

In particular, to achieve success probability at least $2/3$ it is necessary that

$$T \ge \Omega\big(\sqrt{d}\,\big) = \Omega\big(2^{n/2}\big).$$

The rest of this section is devoted to the proof. The strategy is to obtain a uniform, one-sided lower bound on the likelihood ratio $p_{\mathrm{alt}}(\ell)/p_{\mathrm{null}}(\ell)$ at every leaf ℓ of the learning tree, and then apply our likelihood-ratio corollary for total variation distance.

Proof. Fix an arbitrary learning tree $\mathcal{T}$ of depth T. As before, each root-to-leaf path of $\mathcal{T}$ has length T. We now specialize to our two hypotheses. Let

$$\rho_{\mathrm{mm}} := \frac{\mathbb{1}}{d}$$

denote the maximally mixed state, and let μ_{pure} denote the Haar measure over pure states $|\psi\rangle\langle\psi|$ on $\mathcal{H}$. As in the previous section, we define

$$p_{\mathrm{null}}(\ell) := p_{\mathcal{T}}^{\rho_{\mathrm{mm}}}(\ell),$$
$$p_{\mathrm{alt}}(\ell) := \mathbb{E}_{|\psi\rangle\sim\mu_{\mathrm{pure}}}\big[p_{\mathcal{T}}^{|\psi\rangle\langle\psi|}(\ell)\big].$$

We aim to lower bound the likelihood ratio

$$\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)}$$

uniformly over all leaves ℓ.

First, note that for the maximally mixed state we have

$$\langle\phi_{v_t}|\rho_{\mathrm{mm}}|\phi_{v_t}\rangle = \frac{1}{d}\,,$$

and so

$$p_{\mathrm{null}}(\ell) = \prod_{t=0}^{T-1} d\, a_{v_t}\, \frac{1}{d} = \prod_{t=0}^{T-1} a_{v_t}\,.$$

On the other hand, for a pure state $\rho = |\psi\rangle\langle\psi|$ we have

$$p_{\mathcal{T}}^{|\psi\rangle\langle\psi|}(\ell) = \prod_{t=0}^{T-1} d\, a_{v_t}\, |\langle\phi_{v_t}|\psi\rangle|^2.$$

Taking the expectation over a Haar-random $|\psi\rangle$ and dividing, the likelihood ratio becomes

$$\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)} = \frac{\mathbb{E}_{|\psi\rangle}\big[\prod_{t=0}^{T-1} d\, a_{v_t}\, |\langle\phi_{v_t}|\psi\rangle|^2\big]}{\prod_{t=0}^{T-1} a_{v_t}} = d^T\, \mathbb{E}_{|\psi\rangle}\Big[\prod_{t=0}^{T-1} |\langle\phi_{v_t}|\psi\rangle|^2\Big].$$

Thus our task is to bound from below the Haar expectation

$$\mathbb{E}_{|\psi\rangle}\Big[\prod_{t=0}^{T-1} |\langle\phi_{v_t}|\psi\rangle|^2\Big].$$

It is convenient to rewrite this expectation using the T-th moment of a Haar-random pure state. Observe that

$$\prod_{t=0}^{T-1} |\langle\phi_{v_t}|\psi\rangle|^2 = \prod_{t=0}^{T-1} \langle\phi_{v_t}|\psi\rangle\langle\psi|\phi_{v_t}\rangle = \operatorname{tr}\left(|\psi\rangle\langle\psi|^{\otimes T} \bigotimes_{t=0}^{T-1} |\phi_{v_t}\rangle\langle\phi_{v_t}|\right).$$

Taking the expectation over $|\psi\rangle$ and using linearity of the trace we obtain

$$\mathbb{E}_{|\psi\rangle}\Big[\prod_{t=0}^{T-1} |\langle\phi_{v_t}|\psi\rangle|^2\Big] = \operatorname{tr}\big(\mathbb{E}_{|\psi\rangle}\big[|\psi\rangle\langle\psi|^{\otimes T}\big]\, O\big)\,,$$

where we have defined the operator

$$O := \bigotimes_{t=0}^{T-1} |\phi_{v_t}\rangle\langle\phi_{v_t}|$$

which acts on $\mathcal{H}^{\otimes T}$, namely T copies of the Hilbert space.

As discussed earlier, the T-th tensor moment of a Haar-random pure state is proportional to the projector onto the symmetric subspace of $\mathcal{H}^{\otimes T}$. Let Π_{sym} denote this projector, and let

$$\dim(\Pi_{\mathrm{sym}}) = \binom{d+T-1}{T} = \frac{d(d+1)\cdots(d+T-1)}{T!}$$

be the dimension of the symmetric subspace. Then

$$\mathbb{E}_{|\psi\rangle}\big[|\psi\rangle\langle\psi|^{\otimes T}\big] = \frac{\Pi_{\mathrm{sym}}}{\dim(\Pi_{\mathrm{sym}})}.$$

Substituting this into the previous expression and recalling the definition of the likelihood ratio, we find

$$\begin{aligned}\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)} &= d^T \operatorname{tr}\left(\frac{\Pi_{\mathrm{sym}}}{\dim(\Pi_{\mathrm{sym}})}\, O\right) \\ &= \frac{d^T}{\dim(\Pi_{\mathrm{sym}})} \operatorname{tr}\big(\Pi_{\mathrm{sym}}\, O\big).\end{aligned}$$

Next, recall that the projector onto the symmetric subspace has the decomposition

$$\Pi_{\mathrm{sym}} = \frac{1}{T!} \sum_{\pi\in S_T} P_d(\pi),$$

where S_T is the symmetric group on T elements, and $P_d(\pi)$ is the unitary operator on $\mathcal{H}^{\otimes T}$ that permutes the tensor factors according to π. Thus

$$\operatorname{tr}\big(\Pi_{\mathrm{sym}}\, O\big) = \frac{1}{T!} \sum_{\pi\in S_T} \operatorname{tr}\big(P_d(\pi)\, O\big),$$

and therefore

$$\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)} = \frac{d^T}{\dim(\Pi_{\mathrm{sym}})} \cdot \frac{1}{T!} \sum_{\pi\in S_T} \operatorname{tr}\big(P_d(\pi)\, O\big).$$

Using $\dim(\Pi_{\mathrm{sym}}) = \frac{d(d+1)\cdots(d+T-1)}{T!}$, this simplifies to

$$\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)} = \frac{d^T}{d(d+1)\cdots(d+T-1)} \sum_{\pi \in S_T} \mathrm{tr}\big(P_d(\pi)\, O\big). \tag{64}$$

The crucial ingredient is now a lower bound on the sum in (64). Instead of following the original strategy of [CCHL22], we opt for the streamlined strategy of [BCS+25]. To this end we use the following technical lemma.

Lemma 228. *Let $|\psi_0\rangle, \ldots, |\psi_{T-1}\rangle \in \mathcal{H}$ be arbitrary pure states, and let*

$$O = \bigotimes_{t=0}^{T-1} |\psi_t\rangle\langle\psi_t|.$$

Then we have the bound

$$\sum_{\pi \in S_T} \mathrm{tr}\big(P_d(\pi)\, O\big) \geq 1.$$

Proof. Let G be the $T \times T$ Gram matrix of the vectors $|\psi_0\rangle, \ldots, |\psi_{T-1}\rangle$, i.e.

$$G_{ij} := \langle\psi_i|\psi_j\rangle.$$

Using the definition of $P_d(\pi)$ and standard properties of the trace, one checks that

$$\sum_{\pi \in S_T} \mathrm{tr}\big(P_d(\pi)\, O\big) = \sum_{\pi \in S_T} \prod_{t=0}^{T-1} G_{t,\pi^{-1}(t)} = \mathrm{perm}(G),$$

where $\mathrm{perm}(G)$ denotes the permanent of G.

The Gram matrix G is Hermitian positive semidefinite, and its diagonal entries satisfy $G_{tt} = \langle\psi_t|\psi_t\rangle = 1$ for all t. By [GP88], for any such matrix we have

$$\mathrm{perm}(G) \geq 1.$$

Combining these facts yields

$$\sum_{\pi \in S_T} \mathrm{tr}\big(P_d(\pi)\, O\big) = \mathrm{perm}(G) \geq 1,$$

which proves the lemma. □

Applying Lemma 228 to the states $|\phi_0\rangle, \ldots, |\phi_{T-1}\rangle$ along the path to ℓ gives

$$\sum_{\pi \in S_T} \mathrm{tr}\big(P_d(\pi)\, O\big) \geq 1.$$

Substituting this into (64) we obtain the uniform lower bound

$$\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)} \geq \frac{d^T}{d(d+1)\cdots(d+T-1)} = \prod_{t=0}^{T-1} \frac{d}{d+t} = \prod_{t=0}^{T-1} \frac{1}{1+t/d}.$$

We now turn this product into a more explicit bound. Taking logarithms,

$$\log\left(\frac{p_{\mathrm{alt}}(\ell)}{p_{\mathrm{null}}(\ell)}\right) \geq -\sum_{t=0}^{T-1} \log\left(1 + \frac{t}{d}\right).$$

Using the elementary inequality $\log(1+x) \le x$ for all $x \ge 0$, we get

$$\log\left(\frac{p_{\text{alt}}(\ell)}{p_{\text{null}}(\ell)}\right) \ge -\sum_{t=0}^{T-1} \frac{t}{d} = -\frac{T(T-1)}{2d}.$$

Exponentiating both sides and using $e^{-x} \ge 1-x$ for all $x \ge 0$ yields

$$\frac{p_{\text{alt}}(\ell)}{p_{\text{null}}(\ell)} \ge \exp\left(-\frac{T(T-1)}{2d}\right) \ge 1 - \frac{T(T-1)}{2d}.$$

Thus for every leaf ℓ we have the one-sided likelihood ratio bound

$$\frac{p_{\text{alt}}(\ell)}{p_{\text{null}}(\ell)} \ge 1-\delta, \qquad \text{where} \quad \delta := \frac{T(T-1)}{2d}. \tag{65}$$

We are now in a position to apply our likelihood-ratio corollary. Recall that $p_{\text{null}}(\ell) > 0$ for every leaf that can actually be reached under the null hypothesis. (If some leaf has $p_{\text{null}}(\ell) = 0$ then it has $p_{\text{alt}}(\ell) = 0$ as well, so it can be removed from the tree without affecting the experiment.) For all remaining leaves the bound (65) holds.

By the Corollary 226, the condition

$$\frac{p_{\text{alt}}(\ell)}{p_{\text{null}}(\ell)} \ge 1-\delta \quad \text{for all leaves } \ell,$$

with $\delta = T(T-1)/(2d)$, implies

$$d_{\text{TV}}(p_{\text{null}}, p_{\text{alt}}) \le \delta = \frac{T(T-1)}{2d}.$$

Combining this with Le Cam's inequality (applied to the two classical distributions p_{null} and p_{alt}) we obtain

$$\Pr[\text{correct}] \le \frac{1}{2} + \frac{1}{2}\, d_{\text{TV}}(p_{\text{null}}, p_{\text{alt}}) \le \frac{1}{2} + \frac{T(T-1)}{4d},$$

which is exactly the claimed success probability bound.

Finally, suppose an experiment achieves success probability at least $2/3$. Then we must have

$$\frac{1}{2} + \frac{T(T-1)}{4d} \ge \frac{2}{3},$$

which rearranges to

$$T(T-1) \ge \frac{2}{3} d.$$

For all sufficiently large d this forces

$$T \ge \Omega\big(\sqrt{d}\,\big) = \Omega\big(2^{n/2}\big),$$

as claimed in Theorem 227. This completes the proof. □