Quantum Learning Theory

Sitan Chen
Jordan Cotler
Hsin-Yuan Huang

Contents

Preface

We begin our journey by motivating our studies in two different ways, one more pragmatic and one more philosophical.

Machine learning, and in particular its recent manifestation in deep learning in the last two decades, has been transformative for computer science and information technology. The promise, perils, and possibility of generative artificial intelligence have seeped from Silicon Valley to the public discourse, and the ultimate contours of its potential are the subject of intense speculation. Granted all of the recent developments in contemporary machine learning, many of the core ideas derive from *statistical learning theory*, which had its heyday in the 1990's and early 2000's. This is a rigorous mathematical subject which conceives of learning in a probabilistic and often Bayesian manner, drawing on probability theory and empirical process theory, while utilizing information-theoretic concepts from Shannon's foundational work. Since contemporary machine learning is mostly an *empirical* subject pertaining to extraordinarily sophisticated statistical models which defy comprehensive characterization, the particularities of the theorems developed in statistical learning theory are not often used; however, the intuitions these rigorous results provide are essential for designing new neural network architectures, loss functions, training algorithms, and datasets. As such, the afterlife of statistical learning theory is that its quantitative knowledge in mathematically simple settings has been lifted to qualitative but indispensable wisdom about highly complex systems.

Then one motivation for our studies is to develop a quantum version of statistical learning theory (or more succinctly, quantum learning theory), suitable for future application by quantum computers. Our studies will focus on quantum learning for *quantum data* as opposed to classical data, for reasons that will be explained. (Indeed, the latter setting is very interesting but has a somewhat different character.) The subject will necessarily be mathematically rigorous to cement our understanding of quantum data and quantum learning algorithms, as well as to develop robust methods with provable performance guarantees suitable for scientific applications. We emphasize that at this moment in time, quantum learning theory is not chiefly an empirical subject such as contemporary machine learning; this underscores the necessity of mathematical rigor and the importance of the foundational development of basic quantum learning algorithms and methods that future theoretical or empirical inquiries may build on. We will focus on developments in quantum learning theory mostly from 2019 onward, which saw the development of fruitful foundations and applications of the subject.

There is also a second, more philosophical motivation for our study of quantum learning theory. *Epistemology* is the philosophical study of what we can know about the world, and how we come to know it. One of the earlier treatments of the subject goes back over 2000 years to Plato, although among scientists Descartes' maxim

"I think therefore I am" (*cogito ergo sum*) may be more familiar. Specifically, Descartes was concerned with what he could know with certainty about the world, and upon wrestling with various uncertainties he concludes that he knows at the very least that he himself exists, since for him to even render the thought requires his own existence.

A persistent thread in epistemology since the beginning is that there may be aspects of our reality that we can never come to know. A particularly incisive analysis along these lines was developed by Immanuel Kant in the late 18th century, in which he detailed how the physicality of our corporeal beings and the constitution of our minds place a priori fundamental limitations on what we can know about the world, leaving certain truths necessarily out of our reach. While this premise is widely accepted by philosophers, it is often frowned upon by scientists; after all, we are children of the Enlightenment for which scientific knowledge is infinitely extensible and far-reaching. If you feel such an urge to frown on epistemology, consider this more modern example: we live in a universe which is expanding and accelerating. Eventually, the expansion will be so fast that light from the early universe will become so redshifted as to be undetectable. As such, if there is life that develops somewhere in the universe at such a time, they will never be able to empirically determine that there was a Big Bang. Thus a truth about the universe is, to them, out of reach.

In the early 20th century, David Hilbert famously declared that the *mathematical* world was fundamentally knowable, and that every precise mathematical statement was either true or false. This epistemic totalism was shockingly undermined by Kurt Gödel in 1931, when he showed that there must always exist mathematical statements which can neither be proved true nor false. This death blow to Hilbert's (and Bertrand Russell's) conception of mathematical knowledge was concretized by Alan Turing in his foundational work on computer science, the pragmatic heir to mathematical logic. Turing famously showed in 1936 that there is no algorithm (which is guaranteed to terminate in finite time) that can conclusively decide if any given algorithm will halt or not. Thus Turing's theory of undecidability cleaves out facts about the world which are fundamentally unknowable to us, furnishing totally precise examples of epistemic roadblocks.

Decades later starting in the early 1970's, the subject of *computational complexity* began to emerge. Instead of being concerned with whether the solution to a computational problem was knowable or unknowable, the subject focused on the *difficulty* of computational problems. For example, one can show that sorting a list of n items (on a classical computer) requires *at most* $\sim n \log n$ computational steps, but also *at least* $\sim n \log n$ computational steps, thus pinpointing the absolute difficulty of the problem. Some computational problems have polynomial difficulty whereas others have exponential difficulty, and are stratified according to *computational complexity classes*. In this way, computational complexity theory comprises a quantitative form of epistemology, circumscribing how difficult it is to obtain computational knowledge.

Having set the scene, we turn to a deep question: how do we come to learn about the world through scientific inquiry? A key facet is that we interrogate the natural world through experiment, and algorithmically process our collected data to reveal hitherto unknown properties of nature. More formally, we can conceptualize a system in nature – such as a superconductor, a vat of chemicals, a biological

organism, etc. – as a source of *data* which is not fully characterized (or else we would not need to run the experiment); then our experiment comprises a series of interactions with the world to sample data, subsequent computational processing of the data, and possibly additional interactions with the world predicated on the processing of previous data. The ultimate outcome is that we learn a property of the world, such as the charge of the electron, the symmetry of a crystal, etc. In this manner, we see that scientific experiments can be beautifully and precisely abstracted into the framework of learning theory. Therefore, a quantitative study of learning theory can reveal what we can fundamentally come to know about the world, and how difficult it is to do so.

Since the laws of nature are quantum-mechanical, any theory of learning the natural world must take quantum mechanics into account. In particular, the natural systems we seek to understand may be quantum-mechanical; the data we extract can be quantum-mechanical; and our means of processing that data can be quantum-mechanical. Thus we necessitate a quantum theory of learning. Such a theory reveals that there are facets of the natural world which are inaccessible to us unless we can harness quantum computers to couple to natural systems and perform quantum information processing. More bluntly, quantum learning gives us access to properties of the natural world which are otherwise unknowable by classical means. And yet the same theory shows us which properties of the natural world are forever out of reach, even with the aid of vast quantum computational power.

Quantum learning theory circumscribes what is knowable and unknowable about the natural world, providing a quantitative epistemology of the grasp of scientific inquiry. With so much at stake, let us begin.

CHAPTER 1

A Sneak Peek: Learning a Rotation Matrix

Before we dive into the formalism of quantum learning, let us begin with a simple motivating example. You will not need to know any quantum mechanics to understand the setup, but it mirrors, in a stripped–down way, how many real experiments try to learn an unknown quantum process that one can interact with in the lab.

1. Basic Setup

Suppose there is an unknown two-dimensional rotation matrix

$$U = R(\theta) \triangleq \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix},$$

and for simplicity assume that $0 \leq \theta < \pi/2$. Your goal is to figure out what θ is, to within small error.

You can "perform an experiment" on it via the following model. Starting from the first standard basis vector $v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, you can decide in advance upon any collection of "controls" specified by rotation matrices $O_0 = R(\theta_0), \ldots, O_m = R(\theta_m)$, and apply (i.e., left-multiply by) the transformation

$$O_m U O_{m-1} U O_{m-2} \cdots U O_0 \,.$$

This results in some new unit vector $w = \begin{bmatrix} x \\ y \end{bmatrix}$.

Our figure of merit will be **statistical efficiency**, namely we want to learn θ to within some acceptable level of error using as few experiments and "queries" to U (e.g., the above experiment makes m queries to U) as possible.

If we could see the entries of w, then it's not hard to learn U. In fact we don't even need the full flexibility of picking $O_1, \ldots, O_m$: we can simply take $m = 0$ and use no controls whatsoever, so that the above transformation is given by U itself and $w = Uv$. In this case, if $w = (x, y)$, we can simply read off the angle of rotation defining U from $\theta = \arccos(x)$.

There is a crucial catch however: in physical experiments, we never get to see the literal vector w resulting from an experiment. Without getting into the quantum details yet, the reason is that w is a *superposition* between two different states, namely the first standard basis vector and the second standard basis vector. What we can do is **measure** w, at which point we observe one state or the other, but in a probabilistic fashion.

Definition 1 (Born rule – baby version)**.** *Given unit vector $w = (x, y)$, if we* measure *it, we get as output either $\begin{bmatrix}1\\0\end{bmatrix}$ with probability x^2, or $\begin{bmatrix}0\\1\end{bmatrix}$ with probability y^2. Note that as w is a unit vector, this is a valid probability distribution.*

One might thus envision a natural workaround to not having access to the exact entries (x, y) of w. Measuring effectively gives us access to biased coin flips: with probability x^2 we see heads and with probability y^2 we see tails. By repeatedly performing the experiment that results in w and measuring w each time, we can estimate x^2 and y^2 simply by computing the fraction of heads and tails we observe. How many repetitions do we need?

This can be computed with the **Chernoff bound**:

Fact 2 (Chernoff bound)**.** *Let $X_1, \ldots, X_N$ be independent Bernoulli random variables with expectation p. Then for any $t > 0$,*

$$\Pr\Big[\Big|\frac{1}{N}\sum_i X_i - p\Big| > t\Big] \le 2\exp\Big(-\frac{Nt^2}{2p(1-p)}\Big)$$

In our setting, $p = x^2$, and $\hat{X} = \frac{1}{N}\sum_i X_i$, so if we apply the above with $t = \epsilon x\sqrt{1-x^2}$ and $N = 2\log(2/\delta)/\epsilon^2$, the right-hand side of the above bound is δ and we conclude that with

$$O(\log(2/\delta)/\epsilon^2) \tag{1}$$

coin tosses, we can produce an estimate $0 \le \hat{X} \le 1$ such that $|x^2 - \hat{X}| \le \epsilon x\sqrt{1-x^2}$ with probability at least $1-\delta$. We can then output $\arccos(\sqrt{\hat{X}})$ and argue that this is $O(\epsilon)$-close to θ with some elementary calculus (the reader can safely skip this on a first reading without losing any of the core intuition):

Proposition 3. *Let $0 \le \epsilon \le 1/2$. Suppose $0 \le X, X' \le 1$ satisfy $|X - X'| \le \epsilon\sqrt{X(1-X)}$. Then*

$$|\arccos(\sqrt{X}) - \arccos(\sqrt{X'})| \le \epsilon\,.$$

Proof. As $\arccos(\sqrt{X})$ and $\sqrt{X(1-X)}$ are symmetric about $X = 1/2$, we may assume without loss of generality that $|X| \le 1/2$.

If $|X| \ge \frac{\epsilon^2}{\epsilon^2+4}$, then $X/2 \le X' \le X + \epsilon/2$. If we define $f(z) \triangleq \arccos(\sqrt{z})$, then $f'(z) = -\frac{1}{2\sqrt{z(1-z)}}$, so $|f'(z)| \le 2|f'(X)|$ for all z between X and X'. By integrating, we conclude that

$$|\arccos(\sqrt{X}) - \arccos(\sqrt{X'})| \le \frac{1}{\sqrt{X(1-X)}} \cdot \epsilon\sqrt{X(1-X)} \le \epsilon$$

as desired.

If $|X| < \frac{\epsilon^2}{\epsilon^2+4} \le \frac{\epsilon^2}{4}$, then $\arccos(\sqrt{X}) \ge \pi/2 - \epsilon$. If $X' \le X$, then $\arccos(\sqrt{X}) \le \arccos(\sqrt{X'}) \le \pi/2$, so $|\arccos(\sqrt{X}) - \arccos(\sqrt{X'})| \le \epsilon$. If $X' \ge X$, then $|f'(z)| \le 2|f'(X)|$ for all z between X' and X, so the claimed bound follows again by integrating. □

The $1/\epsilon^2$ scaling in Eq. (1) for the number of coin tosses is called the **standard quantum limit** – often it is formulated in the reverse direction, namely using N experiments (sometimes called "shots"), one can estimate the unknown parameter θ to error $\sim 1/\sqrt{N}$.

2. Beating the Standard Quantum Limit

Of course, we are not yet done. While it is an unavoidable fact of life that in the classical world, estimating the bias of a coin to error ϵ requires $\sim 1/\epsilon^2$ coin tosses in general, in the quantum world we are not limited to simply reducing learning rotations to learning the bias of a random coin. Indeed, the approach described above is exceedingly naive: we set $m = 0$ and didn't use any controls O_i whatsoever.

It turns out that by being clever about the choice of experiments, we can do much better, in fact with only $O(\log 1/\epsilon)$ experiments and $O(1/\epsilon)$ queries to U in total across all experiments. The $O(1/\epsilon)$ scaling is called the **Heisenberg limit**: this turns out to be a fundamental barrier that no experimental protocol, no matter how clever, can beat.

The key idea is to "bootstrap." Instead of estimating θ to high precision right off the bat, we are going to gradually refine our estimate. As a thought experiment, imagine we start by getting a relatively crude approximation to θ by running the protocol in the previous section for target precision which is just a small constant, say, $\epsilon_{\mathsf{crude}} = 1/4$. We can accomplish this with probability $1 - \delta$ using only $O(\log(1/\delta))$ experiments and queries to U, with no dependence yet on the final target precision ϵ.

Given this estimate, if we further subtract $\epsilon_{\mathsf{crude}}$ from it, we get an angle $\theta^{(1)}$ which is an underestimate of θ by a margin of at most $2\epsilon_{\mathsf{crude}} \leq 1/2$. To estimate θ, it now suffices to estimate the *residual angle* $\theta - \theta^{(1)}$. So in all subsequent experiments, instead of querying $U = R(\theta)$, we can query

$$U^{(1)} \triangleq UR(\theta^{(1)})^\dagger = R(\theta - \theta^{(1)}) .$$

Here is our main claim:

Lemma 4. *Suppose $\theta - 2^{-k} \leq \theta^{(k)} \leq \theta$ and let $U^{(k)} \triangleq R(\theta - \theta^{(k)})$. Let $\delta_k > 0$. Consider the following protocol:*

- *Repeat the following experiment $C \log 1/\delta_k$ times:*
  - *Apply $U^{(k)}$ a total of 2^k times starting from the first standard basis vector v.*
  - *Measure the resulting vector and record the observation (heads or tails)*
- *Let $\hat{X}$ denote the fraction of heads seen across these experiments.*
- *Define*

$$\theta^{(k+1)} = \theta^{(k)} + \arccos(\sqrt{\hat{X}})/2^k - 1/2^{k+2} .$$

For C a sufficiently large absolute constant, we have $\theta - 1/2^{k+1} \leq \theta^{(k+1)} \leq \theta$ with probability at least $1 - \delta_k$.

Proof sketch. Note that the rotation given by applying $U^{(k)}$ a total of 2^k times is $R(2^k(\theta - \theta^{(k)}))$. By taking C sufficiently large, the argument in the previous section implies that $|\arccos(\sqrt{\hat{X}}) - 2^k(\theta - \theta^{(k)})| \leq 1/4$. Dividing by 2^k on both sides, we conclude that with probability at least $1 - \delta_k$,

$$\theta - 1/2^{k+2} \leq \theta^{(k)} + \arccos(\sqrt{\hat{X}})/2^k \leq \theta + 1/2^{k+2} .$$

Subtracting $1/2^{k+2}$ from all sides and recalling the definition of $\theta^{(k+1)}$ above, we conclude that $\theta^{(k+1)}$ is an underestimate of θ by at most $1/2^{k+1}$ as claimed. □

Continuing in this fashion up to $k = \overline{k} \triangleq \lceil \log_2 1/\epsilon \rceil$, we obtain an angle $\theta^{(\overline{k})}$ which underestimates θ by at most ϵ, with probability at least $1 - \sum_k \delta_k$. Suppose in each round k, we take $\delta_k \triangleq \delta 2^{k-\overline{k}-1}$, so that $\sum_k \delta_k \leq \delta$.

Furthermore, in any round k, we perform

$$2^k \cdot C \log 1/\delta_k = 2^k \cdot C \log 1/\delta + O(2^k \cdot C(\overline{k} + 1 - k))$$

queries to $U^{(k)}$ which amounts to as many queries to U. So the total number of queries made to U is

$$(1 + 2 + \cdots + 2^{\overline{k}}) C \log 1/\delta + \sum_{k=0}^{\overline{k}} O(2^k(\overline{k} + 1 - k)) = O(\log(1/\delta)/\epsilon)$$

as desired.

3. Looking Ahead

3.1. Rotation Learning in the Wild

The rotation learning problem can be thought of as a toy stand–in for a physical process that imprints a phase θ on a two–level system (a qubit, a pair of optical modes, or a two–dimensional invariant subspace inside a larger device).

In precision sensing (gravitational–wave interferometers, atomic clocks, Ramsey spectroscopy, and phase estimation in general), the central task is to learn a small θ as efficiently as possible. Real instruments like LIGO [**AAA+13**] do not literally implement the protocol we analyze here; for example, they inject *squeezed light* to reduce measurement noise rather than concatenating many coherent applications of the same unknown operation. Squeezing is notably more robust to realistic optical losses than schemes that try to amplify phase information solely by repeated coherent evolution or fragile entangled probes. Still, at the level of information flow, many metrology strategies can be idealized as:

$$\text{(prepare a known state)} \xrightarrow{\text{apply } U_\theta\text{, possibly with controls}} \text{(measure and update)}.$$

Our rotation–learning toy model captures precisely this prepare–evolve–measure loop. It allowed us to isolate two ingredients that matter for sample complexity: (i) how coherently we can *accumulate* phase information (e.g. by applying U multiple times or by clever controls), and (ii) how we post-process this information phase into a reliable estimate of the unknown quantum object. The formalism developed in this book will vastly generalize this example and its strategy.

3.2. Extensions

Although we illustrated the bootstrap idea with a 2×2 rotation, in fact the same idea can be extended to learn *any* 2×2 unitary matrix in $O(1/\epsilon)$ total queries.

In fact, one can even extend this beyond 2 dimensions. What is needed is an appropriate generalization of the step where we estimated the bias of a coin toss to constant error $\epsilon_{\mathsf{crude}}$. The relevant ideas for doing this will be introduced later on in this lecture when we discuss **tomography**. When we move from 2 dimensions to d dimensions however, the crucial change is that the number of queries will now depend on d. The intuition is that a completely unknown unitary on a d–dimensional space has d^2 real parameters, so without additional assumptions, one should not expect to learn all of these parameters until the number of queries

scales with $O(d^2)$. Indeed, it was shown recently [**HKOT23**] that the optimal query complexity for learning an arbitrary unitary matrix in d dimensions in the above model is exactly d^2/ϵ, up to constant factors. The argument we presented above is really just a baby version of the argument in that work.

Unfortunately, in the settings we will be interested in, we will always think of d as scaling *exponentially* in the number of "particles" in the system. To avoid exponential scaling, we then need to posit additional *structure* and align the learning task with that structure. For example, one standard and "physically reasonable" choice of structure to assume is that the unknown unitary takes the form of $U = e^{-iH}$, where H is a **local Hamiltonian** on n qubits; we will define this in due time, but for now the intuition to keep in mind is that this Hamiltonian is described by a total number of free parameters that only scales *polynomially* in the number of particles. Under such structural assumptions, one can then hope to develop algorithms that scale much more efficiently – we will cover these in a later unit in this course.

As another preview for what is to come, note that one can consider other models of interaction. In the query model we considered in this lecture, we allowed arbitrary control, and our choice of experiments was adaptive over the different rounds of the learning protocol. One could further enhance this model by, for instance, performing m entangled experiments in parallel, expanding the relevant dimension from d to d^m. While this doesn't end up buying much for the unitary learning problem, in many other quantum learning settings this can make a big difference in the efficiency with which one can learn. In the other direction, one can also consider *weaker* models where, perhaps due to various practical constraints on the experimental apparatus like hardware limitations or noise, we cannot perform arbitrary control. The effect of such constraints on the ultimate efficiency with which we can learn quantum states is another central theme in these notes.

Stepping further back, the rotation–learning example isolates three ingredients that will organize the rest of the book: the *unknown object* (a state, unitary, channel, or Hamiltonian), the *access model* (how we may prepare inputs, interleave known controls, parallelize or reuse the device, and measure), and the *loss metric* (in this case, the "parameter error" with which we estimate θ). Throughout the course of these lectures, we will use these basic ingredients to develop the foundations of a general theory of quantum learning.

Part 1

Quantum Mechanics Toolkit

CHAPTER 2

Essentials of Quantum Mechanics

We begin by building up the basic ingredients of quantum mechanics. This is not meant to be a course on quantum mechanics, and so we will proceed pragmatically and without much fanfare. We will have the luxury of working with finite-dimensional Hilbert spaces (if you do not know what this means, you will soon), since this is the setting of most present applications of quantum learning theory. Our pedagogical approach will be to revisit ordinary probability theory in a suggestive way that naturally generalizes to quantum theory. Our exposition is meant to be accessible to readers with a knowledge of linear algebra and probability theory.

1. Probability theory on vector spaces

1.1. Probability distributions and their transformations

Here we will formulate probability theory on a discrete space, with some additional linear algebraic baggage that will be useful later. If we have a set of size N we can represent a probability distribution over that set as a vector in $\mathbb{R}^N$ given by

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix}$$

where p_i is the probability of the ith item. We have, out of convenience, chosen an ordering on our set of items so that we can organize the probabilities into a vector, but of course this ordering is arbitrary. As usual, we require $p_i \geq 0$ for all i since probabilities cannot be negative, and also $\sum_{i=1}^N p_i = 1$ so that the probabilities are appropriately normalized. There is a natural way of packaging the normalization condition. To this end, consider the row vector

$$\vec{1}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} .$$

Then $\sum_{i=1}^N p_i = 1$ is equivalent to

$$\vec{1}^T \cdot \vec{p} = 1 ,$$

and we will use this more compact expression henceforth. It will sometimes be useful to consider the *probability simplex* Δ_N which is a subset of $\mathbb{R}^N$, where Δ_N consists of all nonnegative vectors with entries summing to one. Then we can write $\vec{p} \in \Delta_N$.

Next we consider a rudimentary version of *dynamics*. That is, what kinds of transformations on $\vec{p}$ will map it into another valid probability distribution? The

simplest kind of transformation we can imagine is a linear one, so let us examine that first. Letting M be an $N \times N$ matrix, we consider the transformation

$$\vec{p}' = M \cdot \vec{p},$$

so that $\vec{p}'$ is the new probability distribution after the transformation. But what conditions do we need to put on M such that $\vec{p}'$ is a bona fide probability distribution for all initial distributions $\vec{p}$? Well, we need for all entries of $\vec{p}'$ to be nonnegative, and for $\vec{1}^T \cdot \vec{p}' = 1$. To ensure the first property, suppose that $\vec{p}$ is all zeroes except for the jth entry which equals one. (That is, we would sample the jth object with probability 1 and never sample anything else.) To introduce some other notation, let $\vec{e}_j$ be vector which is all zeroes except for the jth entry which equals one. Then we have

$$\vec{p}' = M \cdot \vec{e}_j = \begin{bmatrix} M_{1j} \\ M_{2j} \\ \vdots \\ M_{Nj} \end{bmatrix} .$$

In order for all entries of $\vec{p}^{\,'}$ to be nonnegative, we evidently require $M_{ij} \geq 0$ for all j, and i fixed. Varying over i as well, we find the requirement that $M_{ij} \geq 0$ for all i, j, and so M must be a matrix with nonnegative entries. Since we also demand that $\vec{1}^T \cdot \vec{p}' = 1$, we find the condition

$$\vec{1}^T \cdot \vec{p}' = \vec{1}^T \cdot M \cdot \vec{e}_j = \vec{1}^T \cdot \begin{bmatrix} M_{1j} \\ M_{2j} \\ \vdots \\ M_{Nj} \end{bmatrix} = 1 .$$

That is, the jth column of M must sum up to one. Since this must hold for every column, we find the condition

$$\vec{1}^T \cdot M = \vec{1}^T . \tag{2}$$

Thus a nonnegative matrix satisfying (2) will send probability vectors to probability vectors. We honor this finding with a definition:

Definition 5 (Markov matrix)**.** *Let M be an $N \times N$ matrix. We say that M is a* ***Markov matrix*** *if $M_{ij} \geq 0$ for all i, j, and $\vec{1}^T \cdot M = \vec{1}^T$. Then M maps probability vectors to probability vectors.*

A few comments are in order. In many treatments of Markov matrices, there is a different convention in which M is taken to act on probability distributions 'to the left', which would give the transpose our definition above. Our conventions here are chosen to align with those of quantum mechanics, as we will see later on.

We immediately notice that Markov matrices behave nicely under composition. Specifically, we have the useful lemma:

Lemma 6 (Composition of Markov matrices)**.** *If $M_1, M_2, ..., M_k$ are Markov matrices, then $M_k \cdots M_2 \cdot M_1$ is also a Markov matrix.*

The proof of this useful fact follows by a short calculation using the definition (which you should do if you have not thought it through before). The upshot of

this lemma is that we can consider transformations like

$$\vec{p}\,' = M_k \cdots M_2 \cdot M_1 \cdot \vec{p}$$

as instantiating a type of 'circuit', with depth k. That is, we could say the words: starting with $\vec{p}$ we apply M_1 followed by M_2 followed by M_3 and so on, and then finally apply M_k.

Before moving on to increasing levels of sophistication, we consider a simple example:

Example 1 (Bernoulli coin, $N = 2$). We now specialize to a two–outcome space and fix the ordering so that the first coordinate is outcome 0 ("success") and the second is outcome 1 ("failure"). A Bernoulli distribution with success probability θ is therefore represented by

$$\vec{p}_\theta = \begin{bmatrix} \Pr[0] \\ \Pr[1] \end{bmatrix} = \begin{bmatrix} \theta \\ 1-\theta \end{bmatrix}, \qquad \theta \in [0,1].$$

Consider the *bit–flip* dynamics with flip probability $\varepsilon \in [0,1]$,

$$M_\varepsilon = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix},$$

whose entries are nonnegative and whose columns each sum to 1, so M_ε is a Markov matrix in our sense. Acting on $\vec{p}$ produces

$$\vec{p}_{\theta'} = M_\varepsilon\, \vec{p}_\theta = \begin{bmatrix} (1-\varepsilon)\theta + \varepsilon(1-\theta) \\ \varepsilon\theta + (1-\varepsilon)(1-\theta) \end{bmatrix} \quad \Longrightarrow \quad \theta' = (1-2\varepsilon)\,\theta + \varepsilon,$$

where $\theta' = \Pr'[0]$ is the new success probability.

Some immediate checks help build intuition. When $\varepsilon = 0$ the map is the identity; when $\varepsilon = 1$ it deterministically flips $0 \leftrightarrow 1$; and when $\varepsilon = \frac{1}{2}$ it sends every input to the uniform distribution $\vec{p}_{1/2} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ in one step. For any $0 < \varepsilon < 1$, the unique fixed point solves $\theta' = \theta$ and is $\theta_* = \frac{1}{2}$. (To see this, simply solve $\theta_* = (1-2\varepsilon)\theta_* + \varepsilon$ for θ_*). Iterating M_ε a total of k times yields exponential mixing toward the fixed point θ_* at rate $|1-2\varepsilon|$:

$$\theta^{(k)} = \big(1-2\varepsilon\big)^k \Big(\theta^{(0)} - \frac{1}{2}\Big) + \frac{1}{2}\,.$$

Finally, the family M_ε of Markov matrices is closed under composition (illustrating the lemma above): a short calculation shows

$$M_\eta\, M_\varepsilon = M_{\varepsilon+\eta-2\varepsilon\eta}\,,$$

and in particular $M_\varepsilon^k = M_{\varepsilon_{\text{eff}}}$ with

$$\varepsilon_{\text{eff}} = \frac{1-(1-2\varepsilon)^k}{2}.$$

This two–state example already displays dynamics, fixed points, and circuit composition within the linear–algebraic language we have been developing.

Moving on, it is useful to recount a few features of probability distributions. If

we have k probability distributions $\vec{p}_1, ..., \vec{p}_k$, then we can form a new probability distribution by forming a convex combination

$$\vec{p}\,' = \sum_{j=1}^{k} r_j \, \vec{p}_j \tag{3}$$

where $r_j \geq 0$ and $\sum_{j=1}^{k} r_j = 1$. To see this, notice that $\vec{p}\,'$ has nonnegative entries and that $\vec{1}^T \cdot \vec{p}\,' = \sum_{j=1}^{k} r_j \, (\vec{1}^T \cdot \vec{p}_j) = \sum_{j=1}^{k} r_j = 1$. We can interpret $r_1, ..., r_k$ as a probability distribution over k items in its own right, and say of (3) that we have a probabilistic mixture of k probability distributions wherein we sample from $\vec{p}_j$ with probability r_j. That is, $r_1, ..., r_k$ is a probability distribution over probability distributions. (You can use this 'meta' statement to impress your friends, if you like.) To make this concrete, consider the following example:

Example 2 (Sampling two coins, $N = 2$). Suppose we have two Bernoulli coins, represented by the probability vectors $\vec{p}_{1/2}$ and $\vec{p}_{1/3}$, respectively. The first one gives heads with probability $1/2$ and tails with probability $1/2$, and the second gives heads with probability $1/3$ and tails with probability $2/3$. Now suppose I have both coins in my pocket in such a way that when I reach in, I grab the first coin with probability $1/4$ and the second coin with probability $3/4$. Then if I reach in and grab a coin and toss it, what is the probability that I would output heads? This is described by the convex combination

$$\frac{1}{4}\,\vec{p}_{1/2} + \frac{3}{4}\,\vec{p}_{1/3} = \begin{bmatrix} 3/8 \\ 5/8 \end{bmatrix},$$

and so evidently the probability of heads is $3/8$.

So far we have only considered *linear* transformations on $\vec{p}$ that map it into another probability distribution. What if we consider nonlinear transformations? One example would be the nonlinear transformation

$$T(\vec{p}) = \begin{bmatrix} \frac{p_1^2}{\sum_{i=1}^{N} p_i^2} \\ \frac{p_2^2}{\sum_{i=1}^{N} p_i^2} \\ \vdots \\ \frac{p_N^2}{\sum_{i=1}^{N} p_i^2} \end{bmatrix}.$$

Another example would be a Bayesian update. There are clearly a vast infinitude of other possibilities as well. Among this infinitude of transformations there is a natural class that interfaces well with convex combinations of probability distributions. In particular, suppose we mandate that T satisfies

$$T\left(\sum_{j=1}^{k} r_j \, \vec{p}_j\right) = \sum_{j=1}^{k} r_j \, T(\vec{p}_j) \tag{4}$$

for any $\vec{p}_1, ..., \vec{p}_k$ and any valid $r_1, ..., r_k$. In words, we are requiring that a transformation of a probabilistic mixture is a probabilistic mixture of transformations (and specifically, the same transformation). Such T's satisfy a nice structure theorem:

Theorem 7 (Mixture-preserving transformations are Markov matrices)**.** *Suppose that $T : \Delta_N \to \Delta_N$ is a mixture-preserving transformation, namely that (4) is satisfied. Then there exists a Markov matrix M such that $T(\vec{p}) = M \cdot \vec{p}$ for all $\vec{p}$.*

PROOF. Write $\vec{p} = \sum_{j=1}^{N} p_j \, \vec{e}_j$. Using the mixture-preserving property of T, we have

$$T(\vec{p}) = T\left(\sum_{j=1}^{N} p_j \, \vec{e}_j\right) = \sum_{j=1}^{N} p_j \, T(\vec{e}_j) \, .$$

Let M be the matrix whose jth column is $T(\vec{e}_j)$. Then $T(\vec{p}) = M \cdot \vec{p}$. Each column $T(\vec{e}_j)$ is a probability vector, so $M_{ij} \geq 0$ and $\vec{1}^T \cdot M = \vec{1}^T$. Thus M is a Markov matrix, as claimed. □

Mixture-preserving transformations are natural from a physical point of view. Imagine a preparation device that, with probabilities $r_1, ..., r_k$, produces one of the distributions $\vec{p}_1, ..., \vec{p}_k$ by consulting some randomly tossed coins you do not get to see. If dynamics could distinguish whether this randomization happened "before" or "after" the transformation, then the timing of the unseen coin flips would be observable from the output statistics alone. Requiring that they not be observable is exactly the statement of (4).

Two simple consequences are worth keeping in mind. First, the admissible dynamics are closed under randomized control: if with probability r_j you implement a Markov matrix M_j, then the overall map is

$$M' = \sum_{j=1}^{k} r_j \, M_j \, ,$$

which is again a Markov matrix since $\vec{1}^T \cdot M' = \sum_{j=1}^{k} r_j \, (\vec{1}^T \cdot M_j) = \vec{1}^T$ and all entries are nonnegative. Second, if one further insists that deterministic states are carried to deterministic states, so that $\vec{e}_j$ never acquires additional randomness, then each column $T(\vec{e}_j)$ must itself be a basis vector. Equivalently, M has exactly one 1 (and zeros elsewhere) in each column. Such matrices are sometimes called *deterministic* or *functional* Markov matrices. If in addition the mapping $j \mapsto i(j)$ is injective (no two distinct columns point to the same basis vector), then M is a permutation matrix.

By contrast, nonlinear updates arise when you condition on a revealed outcome and then renormalize; the rule in that case depends on which outcome was announced, so it is not a single fixed map on Δ_N and does not represent closed-system dynamics. This classical discussion sets the stage for the quantum case, which we will treat soon. (There, the state space becomes the convex set of density operators, mixture-preserving maps become convex-linear "channels," and the role of Markov matrices is played by completely positive, trace-preserving maps.)

1.2. Joint distributions and tensor products

In probability theory it is essential to consider joint distributions. Here we develop the basic operations of joint distributions in a convenient and illuminating linear algebraic notation. First we require some additional tools on the linear algebra side. Specifically, we will upgrade our linear algebraic toolkit to *multi-linear*

algebra. The key operation will be the **tensor product**, which is an operation for joining two or more vector spaces.

We will proceed by motivating the tensor product informally through simple examples, and then give the abstract definition. It is worth paying close attention as the tensor product will serve as an essential piece of mathematical architecture for almost everything in quantum learning theory.

Consider two vectors $\vec{v}, \vec{w}$ in $\mathbb{R}^N$. We denote their tensor product by $\vec{v} \otimes \vec{w}$. To develop what this means, consider the example below.

Example 3. Let $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Then their tensor product $\vec{v} \otimes \vec{w}$ is represented by

$$\vec{v} \otimes \vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \otimes \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ \\ 2 \cdot \begin{bmatrix} 3 \\ 4 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 6 \\ 8 \end{bmatrix} .$$

In words, $\vec{w}$ gets 'sucked in' to $\vec{v}$. Now let us take the tensor product in the other order, namely $\vec{w} \otimes \vec{v}$:

$$\vec{w} \otimes \vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ \\ 4 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 4 \\ 8 \end{bmatrix} .$$

From this we glean that, in general, $\vec{v} \otimes \vec{w} \neq \vec{w} \otimes \vec{v}$. Moreover, since $\vec{v} \in \mathbb{R}^2$ and $\vec{w} \in \mathbb{R}^2$, we notice that $\vec{v} \otimes \vec{w} \in \mathbb{R}^4$. To this end we write $\vec{v} \otimes \vec{w} \in \mathbb{R}^2 \otimes \mathbb{R}^2 \simeq \mathbb{R}^4$.

Example 4. Suppose $\vec{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ so that $\vec{v} \in \mathbb{R}^2$ and $\vec{w} \in \mathbb{R}^3$. Then

$$\vec{v} \otimes \vec{w} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \\ 8 \\ 10 \end{bmatrix} \in \mathbb{R}^6 ,$$

and we write $\vec{v} \otimes \vec{w} \in \mathbb{R}^2 \otimes \mathbb{R}^3 \simeq \mathbb{R}^6$.

From the previous two examples we see the general rule that if $\vec{v} \in \mathbb{R}^N$ and $\vec{w} \in \mathbb{R}^M$, then $\vec{v} \otimes \vec{w} \in \mathbb{R}^N \otimes \mathbb{R}^M \simeq \mathbb{R}^{NM}$. So upon taking the tensor product of two vector spaces, the dimensions multiply. We can generalize this further by contemplating another example:

Example 5. Let $\vec{v} = \begin{bmatrix}1\\2\end{bmatrix}$, $\vec{w} = \begin{bmatrix}3\\4\end{bmatrix}$, and $\vec{u} = \begin{bmatrix}5\\6\end{bmatrix}$. Then we have

$$\vec{v}\otimes\vec{w}\otimes\vec{u} = (\vec{v}\otimes\vec{w})\otimes\vec{u} = \begin{bmatrix}3\\4\\6\\8\end{bmatrix}\otimes\begin{bmatrix}5\\6\end{bmatrix} = \begin{bmatrix}15\\18\\20\\24\\30\\36\\40\\48\end{bmatrix}$$

and $\vec{v}\otimes\vec{w}\otimes\vec{u} \in \mathbb{R}^2\otimes\mathbb{R}^2\otimes\mathbb{R}^2 \simeq \mathbb{R}^8$.

The above example indicates that

$$\mathbb{R}^{N_1}\otimes\mathbb{R}^{N_2}\otimes\cdots\otimes\mathbb{R}^{N_k} \simeq \mathbb{R}^{N_1N_2\cdots N_k}\,,$$

namely that if we take the tensor product of k vector spaces then the result is a vector space which is the product of the dimensions of the constituents.

We are now ready to define tensor products abstractly, and to really appreciate what it means. Consider the following definition:

Definition 8 (Tensor product)**.** *Let V and W be real vector spaces. A **tensor product** of V and W is a vector space $V\otimes W$ together with a map*

$$\otimes : V\times W \to V\otimes W, \qquad (v,w)\mapsto v\otimes w,$$

that is bilinear in each argument, i.e. for all scalars $a,b,c\in\mathbb{R}$ and vectors $\vec{v},\vec{w},\vec{u}$,

$$(a\,\vec{v}+b\,\vec{w})\otimes\vec{u} = a\,(\vec{v}\otimes\vec{u}) + b\,(\vec{w}\otimes\vec{u}),$$
$$\vec{v}\otimes(b\,\vec{w}+c\,\vec{u}) = b\,(\vec{v}\otimes\vec{w}) + c\,(\vec{v}\otimes\vec{u}),$$

and in particular $(a\vec{v})\otimes\vec{w} = \vec{v}\otimes(a\vec{w}) = a(\vec{v}\otimes\vec{w})$. Concretely, one may construct $V\otimes W$ as the vector space spanned by formal symbols $v\otimes w$ modulo the above bilinearity relations.

To connect this with coordinates, fix bases $\{\vec{e}_i\}_{i=1}^N$ of $\mathbb{R}^N$ and $\{\vec{f}_j\}_{j=1}^M$ of $\mathbb{R}^M$. Then the NM simple tensors $\{\vec{e}_i\otimes\vec{f}_j\}_{i,j}$ form a basis of $\mathbb{R}^N\otimes\mathbb{R}^M$, and so $\dim(\mathbb{R}^N\otimes\mathbb{R}^M) = NM$. If $\vec{v} = \sum_i v_i\,\vec{e}_i$ and $\vec{w} = \sum_j w_j\,\vec{f}_j$, then

$$\vec{v}\otimes\vec{w} = \sum_{i,j} v_iw_j\,(\vec{e}_i\otimes\vec{f}_j)\,,$$

which recovers the stacking rules seen in the earlier examples and realizes the identification $\mathbb{R}^N\otimes\mathbb{R}^M \simeq \mathbb{R}^{NM}$.

Identifying $\mathbb{R}$ with the one–dimensional space spanned by 1, there are canonical isomorphisms $V\otimes\mathbb{R}\simeq V\simeq\mathbb{R}\otimes V$ given by $\vec{v}\otimes a\mapsto a\,\vec{v}$ and $a\otimes\vec{v}\mapsto a\,\vec{v}$. Hence $\mathbb{R}^N\otimes\mathbb{R}^1\simeq\mathbb{R}^N\simeq\mathbb{R}^1\otimes\mathbb{R}^N$.

Linear maps interact nicely with tensor products. If $A:\mathbb{R}^N\to\mathbb{R}^{N'}$ and $B:\mathbb{R}^M\to\mathbb{R}^{M'}$ are linear, there is a linear map $A\otimes B:\mathbb{R}^N\otimes\mathbb{R}^M\to\mathbb{R}^{N'}\otimes\mathbb{R}^{M'}$ defined by

$$(A\otimes B)(\vec{v}\otimes\vec{w}) = (A\vec{v})\otimes(B\vec{w})$$

which in matrix form is the familiar Kronecker product.

Remark 9 (Associativity of tensor products)**.** *For our purposes, it does not matter whether we first form $(V \otimes W)$ and then tensor with U from the right, or first form $(W \otimes U)$ and then tensor with V from the left. There is a canonical identification between*

$$(V \otimes W) \otimes U \quad \text{and} \quad V \otimes (W \otimes U),$$

and so we will simply write

$$V \otimes W \otimes U$$

without worrying about parentheses. This scales to many tensor factors. For a vector space V we write

$$V^{\otimes k} := \underbrace{V \otimes \cdots \otimes V}_{k \text{ copies}},$$

which has dimension $(\dim V)^k$ and a basis $\{\vec{e}_{i_1} \otimes \cdots \otimes \vec{e}_{i_k}\}$. We will use this to model multi–part systems: for example, a register of k N-ary variables naturally lives in $(\mathbb{R}^N)^{\otimes k} \simeq \mathbb{R}^{N^k}$.

As a word of caution, order still matters*. As we explained before, in general we have $\vec{v} \otimes \vec{w} \neq \vec{w} \otimes \vec{v}$. When we want to swap the order of a tensor product we will use the linear map $\mathsf{SWAP} : V \otimes W \to W \otimes V$, acting by*

$$\mathsf{SWAP} \cdot (\vec{v} \otimes \vec{w}) = \vec{w} \otimes \vec{v}\,.$$

In summary, associativity lets us ignore parentheses; SWAP lets us reorder factors when needed.

Going from the abstract back to the concrete, we have the example below:

Example 6. Suppose you are faced with this mess:

$$(a\,\vec{v} + b\,\vec{w}) \otimes (c\,\vec{s} + d\,\vec{t} + e\,\vec{u}) \otimes (f\,\vec{q} + g\,\vec{r})\,.$$

To expand it, what do you do? *Don't panic.* If you have a long list of things to do, just do them *one at a time*. Specifically in this case, use associativity to expand the bracketed terms first:

$$\begin{aligned}&\underbrace{(a\,\vec{v} + b\,\vec{w}) \otimes (c\,\vec{s} + d\,\vec{t} + e\,\vec{u})} \otimes (f\,\vec{q} + g\,\vec{r}) \\ &= (ac\,\vec{v} \otimes \vec{s} + ad\,\vec{v} \otimes \vec{t} + ae\,\vec{v} \otimes \vec{u} + bc\,\vec{w} \otimes \vec{s} + bd\,\vec{w} \otimes \vec{t} + be\,\vec{w} \otimes \vec{u}) \otimes (f\,\vec{q} + g\,\vec{r}).\end{aligned}$$

Now you can multiply through and expand the rest of the terms as

$$\begin{aligned}&acf\,\vec{v} \otimes \vec{s} \otimes \vec{q} + acg\,\vec{v} \otimes \vec{s} \otimes \vec{r} + adf\,\vec{v} \otimes \vec{t} \otimes \vec{q} + adg\,\vec{v} \otimes \vec{t} \otimes \vec{r} \\ &\quad + aef\,\vec{v} \otimes \vec{u} \otimes \vec{q} + aeg\,\vec{v} \otimes \vec{u} \otimes \vec{r} + bcf\,\vec{w} \otimes \vec{s} \otimes \vec{q} + bcg\,\vec{w} \otimes \vec{s} \otimes \vec{r} \\ &\quad + bdf\,\vec{w} \otimes \vec{t} \otimes \vec{q} + bdg\,\vec{w} \otimes \vec{t} \otimes \vec{r} + bef\,\vec{w} \otimes \vec{u} \otimes \vec{q} + beg\,\vec{w} \otimes \vec{u} \otimes \vec{r}\,,\end{aligned}$$

which is the desired expansion.

With some basic tensor product definitions at hand, we can now leverage them to discuss joint probability distributions in a slick vector space formalism.

Respecting historical tradition,[1] suppose we have two urns, where the first urn has N objects and the second urn has M objects. Suppose that the probability that we select one of the N items in the first urn is described by the probability

[1] See *Ars Conjectandi* by Jacob Bernoulli, published posthumously in 1713.

vector $\vec{p} \in \mathbb{R}^N$, and the probability that we select one of the M items in the second urn is described by the probability vector $\vec{q} \in \mathbb{R}^M$. Then if we select an item from the first urn followed by the second urn, what is the probability that we sampled item i from the first urn *and* item j from the second urn? The answer is encoded in the tensor product $\vec{p} \otimes \vec{q}$, and in particular its $(i-1)M + j$th entry:

$$[\vec{p} \otimes \vec{q}]_{(i-1)M+j} = p_i q_j \, .$$

We can extract this entry by dotting $\vec{p} \otimes \vec{q}$ against $\vec{e}_i^{\,T} \otimes \vec{e}_j^{\,T}$, namely

$$(\vec{e}_i^{\,T} \otimes \vec{e}_j^{\,T}) \cdot (\vec{p} \otimes \vec{q}) = p_i q_j \, .$$

The vector $\vec{p} \otimes \vec{q}$ is itself a probability vector living in $\Delta_{NM} \subset \mathbb{R}^{NM}$; thus it is a probability distribution on NM outcomes, as we wanted.

So far we have examined $\vec{p} \otimes \vec{q}$ which is a product distribution, assuming in our example that our sampling from each of the two urns is uncorrelated. Below we show in an example that convex combinations of tensor products can represent a correlated, joint distribution.

Example 7. Suppose the first urn has two items ($N = 2$), say a ring and a watch, and the second urn has three items ($M = 3$), say a tissue, a match, and a rubber band. The urns were prepared by the ghost of Jacob Bernoulli. We are told that with probability $1/3$ he put a ring in the first urn *and* a rubber band in the second urn, and with probability $2/3$ he put a watch in the first urn *and* a match in the second urn. Then the joint distribution over the urns is described by

$$\frac{1}{3} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1/3 \\ 0 \\ 2/3 \\ 0 \end{bmatrix} .$$

This distribution does not factorize into a tensor product of two individual vectors.

We abstract this example in the following remark.

Remark 10 (Joint distributions and multi-index notation)**.** *Given k probability spaces represented by $\Delta_{N_i} \subset \mathbb{R}^{N_i}$ for $i = 1, ..., k$, a distribution on the joint space is represented by*

$$\Delta_{N_1 \cdots N_k} \subset \mathbb{R}^{N_1 \cdots N_k} \simeq \mathbb{R}^{N_1} \otimes \cdots \otimes \mathbb{R}^{N_k}.$$

Product (independent) distributions have the special form $\vec{p}^{\,(1)} \otimes \vec{p}^{\,(2)} \otimes \cdots \otimes \vec{p}^{\,(k)}$, and general joint distributions are convex combinations of such products. For example, if $\vec{p}_i^{\,(j)}$ represents a distribution in $\mathbb{R}^{N_j}$, then

$$\sum_{i_1, i_2, \ldots, i_k} r_{i_1 i_2 \cdots i_k} \, \vec{p}_{i_1}^{\,(1)} \otimes \vec{p}_{i_2}^{\,(2)} \otimes \cdots \otimes \vec{p}_{i_k}^{\,(k)}$$

is a joint distribution so long as $r_{i_1 i_2 \cdots i_k} \geq 0$ for all $i_1, i_2, ..., i_k$ and additionally $\sum_{i_1, i_2, \ldots, i_k} r_{i_1 i_2 \cdots i_k} = 1$. Here we have used a multi-index notation, in which we are putting subscripts on subscripts; this is to avoid notation like $\sum_{a,b,c,\ldots} r_{abc\cdots}$ which do not specify the total number of subscripts, which in our case is k. (Moreover, there are only 26 letters of the Latin alphabet.) Multi-index notation may initially

seem like gross notation, but you will soon grow accustomed to it, like generations have before you.

Joint distributions interface nicely with the $\vec{1}^T$ row vector in a number of ways. For clarity, let us write $\vec{1}_N^T$ to denote the all-ones row vector with N entries. Then we have the nice identity

$$\vec{1}_{N_1}^T \otimes \vec{1}_{N_2}^T \otimes \cdots \otimes \vec{1}_{N_k}^T = \vec{1}_{N_1 N_2 \cdots N_k}^T .$$

Thus if $\vec{p}$ is a joint distribution living in $\Delta_{N_1 N_2 \cdots N_k}$, then we have

$$(\vec{1}_{N_1}^T \otimes \vec{1}_{N_2}^T \otimes \cdots \otimes \vec{1}_{N_k}^T) \cdot \vec{p} = \vec{1}_{N_1 N_2 \cdots N_k}^T \cdot \vec{p} = 1 .$$

We can also use the all-one row vector to formulate a nice way of computing marginal distributions. To illustrate, we proceed with the example below.

Example 8. Consider a joint distribution on $\Delta_6 \subset \mathbb{R}^2 \otimes \mathbb{R}^3$. Let us denote the joint distribution by $\vec{p}_{AB}$ where A represents the first subsystem of two items, and B represents the second subsystems of three items. Then we can write $\vec{p}_{AB}$ as

$$\vec{p}_{AB} = \begin{bmatrix} p_{AB}(1,1) \\ p_{AB}(1,2) \\ p_{AB}(1,3) \\ p_{AB}(2,1) \\ p_{AB}(2,2) \\ p_{AB}(2,3) \end{bmatrix} .$$

Suppose we want to marginalize over the second probability space (the one over three items). Letting $\mathbb{1}_N$ denote the $N \times N$ identity matrix, we marvel at the linear operator $\mathbb{1}_2 \otimes \vec{1}_3^T$ which maps $\mathbb{R}^2 \otimes \mathbb{R}^3 \to \mathbb{R}^2$. We marvel at it because applying the operator to $\vec{p}_{AB}$ we find

$$(\mathbb{1}_2 \otimes \vec{1}_3^T) \cdot \vec{p}_{AB} = \begin{bmatrix} p_{AB}(1,1) + p_{AB}(1,2) + p_{AB}(1,3) \\ p_{AB}(2,1) + p_{AB}(2,2) + p_{AB}(2,3) \end{bmatrix} = \begin{bmatrix} p_A(1) \\ p_A(2) \end{bmatrix} = \vec{p}_A$$

where $\vec{p}_A$ is the marginal distribution on the first subsystem A, which has two items.

The insight in the above example generalizes in the following way.

Remark 11 (Marginalizing any subset of subsystems)**.** *Let $\vec{p} \in \Delta_{N_1 \cdots N_k}$ be a joint distribution on k subsystems with sizes $N_1, ..., N_k$. For any subset $S \subseteq \{1, ..., k\}$, define the linear "marginalization" map*

$$\mathcal{M}_S := \bigotimes_{j=1}^{k} K_j = K_1 \otimes K_2 \otimes \cdots \otimes K_k , \qquad K_j = \begin{cases} \mathbb{1}_{N_j} & \text{if } j \in S \\ \vec{1}_{N_j}^T & \text{if } j \notin S \end{cases} ,$$

and so $\mathcal{M}_S : \mathbb{R}^{N_1 \cdots N_k} \to \mathbb{R}^{Prod_{j \in S} N_j}$. Then $\mathcal{M}_S \cdot \vec{p}$ is the marginal on the subsystems indexed by S by marginalizing over the subsystems indexed by $\{1, ..., k\} \setminus S$.

To summarize, we have recast ordinary probability theory (on discrete probability spaces) in a linear-algebraic language, which has motivated us to develop the fundamentals of multi-linear algebra and tensor products. This mathematical technology certainly illuminates aspects of multi-linearity lurking in ordinary probability theory. But our true motivation was to set up probability theory in such a way as to make (finite-dimensional) quantum mechanics appear as a natural generalization, using many of the same ingredients. In this next section when

we introduce quantum mechanics, we will relentlessly capitalize on parallels with probability theory, but also take care to point out where such parallels break down.

2. Quantum theory in finite dimensions

We begin with a very brief history of quantum theory. Circa 1900 Max Planck studied blackbody radiation, and solved an inadequacy in the extant equations by stipulating that energy is quantized in units of his eponymous constant. Then in 1905, Einstein suggests that light itself is quantized as "photons", providing an explanation for the photoelectric effect. In the ensuing decade, Bohr makes a first pass at quantum theory (the so-called 'old' quantum theory), and correctly predicts the spectral lines of hydrogen. This first pass at quantum theory only goes so far, and a second pass is made in the 1920's. In 1924, de Broglie postulates that a particle with momentum p has 'wavelength' $\lambda = h/p$, which is soon confirmed by electron diffraction experiments. Thereafter, Heisenberg, Born, and Jordan developed matrix mechanics in 1925 (although they did not yet understand the connection to de Broglie). In 1926, Schrödinger leveraged de Broglie's insight to develop wave mechanics, and that same year showed the equivalence with matrix mechanics. That year as well, Born gave a 'probabilistic' interpretation of quantum mechanics which clarified its connections to measurable quantities in experiments. In 1927, Heisenberg wrote down his famous uncertainty principle. Most of the abstract mathematical foundations of quantum mechanics were consolidated by Dirac and von Neumann in the early 1930's, and Einstein-Podolsky-Rosen as well as Schrödinger highlighted the importance of entanglement in 1935. The year after in 1936, Birkoff and von Neumann investigated how quantum mechanics leads to a new form of logical reasoning that goes beyond classical Boolean logic; in hindsight this may be regarded as the first hint of the possibility of quantum computing (although it was not understood as such at the time).

Having completed our brief historical diegesis, we now turn to presenting the axioms of quantum mechanics. There are various ways of 'motivating' the axioms of quantum mechanics, although at some level they were *guessed* by very clever people and experimentally confirmed by very clever people (sometimes in the opposite order). We will, however, give some intuition. But first, a word of caution. When someone asks for a motivation for quantum mechanics in terms of classical mechanics, this is philosophically backwards; it would be like asking for a derivation of special relativity starting from Newton's equations. Indeed, just as special relativity reduces to Newtonian physics in a certain regime of validity, so too does quantum mechanics reduce to classical mechanics in a certain regime of validity. Nonetheless, we will proceed with an idiosyncratic way of 'guessing' some of the axioms of quantum mechanics starting from classical intuitions.

2.1. Mechanics on ℓ^p spaces: from classical to quantum

Let us begin by contemplating the salient mathematical structures undergirding the dynamics of probability distributions discussed above. For this, it is useful to have the following definition:

Definition 12 (Normed vector space)**.** *Let V be a vector space over a field K; we will consider either $V = \mathbb{R}^N$ (with $K = \mathbb{R}$), or $V = \mathbb{C}^N$ (with $K = \mathbb{C}$). A* ***normed***

vector space *is a pair* $(V, \|\cdot\|)$ *where* $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ *is the* ***norm*** *which satisfies the following three properties:*

(1) (Positive definiteness) $\|\vec{v}\| = 0$ *if and only if* $\vec{v}$ *is the zero vector.*
(2) (Absolute homogeneity) $\|a\,\vec{v}\| = |a|\|\vec{v}\|$ *for any* $a \in K$ *and any* $\vec{v} \in V$*.*
(3) (Triangle inequality) $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$ *for any* $\vec{v}, \vec{w} \in V$*.*

Then we can define a very useful class of norms as follows:

Definition 13 (ℓ^p norms)**.** *The* ℓ^p ***norm****, defined over* $\mathbb{R}^N$ *or* $\mathbb{C}^N$ *for* $p \geq 1$*, is*

$$\|\vec{v}\|_p := \left(\sum_{j=1}^{N} |v_j|^p\right)^{\frac{1}{p}}. \tag{5}$$

One can show that (5) is indeed a norm in the sense of Definition 12 above. (It is immediate to verify positive definiteness and absolute homogeneity; verifying the triangle inequality involves a more delicate proof leveraging Hölder's inequality.)

A special case of the ℓ^p norm is when $p = 1$, giving $\|\vec{v}\|_1 = \sum_{j=1}^{N} |v_j|$. Then when $\vec{p}$ describes a probability distribution, the normalization of probability distributions is equivalent to the condition $\|\vec{p}\|_1 = 1$. Then our characterization of Markov matrices can be equivalently phrased as follows: M is a Markov matrix if and only if

$$\|M \cdot \vec{p}\|_1 = \|\vec{p}\|_1$$

for all $\vec{p}$ describing probability distributions. In fact, using absolutely homogeneity, we also have the slightly weaker statement that M is a Markov matrix if and only if $\|M \cdot \vec{v}\|_1 = \|\vec{v}\|_1$ where all entries of $\vec{v}$ have the same sign. But then we might ask: what are the matrices A such that $\|A \cdot \vec{v}\|_1 = \|\vec{v}\|_1$ for all $\vec{v} \in \mathbb{R}^N$? Interestingly, such matrices A, called ℓ^1-isometries, are highly restricted:

Theorem 14 (ℓ^1-isometries)**.** *Let* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ *and* $A \in \mathbb{K}^{N \times N}$*. The following are equivalent:*

(1) $\|A \cdot \vec{v}\|_1 = \|\vec{v}\|_1$ *for all* $\vec{v} \in \mathbb{K}^N$*.*
(2) $A = P \cdot \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)$ *where* P *is a permutation matrix and* $|\varepsilon_j| = 1$ *for all* j *(so* $\varepsilon_j = \pm 1$ *if* $\mathbb{K} = \mathbb{R}$*).*

In the proof below, for a vector $\vec{v} = (v_1, \ldots, v_N) \in \mathbb{K}^N$ we write

$$\mathrm{supp}(\vec{v}) := \{\, k \in \{1, \ldots, N\} : v_k \neq 0 \,\}$$

for its *support*. We say two vectors have *disjoint supports* if their supports are disjoint sets.

Proof. Write $\vec{a}_j := A \cdot \vec{e}_j$ for the jth column of A. Then $\|\vec{a}_j\|_1 = \|A \cdot \vec{e}_j\|_1 = \|\vec{e}_j\|_1 = 1$.

Fix $i \neq j$. In the real case,

$$\|\vec{a}_i \pm \vec{a}_j\|_1 = \|A \cdot (\vec{e}_i \pm \vec{e}_j)\|_1 = \|\vec{e}_i \pm \vec{e}_j\|_1 = 2.$$

By the triangle inequality we always have $\|\vec{a}_i \pm \vec{a}_j\|_1 \leq \|\vec{a}_i\|_1 + \|\vec{a}_j\|_1 = 2$; equality of sums forces equality *coordinate-wise*. Thus for every coordinate k,

$$|a_i(k) \pm a_j(k)| = |a_i(k)| + |a_j(k)|.$$

For reals, the '+' equality enforces same sign (or a zero), and the '−' equality enforces opposite sign (or a zero); both can hold only if $a_i(k)a_j(k) = 0$. Hence $\mathrm{supp}(\vec{a}_i) \cap \mathrm{supp}(\vec{a}_j) = \varnothing$.

In the complex case, use

$$\|\vec{a}_i + \vec{a}_j\|_1 = \|\vec{a}_i + i\,\vec{a}_j\|_1 = 2.$$

Again equality is coordinate-wise, so with $z = a_i(k)$ and $w = a_j(k)$,

$$|z + w| = |z| + |w|, \qquad |z + i\,w| = |z| + |w|\,.$$

Each equality in $\mathbb{C}$ holds if and only if the summands share an argument; the first says z and w are collinear, the second says z and $i\,w$ are collinear. This is impossible unless $z = 0$ or $w = 0$. Thus the supports of distinct columns are disjoint in the complex case as well.

We now have N nonempty, pairwise-disjoint subsets $S_j := \mathrm{supp}(\vec{a}_j) \subseteq \{1, \ldots, N\}$. Therefore

$$N \;\le\; \sum_{j=1}^{N} |S_j| = \Big|\bigcup_{j=1}^{N} S_j\Big| \;\le\; N,$$

so $|S_j| = 1$ for all j. Hence $\vec{a}_j = \varepsilon_j\,\vec{e}_{\sigma(j)}$ for some permutation σ and some $\varepsilon_j \neq 0$. From $\|\vec{a}_j\|_1 = |\varepsilon_j| = 1$ we get $|\varepsilon_j| = 1$, and writing P for the permutation matrix of σ gives $A = P \cdot \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)$. The converse is immediate. □

Remark 15. *Equivalently, the ℓ^1-isometries are the **signed permutation matrices** when $\mathbb{K} = \mathbb{R}$ and the **monomial matrices** with unimodular entries (i.e. their absolute value equals one) when $\mathbb{K} = \mathbb{C}$. If one further assumes $A_{ij} \geq 0$, then necessarily $\varepsilon_j = 1$ for all j, so A is a permutation matrix.*

The upshot of Theorem 14 is that the only linear maps that preserve the ℓ^1 norm on all of $\mathbb{R}^N$ (or $\mathbb{C}^N$) are signed-permutation (or monomial) matrices. Thus, if we insist on global ℓ^1–isometries, the dynamics amount only to relabeling coordinates and multiplying by signs (or phases). A nontrivial theory appears when we restrict attention to the positive cone and, in particular, to the probability simplex Δ_N: requiring a linear map M to send probability vectors to probability vectors yields precisely the column–stochastic (Markov) matrices introduced above. Moreover, Theorem 14 generalizes as follows:

Theorem 16 (ℓ^p-isometries for $p \neq 2$)**.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and $A \in \mathbb{K}^{N\times N}$. Then for $p \geq 1$ and $p \neq 2$, the following are equivalent:*

(1) $\|A \cdot \vec{v}\|_p = \|\vec{v}\|_p$ for all $\vec{v} \in \mathbb{K}^N$.

(2) $A = P \cdot \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_N)$ where P is a permutation matrix and $|\varepsilon_j| = 1$ for all j (so $\varepsilon_j = \pm 1$ if $\mathbb{K} = \mathbb{R}$).

A proof can be found in [**Aar04**] (although the original proof goes back to at least Banach). The theorem above shows that for $p \neq 2$ the only linear $\|\cdot\|_p$–isometries are monomial matrices, so there is no norm-preserving linear dynamics that mixes coordinates beyond permutations (and multiplicative sign or phase factors). The case $p = 1$ is special only in that, after restricting to the positive cone, we can relax from "isometry on all vectors" to the weaker requirement "maps the probability simplex to itself"; this yields the rich class of Markov matrices. For $p > 1$ and not equal to 2, no analogous stochastic family exists. By contrast, when $p = 2$ the isometries form a continuous group providing genuinely nontrivial linear dynamics.

We have already explicated how the $p = 1$ case corresponds to classical mechanics; we will see that the $p = 2$ case corresponds to quantum mechanics.

First let us give a structure theorem for the ℓ^2-isometries. We start with the following definition.

Definition 17 (Orthogonal and unitary groups)**.** *A matrix $R \in \mathbb{R}^{N\times N}$ is an **orthogonal matrix** if and only if it satisfies $R^T R = RR^T = \mathbb{1}$. This set of matrices is closed under multiplication and inverses, and forms the **orthogonal group** $O(N)$. Similarly, a matrix $U \in \mathbb{C}^{N\times N}$ is a **unitary matrix** if and only if it satisfies $U^\dagger U = UU^\dagger = \mathbb{1}$. This set of matrices is closed under multiplication and inverses, and forms the **unitary group** $U(N)$.*

Then our structure theorem for ℓ^2-isometries is as follows.

Theorem 18 (ℓ^2-isometries)**.** *Let $R \in \mathbb{R}^{N\times N}$. The following are equivalent.*

(1) $\|R \cdot \vec{v}\|_2 = \|\vec{v}\|_2$ for all $\vec{v} \in \mathbb{R}^N$.
(2) $R \in O(N)$.

Similarly, let $U \in \mathbb{C}^{N\times N}$. The following are equivalent.

(1) $\|U \cdot \vec{v}\|_2 = \|\vec{v}\|_2$ for all $\vec{v} \in \mathbb{C}^N$.
(2) $U \in U(N)$.

We defer the proof until later, when additional mathematical tools will allow us to present it more simply.

In the same way that

$$\vec{p}\,' = M_k \cdots M_2 \cdot M_1 \cdot \vec{p}$$

for the M_i being Markov matrices constitutes ℓ^1-preserving dynamics on $\Delta_N \subset \mathbb{R}^N$, then e.g.

$$\vec{\Psi}' = U_k \cdots U_2 \cdot U_1 \cdot \vec{\Psi} \tag{6}$$

for $\vec{\Psi}, \vec{\Psi}' \in \mathbb{C}^N$ and the U_i being unitary matrices constitutes ℓ^2-preserving dynamics on $\mathbb{C}^N$. Just as probability distributions $\vec{p} \in \Delta_N \subset \mathbb{R}^N$ play a starring role in classical mechanics, the **wavefunction** plays a starring role in quantum mechanics. In its simplest form, a wavefunction is a vector $\vec{\Psi} \in \mathbb{C}^N$. (The fact that $\vec{\Psi}$ lives on $\mathbb{C}^N$ as opposed to $\mathbb{R}^N$ is an empirical fact with measurable consequences.) In particular, the wavefunction will provide a description of the 'state' of a quantum system, and so often the words 'wavefunction' and 'state' are used interchangeably.

Quantum mechanics is essentially the study of dynamics of the form (6) on $\mathbb{C}^N$, along with additional physical input that relates that dynamics to observable reality. Other physical inputs can constrain the form of the unitaries which are used. Before delving into these 'physical' considerations below, it is first worth explicating a bit more of the mathematical structure of ℓ^2 spaces, since they will be our stomping grounds for the entirety of this book.[2]

So far we have introduced the structure of an ℓ^2 norm on $\mathbb{C}^N$, in Definitions 12 and 13 (taking $p = 2$ in the latter). A nice feature of the ℓ^2 norm is that it gives us a very nice additional structure on $\mathbb{C}^N$, namely an inner product space. We define inner product spaces below, and then explain how the ℓ^2 norm allows us to define a canonical inner product space.

[2]They are also, more generally, the stomping grounds for our physical reality.

Definition 19 (Inner product and inner product space)**.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be a vector space over $\mathbb{K}$. An **inner product** on V is a map*

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$$

such that for all $u, v, w \in V$ and $a, b \in \mathbb{K}$:

(1) (Conjugate symmetry) $\langle v, w \rangle = \overline{\langle w, v \rangle}$.

(2) (Sesquilinearity) $\langle u, av + bw \rangle = a \langle u, v \rangle + b \langle u, w \rangle$ *and* $\langle au + bv, w \rangle = \overline{a} \langle u, w \rangle + \overline{b} \langle v, w \rangle$. *Equivalently, the inner product is linear in its* second *argument and conjugate–linear in its first.*[3]

(3) (Positive definiteness) $\langle v, v \rangle \geq 0$, *with equality if and only if $v = 0$.*

*A pair $(V, \langle \cdot, \cdot \rangle)$ is called an **inner product space**. The inner product induces a norm by*

$$\|v\| := \sqrt{\langle v, v \rangle}\,.$$

To fully bring you into the fold, we introduce a slightly more refined notion of inner product spaces due to Hilbert.

Definition 20 (Hilbert space)**.** *A (complex) **Hilbert space** is a complex inner product space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ that is complete*[4] *with respect to the induced norm $\|v\| = \sqrt{\langle v, v \rangle}$.*

Remark 21 (Finite-dimensional case and notation)**.** *When $\dim \mathcal{H} < \infty$, completeness is automatic, so every complex inner product space is a Hilbert space. In this book we work exclusively with finite-dimensional Hilbert spaces, typically written $\mathcal{H} \simeq \mathbb{C}^N$ equipped with the ℓ^2 inner product. We will often write $\vec{\Psi} \in \mathcal{H}$ for a state vector ("wavefunction"), and linear maps on $\mathcal{H}$ are represented by matrices; those that preserve the inner product are precisely the unitary operators $U : \mathcal{H} \to \mathcal{H}$.*

As such, an inner product space can be thought of as a normed space, with additional structure. Below we explain how the ℓ^2 norm is induced by an inner product.

Definition 22 (ℓ^2 inner product)**.** *On $\mathbb{C}^N$ we take the standard inner product to be the **ℓ^2 inner product***

$$\langle v, w \rangle := v^\dagger w = \sum_{j=1}^{N} \overline{v}_j w_j,$$

which on $\mathbb{R}^N$ reduces to $v^T w$. The induced norm is $\|v\| = \sqrt{\langle v, v \rangle} = \left(\sum_{j=1}^{N} |v_j|^2\right)^{1/2} = \|\vec{v}\|_2$, which is precisely the ℓ^2 norm.

A useful notion is (Hermitian) conjugation, which we define below.

Definition 23 (Conjugation and Hermitian adjoint)**.** *For a complex number $a \in \mathbb{C}$, its complex conjugate is a^*. For a vector $\vec{v} \in \mathbb{C}^N$, write $\vec{v}^*$ for entrywise conjugation and define the **conjugate transpose** (or **Hermitian conjugate**) by*

$$\vec{v}^\dagger := (\vec{v}^*)^T\,.$$

---

[3] This is the convention commonly used in physics. Over $\mathbb{R}$ it reduces to bilinearity.

[4] "Complete" means that every Cauchy sequence in $\mathcal{H}$ (with respect to the metric $d(u, v) = \|u - v\|$ induced by the inner product) converges to a limit *in* $\mathcal{H}$: for all $\varepsilon > 0$ there exists N such that $m, n \geq N$ implies $\|x_m - x_n\| < \varepsilon$, and there is $x \in \mathcal{H}$ with $\|x_n - x\| \to 0$. Intuitively, there are no 'holes' in the space.

For a matrix $A \in \mathbb{C}^{M\times N}$, write A^* for entrywise conjugation and define its ***Hermitian adjoint*** *(conjugate transpose) by*

$$A^\dagger := (A^*)^T \in \mathbb{C}^{N\times M}.$$

Equivalently, using the ℓ^2 inner product $\langle u, v\rangle = u^\dagger v = \sum_{j=1}^N u_j^* v_j$, we have that $A^\dagger$ is the unique linear map satisfying

$$\langle x, Ay\rangle = \langle A^\dagger x, y\rangle \quad \text{for all } x \in \mathbb{C}^M,\ y \in \mathbb{C}^N.$$

The adjoint obeys, for all compatible A, B and scalars $\alpha, \beta \in \mathbb{C}$,

$$(AB)^\dagger = B^\dagger A^\dagger, \qquad (\alpha A + \beta B)^\dagger = \alpha^* A^\dagger + \beta^* B^\dagger, \quad (A^\dagger)^\dagger = A.$$

Over $\mathbb{R}$, complex conjugation is trivial ($a^* = a$), so $A^\dagger = A^T$. Additionally, a matrix H is ***Hermitian*** *(or* ***self–adjoint****) if $H^\dagger = H$.*

Having defined the ℓ^2 inner product as well as the Hermitian adjoint, we can rephrase Theorem 18 as:

Theorem 24 (ℓ^2-isometries, reprise)**.** *Let $R \in \mathbb{R}^{N\times N}$. The following are equivalent.*

(1) $\langle R\vec{v}, R\vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle$ for all $\vec{v} \in \mathbb{R}^N$.
(2) $R \in O(N)$.

Similarly, let $U \in \mathbb{C}^{N\times N}$. The following are equivalent.

(1) $\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle$ for all $\vec{v} \in \mathbb{C}^N$.
(2) $U \in U(N)$.

With our inner product definitions at hand, we are now equipped to provide a simple proof of Theorem 24 and thus Theorem 18.

Proof. We give the argument for $\mathbb{C}^N$; the real case is analogous with † replaced by ^{T} and i replaced by ± 1.

Assume (1): $\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle$ for all $\vec{v} \in \mathbb{C}^N$. Write

$$\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, U^\dagger U\, \vec{v}\rangle,$$

so for every $\vec{v}$,

$$\langle \vec{v},\, (U^\dagger U - \mathbb{1})\, \vec{v}\rangle = 0.$$

Set $H := U^\dagger U - \mathbb{1}$. Then $\langle \vec{v}, H\vec{v}\rangle = 0$ for all $\vec{v}$. For arbitrary $\vec{x}, \vec{y}$ we compute (using conjugate–linearity in the first argument and linearity in the second):

$$\begin{aligned} 0 &= \langle \vec{x} + \vec{y},\, H(\vec{x} + \vec{y})\rangle = \langle \vec{x}, H\vec{x}\rangle + \langle \vec{x}, H\vec{y}\rangle + \langle \vec{y}, H\vec{x}\rangle + \langle \vec{y}, H\vec{y}\rangle \\ &= \langle \vec{x}, H\vec{y}\rangle + \langle \vec{y}, H\vec{x}\rangle, \\ 0 &= \langle \vec{x} + i\vec{y},\, H(\vec{x} + i\vec{y})\rangle = \langle \vec{x}, H\vec{x}\rangle + i\langle \vec{x}, H\vec{y}\rangle - i\langle \vec{y}, H\vec{x}\rangle + \langle \vec{y}, H\vec{y}\rangle \\ &= i\langle \vec{x}, H\vec{y}\rangle - i\langle \vec{y}, H\vec{x}\rangle. \end{aligned}$$

Solving these two equations gives $\langle \vec{x}, H\vec{y}\rangle = \langle \vec{y}, H\vec{x}\rangle = 0$ for all $\vec{x}, \vec{y}$. Fixing $\vec{y}$ and taking $\vec{x} = H\vec{y}$ yields $\|H\vec{y}\|^2 = 0$, so $H\vec{y} = 0$ for all $\vec{y}$ and hence $H = 0$. Therefore $U^\dagger U = \mathbb{1}$. In particular, the columns of U are orthonormal, so U is invertible and $U^{-1} = U^\dagger$; hence also $UU^\dagger = \mathbb{1}$ and $U \in U(N)$, establishing (2).

Conversely, if $U \in U(N)$ then $U^\dagger U = \mathbb{1}$, and for all $\vec{v}$,

$$\langle U\vec{v}, U\vec{v}\rangle = \langle \vec{v}, U^\dagger U\, \vec{v}\rangle = \langle \vec{v}, \vec{v}\rangle\,,$$

which is (1). This completes the proof. □

Let us pause to summarize what we have done so far in this Subsection. First, we recognized that dynamics on (finite) probability distributions is dynamics that preserves ℓ_1. We then contemplated what dynamics would look like that preserves ℓ_p for $p > 1$, and found that the only interesting option is $p = 2$, for which unitary dynamics does the job. We then explained that the ℓ_2 is produced by a natural inner product, which also interfaces nicely with unitary dynamics. Below, we will show how ℓ_2-preserving dynamics is essentially (finite-dimensional) quantum mechanics, along with some additional mathematical baggage which relates the dynamics to observable measurements. Then let us commence below with the axioms of quantum mechanics.

2.2. The axioms of quantum mechanics

Quantum mechanics was presented essentially its contemporary form by Paul Dirac in 1930 [**Dir81**] and placed on a rigorous Hilbert space footing by John von Neumann in 1932 [**VN18**]. The reader might be surprised to discover that Dirac's book [**Dir81**] remains foundational for quantum-mechanics courses nearly a century later.

2.2.1. *Bra-ket notation*

Before giving the axioms, we introduce Dirac's famous **bra-ket notation**, much beloved by physicists (and sometimes unfairly despised by mathematicians). Consider the $\mathbb{C}^N$, viewed as a Hilbert space with ℓ^2 inner product. In the future, we will simply say "consider the Hilbert space $\mathcal{H} \simeq \mathbb{C}^N$". Recall that if $\vec{\psi}, \vec{\phi} \in \mathcal{H}$ then their inner product is

$$\langle \vec{\psi}, \vec{\phi} \rangle = \sum_{j=1}^{N} \psi_j^* \phi_j = \vec{\psi}^\dagger \cdot \vec{\phi}\,.$$

The far right-hand side demonstrates that we can think of the inner product as a bilinear map from $\mathcal{H}^* \otimes \mathcal{H} \to \mathbb{C}$, where $\mathcal{H}^*$ is the space of row vectors. There is a canonical isomorphism from $\mathcal{H}$ to $\mathcal{H}^*$ given by Hermitian conjugation. This is all just a fancy way of saying the following: to take the inner product $\langle \vec{\psi}, \vec{\phi} \rangle$ of $\vec{\psi}$ and $\vec{\phi}$, we just take the Hermitian conjugate of $\vec{\psi}$ and dot that with $\vec{\phi}$.

The far left-hand side of 2.2.1 takes the notational form of a 'bracket'. Dirac suggests that we enclose vectors in $\mathcal{H}$ by $|\cdot\rangle$, so that instead of writing $\vec{\phi}$ we write $|\phi\rangle$. Such an object is called a 'ket'. In similar spirit, a column vector $\vec{\phi}^\dagger \in \mathcal{H}^*$ is enclosed by $\langle \cdot |$, so that instead of writing $\vec{\psi}^\dagger$ we write $\langle \psi |$. Such an object is called a 'bra'. Then bras and kets are related via Hermitian conjugation, namely

$$|\psi\rangle^\dagger = \langle \psi |\,.$$

Finally, we can put together bras and kets to form

$$\langle \psi | \phi \rangle := \langle \vec{\psi}, \vec{\phi} \rangle = \sum_{j=1}^{N} \psi_j^* \phi_j = \vec{\psi}^\dagger \cdot \vec{\phi}\,,$$

which is a...(drum roll please) 'bra-ket'! Get it?[5]

[5]Famously, Dirac was not known for his sense of humor.

Besides being somewhat whimsical, Dirac's bra-ket notation is in fact extremely useful. The reason is not so much mathematical, but rather visual. As you yourself will experience, bra-ket notation is visually suggestive of how to organize and manipulate certain equations (especially compared with arrows and daggers), and eases the mind towards simplifying complicated expressions in multi-linear algebra. That is, Dirac found a notation which resonates with the structure of our minds.

Let us develop Dirac's notation a bit further. In addition to forming 'inner products' $\langle\psi|\phi\rangle = \vec{\psi}^\dagger \cdot \vec{\phi}$, we can also form 'outer products' $|\phi\rangle\langle\psi| = \vec{\phi} \cdot \vec{\psi}^\dagger$. Here $|\phi\rangle\langle\psi|$ is evidently a rank 1, $N \times N$ matrix. Then the trace of this matrix is evidently

$$\mathrm{tr}(|\phi\rangle\langle\psi|) = \langle\psi|\phi\rangle\,.$$

Since Hermitian conjugation for a scalar is the same as complex conjugation, we have the useful identity

$$(\langle\psi|\phi\rangle)^\dagger = (\langle\psi|\phi\rangle)^* = \langle\phi|\psi\rangle\,,$$

where we observe that the ψ and ϕ have switched sides.

It is useful to show a few examples to get you fully acquainted with bra-ket notation. Consider the standard orthonormal basis $\{\vec{e}_i\}_{i=1}^N$ of $\mathbb{C}^N$, which we denote by $\{|i\rangle\}_{i=1}^N$ in our new notation. The orthonormality of the basis elements can be expressed as

$$\langle i|j\rangle = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

where δ_{ij} is called the **Kronecker delta**. Recalling that the identity matrix is $\mathbb{1} = \sum_{i=1}^N \vec{e}_i \cdot \vec{e}_i^T$, in bra-ket notation we have

$$\mathbb{1} = \sum_{i=1}^N |i\rangle\langle i|\,.$$

Then given a state $|\psi\rangle$, we have

$$|\psi\rangle = \mathbb{1}|\psi\rangle = \left(\sum_{i=1}^N |i\rangle\langle i|\right)|\psi\rangle = \sum_{i=1}^N |i\rangle \underbrace{\langle i|\psi\rangle}_{=:\psi_i} = \sum_{i=1}^N \psi_i |i\rangle\,, \tag{7}$$

where ψ_i are the coefficients of $|\psi\rangle$ in the $|i\rangle$-basis. (Note also that $(\langle i|\psi\rangle)^\dagger = \langle i|\psi\rangle)^* = \langle\psi|i\rangle = \psi_i^*$, and so the coefficients of $\langle\psi|$ in the $\langle i|$-covector basis are ψ_i^*.) Similarly, for a matrix M, we have

$$M = \mathbb{1}\cdot M\cdot\mathbb{1} = \left(\sum_{i=1}^N |i\rangle\langle i|\right) M \left(\sum_{j=1}^N |j\rangle\langle j|\right) = \sum_{i,j=1}^N |i\rangle \underbrace{\langle i|M|j\rangle}_{=:M_{ij}}\langle j| = \sum_{i,j=1}^N M_{ij}\,|i\rangle\langle j|\,, \tag{8}$$

where M_{ij} are the matrix elements of M in the $|i\rangle$-basis. As a check of our notation, let us compute $M|\psi\rangle$ using the far-right hand sides of both (7) and (8):

$$M|\psi\rangle = \left(\sum_{i,j=1}^N M_{ij}\,|i\rangle\langle j|\right)\sum_{k=1}^N \psi_k|k\rangle = \sum_{i,j,k=1}^N M_{ij}\psi_k|i\rangle\underbrace{\langle j|k\rangle}_{=\delta_{jk}} = \sum_{i=1}^N \left(\sum_{j=1}^N M_{ij}\psi_j\right)|i\rangle\,.$$

So we see that the coefficients of $M|\psi\rangle$ in the $|i\rangle$-basis are $\sum_{j=1}^N M_{ij}\psi_j$, exactly as expected using the standard rules of matrix multiplication.

For our final flourish, we present the **spectral theorem** in finite dimensions in bra-ket notation. The spectral theorem will play a crucial role in the formulation of quantum mechanics.

Theorem 25 (Spectral theorem for normal matrices in finite dimensions)**.** *Let $A : \mathcal{H} \to \mathcal{H}$ be a linear operator on a finite dimensional complex Hilbert space $\mathcal{H} \simeq \mathbb{C}^N$. The following are equivalent*

(1) A is normal, meaning $A^\dagger A = AA^\dagger$.

(2) There exists an orthonormal basis of eigenstates $|v_1\rangle, \ldots, |v_N\rangle$ and complex numbers $\lambda_1, \ldots, \lambda_N$ such that $A = \sum_{j=1}^N \lambda_j |v_j\rangle\langle v_j|$ Equivalently, if U is the unitary with columns $|v_j\rangle$ then $U^\dagger A U = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$.

We will break up the proof into a few lemmas:

Lemma 26. *Let $A : \mathcal{H} \to \mathcal{H}$ be a normal matrix for $\mathcal{H} \simeq \mathbb{C}^N$. Then A has at least one eigenvector $|v\rangle$. Moreover, if $A|v\rangle = \lambda|v\rangle$, then $A^\dagger|v\rangle = \lambda^*|v\rangle$.*

Proof. By the fundamental theorem of algebra the characteristic polynomial $p_A(\lambda) = \det(A - \lambda\mathbb{1})$ has a complex root. If λ is such a root, then $A - \lambda\mathbb{1}$ has a non-trivial nullspace, meaning that A has an eigenvalue λ and at least one nonzero eigenstate $|v\rangle$ with $A|v\rangle = \lambda|v\rangle$. Without loss of generality we take $|v\rangle$ to be normalized so that $\langle v|v\rangle = 1$. Now notice that

$$\underbrace{\langle v|A^\dagger}_{=\lambda^*\langle v|} |v\rangle = \lambda^* \,. \tag{9}$$

Recall that the Cauchy-Schwarz inequality $|\langle\psi|\phi\rangle| \leq \sqrt{\langle\psi|\psi\rangle}\sqrt{\langle\phi|\phi\rangle}$ achieves equality only when $|\psi\rangle$ is proportional to $|\phi\rangle$. Assuming A is normal, we have

$$\begin{aligned} |\lambda| &= |\langle v|A^\dagger|v\rangle| \\ &\leq \sqrt{\langle v|v\rangle}\sqrt{\langle v|A^\dagger A|v\rangle} \\ &\leq \sqrt{\langle v|AA^\dagger|v\rangle} \\ &= |\lambda| \,, \end{aligned}$$

where we have used Cauchy-Schwarz in the first equality and normality of A in the equality thereafter. We thus see that Cauchy-Schwarz is tight in the above setting, implying that $A^\dagger|v\rangle$ is proportional to $|v\rangle$. In light of (9), we find that $A^\dagger|v\rangle = \lambda^*|v\rangle$, and so $|v\rangle$ is an eigenstate of $A^\dagger$ with eigenvalue λ^*. □

Lemma 27. *Let $A : \mathcal{H} \to \mathcal{H}$ be a normal matrix for $\mathcal{H} \simeq \mathbb{C}^N$. If A has two eigenvectors $|v\rangle, |w\rangle$ with distinct eigenvalues λ, μ, then $\langle v|w\rangle = 0$, i.e. $|v\rangle$ and $|w\rangle$ are orthogonal.*

Proof. Without loss of generality we can take $\langle v|v\rangle = \langle w|w\rangle = 1$. Using Lemma 26, $A^\dagger|w\rangle = \mu^*|w\rangle$. Then

$$(\lambda - \mu)\langle v|w\rangle = \langle (A - \mu\mathbb{1})v \,|\, w\rangle = \langle v \,|\, (A^\dagger - \mu^*\mathbb{1})w\rangle = 0 \,,$$

and so we find $\langle v|w\rangle = 0$. Thus eigenstates with distinct eigenvalues are orthogonal. □

Lemma 28 (Invariance of an eigenspace and its orthogonal complement)**.** *Let A be a normal operator on a finite–dimensional complex Hilbert space $\mathcal{H}$ and let*

$$E_\lambda := \ker\big(A - \lambda\,\mathbb{1}\big)$$

be the λ–eigenspace of A. Then E_λ and $E_\lambda^\perp$ are each invariant under both A and $A^\dagger$. In particular, the restriction

$$A\big|_{E_\lambda^\perp}$$

is normal on the Hilbert space $E_\lambda^\perp$.

Proof. By Lemma 26, if $|y\rangle \in E_\lambda$ then $A^\dagger|y\rangle = \lambda^*|y\rangle$. Now let $|x\rangle \in E_\lambda^\perp$ and $|y\rangle \in E_\lambda$. Then we have $\langle y|A|x\rangle = \langle A^\dagger y|x\rangle = \lambda^*\langle y|x\rangle = 0$. Since $\langle y|A|x\rangle = 0$ for every $|y\rangle \in E_\lambda$, we have $A|x\rangle \in E_\lambda^\perp$. Thus A leaves $E_\lambda^\perp$ invariant. The same calculation with A and $A^\dagger$ interchanged shows $A^\dagger$ leaves $E_\lambda^\perp$ invariant as well. Trivially A leaves E_λ invariant and from the first step $A^\dagger$ leaves E_λ invariant too.

Finally set $B := A\big|_{E_\lambda^\perp}$. Since both A and $A^\dagger$ leave $E_\lambda^\perp$ invariant, the adjoint of B with respect to the inner product on $E_\lambda^\perp$ is $B^\dagger = A^\dagger\big|_{E_\lambda^\perp}$. Hence

$$B^\dagger B = \big(A^\dagger A\big)\big|_{E_\lambda^\perp} = \big(AA^\dagger\big)\big|_{E_\lambda^\perp} = BB^\dagger,$$

so B is normal on $E_\lambda^\perp$. □

With the above lemmas at hand, we finally turn to the proof of Theorem 25.

Proof of Theorem 25. We prove (1) ⇒ (2) by induction on N. The case $N = 1$ is immediate. Assume the claim holds for all dimensions smaller than N.

By Lemma 26 the operator A has an eigenvalue λ and a nonzero eigenstate. Let $E_\lambda = \ker(A - \lambda\,\mathbb{1})$ and choose an orthonormal basis $\{|v_1\rangle, \ldots, |v_r\rangle\}$ of E_λ. By Lemma 28 the orthogonal complement $E_\lambda^\perp$ is invariant under both A and $A^\dagger$. Hence the restriction

$$B := A\big|_{E_\lambda^\perp}$$

is a normal operator on the Hilbert space $E_\lambda^\perp$ whose dimension is $N - r$. By the induction hypothesis there exists an orthonormal basis $\{|v_{r+1}\rangle, \ldots, |v_N\rangle\}$ of $E_\lambda^\perp$ consisting of eigenstates of B, hence of A. Together with $\{|v_1\rangle, \ldots, |v_r\rangle\}$ this gives an orthonormal eigenbasis of $\mathcal{H}$. Writing A in this basis yields

$$A = \sum_{j=1}^{N} \lambda_j\, |v_j\rangle\langle v_j|\,,$$

with $\lambda_j = \lambda$ for $j \le r$ and λ_j equal to the eigenvalues of B for $j > r$. This proves (2).

For (2) ⇒ (1) we compute

$$A^\dagger = \sum_{j=1}^{N} \lambda_j^*\, |v_j\rangle\langle v_j| \quad \text{and} \quad A^\dagger A = \sum_{j=1}^{N} |\lambda_j|^2\, |v_j\rangle\langle v_j| = AA^\dagger\,,$$

and so A is normal. This completes the proof. □

Remark 29 (Hermitian and unitary cases)**.** *If $A = A^\dagger$ then every λ_j is real and $A = \sum_j \lambda_j |v_j\rangle\langle v_j|$. If A is unitary then every λ_j has $|\lambda_j| = 1$ and $A = \sum_j e^{i\theta_j} |v_j\rangle\langle v_j|$*

As an application, consider the following definition.

Definition 30 (Projector)**.** *A **projector** P on $\mathcal{H}$ is a Hermitian idempotent: $P = P^\dagger = P^2$. Equivalently, $P \succeq 0$ and its eigenvalues lie in $\{0, 1\}$.*

Hermiticity implies that all of the eigenvalues of P are real and positive semi-definiteness implies that all of the eigenvalues are nonnegative. Then $P^2 = P$ means that the eigenvalues are either 0 or 1. Supposing $\mathcal{H} \simeq \mathbb{C}^N$, we can use the spectral decomposition to write P as

$$P = \sum_{i=1}^{r} 1 \cdot |v_i\rangle\langle v_i| + \sum_{i=r+1}^{N} 0 \cdot |v_i\rangle\langle v_i| = \sum_{i=1}^{r} |v_i\rangle\langle v_i|$$

for some orthonormal basis $\{|v_i\rangle\}_{i=1}^N$, where r is the rank of the projector. Then P is a projector onto the r-dimensional subspace of $\mathcal{H}$ spanned by $\{|v_i\rangle\}_{i=1}^r$. We can check that $P^\perp = \mathbb{1} - P$ is also a projector onto the orthogonal complement.

Having covered the essence of bra-ket notation, we turn to presenting the axioms of quantum mechanics a la Dirac (with some refinements).

2.2.2. *The axioms*

Here we give the standard axioms of quantum mechanics, with some commentary. The axioms describe the basic mathematical objects of quantum theory, and tether them to observable reality. In the form presented below, the axioms may seem somewhat abstract, and we will discuss this unusual feature shortly. We have tailored the axioms to the finite-dimensional setting for clarity.

(1) **(Quantum states fully describe a system at fixed time.)** At a fixed moment in time, a quantum system about which we have maximal information is fully described by some state vector $|\psi\rangle$ with unit norm living in a Hilbert space $\mathcal{H}$.

(2) **(Time evolution of a closed system is unitary.)** If a quantum system is closed (i.e. it is not interacting with any external system) and starts in an initial state $|\psi_0\rangle$, then at any later time T the state $|\psi_T\rangle$ will be related to the original one by some unitary, that is $|\psi_T\rangle = U|\psi_0\rangle$ for some unitary U that may depend on T.

(3) **(Physical properties have associated projectors.)** Any measurable physical property (such as "spin-up along the z-axis", or "the particle is in region R") has an associated projector P. Such an operator P is an example of an **observable**.

(4) **(Measurement and the Born rule.)** Suppose we have a property corresponding to a projector P, and measure whether or not a system with state vector $|\psi\rangle$ (with unit norm) has that property. Then the probability that we measure $|\psi\rangle$ to have the given property is $\langle\psi|P|\psi\rangle$. This is called the **Born rule**. If $|\psi\rangle$ is measured to have the property, then after measurement the new state of the system is

$$\frac{P|\psi\rangle}{\sqrt{\langle\psi|P^\dagger P|\psi\rangle}} = \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}},$$

which also has unit norm (assuming $P|\psi\rangle \neq 0$, in which case we would never have measured $|\psi\rangle$ to have the given property anyway.)

Now we have a number of comments to make in order to unpack the axioms. The first two axioms were motivated by our previous discussions, in which quantum mechanics is framed as norm-preserving dynamics on ℓ^2. The first axiom codifies that a (normalized) vector in a Hilbert space contains everything there is to know about a quantum state, and the second axioms explains that the dynamics of an isolated system is described by unitary dynamics. Unitary dynamics is reversible since (i.e. unitary matrices are invertible), and so in a closed system the future is completely determined by the past and the past is completely determined by the future. An interesting feature of the second axiom is that it does not tell us *which* unitaries we should use. Indeed, given a classical system, we might wonder what kinds of quantum unitary dynamics can reproduce the the classical dynamics in the appropriate regime. This is a subtle question which goes beyond the axioms, and requires additional empirical input.

The first axiom's proviso "about which we have maximal information" deserves explanation. Consider flipping an unbiased coin to decide whether to prepare a system in state $|\psi_0\rangle$ or $|\psi_1\rangle$. After the flip, the system is in state $|\psi_0\rangle$ with probability $1/2$ or state $|\psi_1\rangle$ with probability $1/2$. This probabilistic description reflects our classical ignorance, not any fundamental quantum uncertainty. The system is definitely in one state or the other; *we* simply do not know which. There is a useful formalism for handling such incomplete knowledge, which we will introduce later.

The second axiom's restriction to "closed" systems is similarly important. A closed system does not interact with any external environment. If such interactions were present, we would need to account for our incomplete knowledge of the environment, which we will address later. When a system couples to an external environment, its dynamics can become non-unitary: information leaks irreversibly from our system into the environment, where it becomes inaccessible to us. Despite being non-unitary, these dynamics can be nicely characterized.

While the first and second axioms specify the basic mathematical objects at play, the third and fourth axioms tether those mathematical objects to empirical reality. This is differently structured than e.g. Newton's axioms of classical mechanics, which specify properties like position and momentum but do not explain what it means to measure them, or how to do so.[6]

Now we turn to the third axiom. The third axiom assigns yes/no properties of a quantum system to linear subspaces of the Hilbert space, via projectors onto those subspaces. For instance, the property 'the spin points up in the z-direction' corresponds to some projector P. The opposite property corresponds to the projector $P^\perp = \mathbb{1} - P$ onto the orthogonal complement. If we have a collection of properties corresponding to projectors $P_1, \ldots, P_k$, we call them **compatible** if the corresponding subspaces are mutually orthogonal, i.e. $P_i P_j = 0$ for $i \neq j$. This orthogonality implies $[P_i, P_j] = 0$. Under orthogonality, if a state answers 'yes' to one property

[6]Part of the reason is that position and momentum, at least in some informal sense, were already known to empiricists in Newton's time. Thus people already knew how to measure them. Interestingly, as we all know, one can use Newton's laws to build devices to better measure position and momentum. You might wonder if this would lead to a circular argument: can we use devices, built using principles from Newton's laws, to then do experiments to test Newton's laws? In short, the answer is 'yes', if we (correctly) conceive of such experiments as testing the *consistency* of Newton's laws with empirical reality. Indeed, since measurements of quantities in Newton's theory require Newton's theory for their specification and possibly design, and there is no clear sense in which one can use empirical findings to test Newton's laws *ex nihilo*.

(i.e. $P_i|\psi\rangle = |\psi\rangle$), it automatically answers 'no' to all others (i.e. $P_j|\psi\rangle = 0$ for $j \neq i$). The projectors $P_1, \ldots, P_k$ are **complete** if their corresponding subspaces span all of $\mathcal{H}$, which is equivalent to $P_1 + \cdots + P_k = \mathbb{1}$. Completeness means that a state will always answer 'yes' to at least one property. Then compatibility and completeness together mean that the state will answer 'yes' to exactly one property in the list (and thus 'no' to all others in the list). The following remark captures some useful nomenclature.

Remark 31 (Hermitian observables)**.** *Let $P_1, \ldots, P_k$ correspond to compatible and complete properties. Suppose that my detector registers the real number a_j to indicate 'yes' for property j. (For instance, if the jth property is 'the particle is at position j', then the detector might just output the number j for the position.) Then we can construct the Hermitian observable*

$$A = \sum_{j=1}^{k} a_j \, P_j \tag{10}$$

which encodes measurement outcomes with respect to our list of properties. In particular,

$$\langle\psi|A|\psi\rangle = \sum_{j=1}^{k} a_j \, \langle\psi|P_j|\psi\rangle = \sum_{j=1}^{k} a_j \, \text{Prob}[\text{measure outcome } j] \, ,$$

*where in the last equality we used the Born rule from the fourth axiom. The resulting number is the expectation value of the output of our detector. Since by the spectral theorem all Hermitian operators A can be written in the form (10) for some choice of compatible and complete properties, we call Hermitian operators **observables**, with the understanding that their physical interpretation in terms of properties comes from their spectral decomposition.*

A consequence of our discussion above is that certain properties may be *incompatible*, i.e. correspond to non-orthogonal subspaces. For instance, properties corresponding to projectors P and Q are said to be incompatible if $[P, Q] \neq 0$. In this case the two measurements do not admit a common eigenbasis, so in general one cannot ascribe sharp values to both properties simultaneously. Typically, if a state has a definite value for the property corresponding to P, then measuring the property corresponding to Q will yield (in light of the fourth axiom) probabilistic results, and the act of measurement can disturb the system so that P is no longer definite. This lack of joint sharpness is the essence of incompatibility, underlies the uncertainty principle, and is one of the distinguishing features of quantum mechanics vis-à-vis classical mechanics.

The fourth axiom is, in a sense, the most mysterious. While the third axioms abstractly explains the relationship between properties of a system and the quantum state of a system, the fourth axiom tethers these properties to probabilistic observable outcomes. To begin, recall that we said that a state $|\psi\rangle$ has the property corresponding to P if $P|\psi\rangle = |\psi\rangle$, and so not have the property if $(\mathbb{1} - P)|\psi\rangle = P^{\perp}|\psi\rangle = |\psi\rangle$ (or equivalently $P|\psi\rangle = 0$). So far we have accounted for the possibilities $P|\psi\rangle = |\psi\rangle$ or 0, but if $|\psi\rangle$ is neither in the subspace corresponding to P or orthogonal to it, then $P|\psi\rangle \neq |\psi\rangle$ and $\neq 0$. The Born rule tells us that we should interpret the norm squares of the projection of $|\psi\rangle$ into P, namely $\langle\psi|P^{\dagger}P|\psi\rangle = \langle\psi|P|\psi\rangle$, as the probability that $|\psi\rangle$ has that property. More peculiar

is that when we affirmatively measure $|\psi\rangle$ to have that property, the $|\psi\rangle$ assumes the new state $\frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}}$. This state now *has* the property, since

$$P \cdot \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}} = \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}}$$

Said another way, if we measure a state to affirmatively have a property (whether or not it definitely had the property before), it subsequently *assumes* that property. This is different from classical mechanics: for example, classical mechanics stipulates that if we measure a particle to have position x then it definitely had position x before. In quantum mechanics, by contrast, measuring a particle to be in position x just tells us that the particle is in position x now, even though it might not 'definitively' have had that property before.

We notice another peculiarity of the fourth axiom, which is that the map

$$|\psi\rangle \longmapsto \frac{P|\psi\rangle}{\sqrt{\langle\psi|P|\psi\rangle}} \tag{11}$$

is not in general unitary (unless $P = \mathbb{1}$ in which case the map is the identity since $|\psi\rangle$ has unit norm). This would appear to violate the second axiom, which necessitates unitary dynamics. However, we were careful in the second axiom to specify that unitary dynamics happens for *closed* systems; in ordinary circumstances, the measurement apparatus is external to the system that it interrogates, and so the non-unitary of (11) is not in conflict with the second axiom. However, the fourth axiom tempts us to consider the following: if we described the detector (which itself is quantum-mechanical) as *part of* the closed system, then the total detector-system dynamics must be unitary; then can the fourth axiom somehow be derived from the other three? This question is both challenging and profound. Its core difficulty is that the first three axioms do not speak of probability whereas the fourth axioms does speak of probability; as such, the question posed would mandate that probability is *emergent* in quantum mechanics. There have been a vast number of attempts to weaken the fourth axiom or to in some sense 'derive' it from the other three (which often involves covertly bringing in a weakening of the fourth axiom anyway). For our purposes, we can think of the fourth axiom is *pragmatic*, in that it tells us what happens, *in practice*, when we measure a quantum system with an external measurement device.[7]

Having abstractly discussed the axioms, some examples are in order.

Example 9 (Dynamics and projective measurements for a single qubit). We work in the two-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$ with the *computational basis*

[7]Related to the previous footnote, we might wonder how we can test quantum mechanics as a theory if we require quantum theory to build the measurement apparatus needed for the tests themselves. As before, the answer is that we are testing the *consistency* of quantum mechanics, and its alignment with empirical reality. One cannot generally test quantum mechanics with detectors solely intelligible through Newtonian mechanics, i.e. you cannot solely use classical to test quantum (see [**Mah18**] for a quantum cryptographic wrinkle in this story). But it is fine to use quantum to test quantum, so long as it all works out empirically. And it very much does.

$\{|0\rangle, |1\rangle\}$ where $|0\rangle := \begin{bmatrix}1\\0\end{bmatrix}$ and $|1\rangle := \begin{bmatrix}0\\1\end{bmatrix}$. We introduce the **Pauli matrices**[8]

$$X = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} = |0\rangle\langle 1| + |1\rangle\langle 0|$$
$$Y = \begin{bmatrix}0 & -i\\ i & 0\end{bmatrix} = -i\,|0\rangle\langle 1| + i\,|1\rangle\langle 0|$$
$$Z = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} = |0\rangle\langle 0| - |1\rangle\langle 1|.$$

They are Hermitian, satisfy $X^2 = Y^2 = Z^2 = \mathbb{1}$, and obey

$$[\sigma_j, \sigma_k] = 2i\,\varepsilon_{jk\ell}\,\sigma_\ell, \qquad \{\sigma_j, \sigma_k\} = 2\,\delta_{jk}\,\mathbb{1},$$

where $(\sigma_1, \sigma_2, \sigma_3) = (X, Y, Z)$. Their eigenvalues are ± 1, with $Z|0\rangle = |0\rangle$ and $Z|1\rangle = -|1\rangle$.

Measuring "spin along z" corresponds to the compatible, complete pair of projectors

$$P_0 = |0\rangle\langle 0| = \frac{\mathbb{1} + Z}{2}, \qquad P_1 = |1\rangle\langle 1| = \frac{\mathbb{1} - Z}{2}.$$

Likewise, "spin along x" has eigenstates $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ with projectors

$$P_+^{(x)} = |+\rangle\langle +| = \frac{\mathbb{1} + X}{2}, \qquad P_-^{(x)} = |-\rangle\langle -| = \frac{\mathbb{1} - X}{2}.$$

More generally, for any unit vector $\hat{n} = (n_x, n_y, n_z) \in \mathbb{R}^3$ we have

$$P_\pm^{(\hat{n})} = \frac{\mathbb{1} \pm \hat{n}\cdot\vec{\sigma}}{2}, \qquad \hat{n}\cdot\vec{\sigma} := n_x X + n_y Y + n_z Z,$$

which indeed satisfy the properties of projectors.

For dynamics, consider unitary rotations generated by the Pauli matrices. For any unit vector $\hat{n}$ and real angle θ, define

$$R_{\hat{n}}(\theta) := \exp\!\left(-\,i\,\frac{\theta}{2}\,\hat{n}\cdot\vec{\sigma}\right) = \cos\!\left(\frac{\theta}{2}\right)\mathbb{1} - i\,\sin\!\left(\frac{\theta}{2}\right)\hat{n}\cdot\vec{\sigma}.$$

Physically, $R_{\hat{n}}(\theta)$ is the time-t propagator of a closed qubit with Hamiltonian $H = \frac{\Omega}{2}\,\hat{n}\cdot\vec{\sigma}$ and $\theta = \Omega\,t$. That is, $R_{\hat{n}}(\theta)$ can be written as e^{-iHt} for the above choices of H and t.

Suppose we prepare the qubit in the $+1$ eigenstate of Z, namely $|\psi_0\rangle = |0\rangle$. If the system evolves under the Hamiltonian $H = \frac{\Omega}{2}Y$ for time t, the unitary $U(t) = R_y(\theta)$ acts with $\theta = \Omega\,t$. Acting on $|0\rangle$ and using $Y|0\rangle = i|1\rangle$, the evolved state is

$$|\psi_t\rangle = \cos\!\left(\tfrac{\theta}{2}\right)|0\rangle + \sin\!\left(\tfrac{\theta}{2}\right)|1\rangle.$$

Now consider measuring in the Z basis. The Born rule with projectors P_0, P_1 gives

$$p_Z(0\,|\,t) = \cos^2\!\left(\tfrac{\theta}{2}\right), \qquad p_Z(1\,|\,t) = \sin^2\!\left(\tfrac{\theta}{2}\right).$$

[8]The hardest one to remember is Y, in particular the placement of the minus sign in the matrix elements. High energy physicist Howard Georgi has a useful mnemonic: the 'minus i' is lighter so it floats all the way to the top. Now hopefully you will never forget where the minus sign goes.

If outcome 0 is observed, the state collapses to $|0\rangle$; if outcome 1 is observed, it collapses to $|1\rangle$.

If instead we measure in the X basis, the probabilities are

$$p_X(\pm \,|\, t) = \tfrac{1}{2}(1 \pm \langle\psi_t|X|\psi_t\rangle)\,.$$

Since $\langle\psi_t|X|\psi_t\rangle = \sin\theta$, we find

$$p_X(+\,|\,t) = \tfrac{1+\sin\theta}{2}, \qquad p_X(-\,|\,t) = \tfrac{1-\sin\theta}{2}.$$

To connect with the Bloch sphere, define for any $|\psi\rangle$ the triple

$$\vec{r} = (\langle X\rangle, \langle Y\rangle, \langle Z\rangle) \in \mathbb{R}^3.$$

For the state $|\psi_t\rangle$, we obtain $\vec{r}(t) = (\sin\theta, 0, \cos\theta)$, a unit vector rotating about the y-axis. The Born rule in this language becomes

$$\Pr[\text{outcome } \pm \text{ along } \hat{n}] = \frac{1 \pm \hat{n}\cdot\vec{r}}{2}.$$

Armed with our basic examples, we next examine some additional mathematical structures in quantum mechanics.

2.3. Additional mathematical structures

Here we will introduce some additional mathematical apparatus which we can view as additional tools for the applications of the axioms of quantum mechanics presented above.

2.3.1. *Tensor products and density matrices*

We now carry the tensor–product technology into the quantum setting and introduce the operator language that lets us handle classical uncertainty and open–system effects in a clean way. When two systems are modeled by Hilbert spaces $\mathcal{H}_A \simeq \mathbb{C}^{N_A}$ and $\mathcal{H}_B \simeq \mathbb{C}^{N_B}$, their composite is described by the tensor product

$$\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^{N_A N_B}.$$

Choose orthonormal bases $\{|i\rangle_A\})_{i=1}^{N_A}$ and $\{|j\rangle_B\}_{j=1}^{N_B}$. The product kets $\{|i\rangle_A \otimes |j\rangle_B\}_{i,j}$ form an orthonormal basis of $\mathcal{H}_{AB}$. As in the classical case, linear maps respect tensoring. If X_A acts on $\mathcal{H}_A$ and Y_B acts on $\mathcal{H}_B$, then

$$(X_A \otimes Y_B)\big(|\psi\rangle_A \otimes |\phi\rangle_B\big) = (X_A|\psi\rangle_A) \otimes (Y_B|\phi\rangle_B).$$

Operations on a single part are written $X_A \otimes \mathbb{1}_B$ or $\mathbb{1}_A \otimes Y_B$.

A pure state $|\Psi\rangle \in \mathcal{H}_{AB}$ is called a **product state** if it factors as $|\Psi\rangle = |\psi\rangle_A \otimes |\phi\rangle_B$. Otherwise it is **entangled**. The following normal form is indispensable.

Theorem 32 (Schmidt decomposition)**.** *For any unit vector $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ there exist orthonormal sets $\{|k\rangle_A\}$ and $\{|k\rangle_B\}$ together with nonnegative numbers $\{\lambda_k\}$ that sum to one such that*

$$|\Psi\rangle = \sum_{k=1}^{r} \sqrt{\lambda_k}\,|k\rangle_A \otimes |k\rangle_B, \quad r \le \min\{N_A, N_B\}\,.$$

*The number r is uniquely defined and is called the **Schmidt rank**.*

This is really just another way of stating the linear algebraic fact that every $N_A \times N_B$ matrix (in this case the entries of $|\Psi\rangle$ reshaped into such a matrix) has a singular value decomposition, so we defer the proof until a bit later.

Up to this point, our description of a single system has used a unit vector $|\psi\rangle$. That choice corresponds to maximal information. In many situations there is additional classical uncertainty. Perhaps a device prepares $|\psi_j\rangle$ with probability r_j. It is convenient to package such ensembles into a single object, the **density operator** (or **density matrix**)

$$\rho := \sum_j r_j \, |\psi_j\rangle\langle\psi_j| \in \mathcal{S}(\mathcal{H}), \tag{12}$$

which is Hermitian, positive semidefinite, and satisfies $\mathrm{tr}(\rho) = 1$. In fact, any operator which is Hermitian, positive semidefinite, and satisfies $\mathrm{tr}(\rho) = 1$ can be written in the form (12), and so we define:

Definition 33 (Density operator)**.** *A* ***density operator*** $\rho \in \mathcal{S}(\mathcal{H})$ *is a linear operator on* $\mathcal{H}$ *which satisfies* $\rho = \rho^\dagger$*,* $\mathrm{tr}(\rho) = 1$*, and* $\rho \succeq 0$*.*

We say that a state is **pure** when $\rho = |\psi\rangle\langle\psi|$, equivalently $\rho^2 = \rho$ and $\mathrm{tr}(\rho^2) = 1$, and otherwise it is **mixed**. A pure state corresponds to a rank 1 density matrix, and a mixed state corresponds to rank greater than 1. The Born rule extends linearly. Specifically, for a projector P,

$$\Pr[\text{“yes” on } P \text{ given } \rho] = \mathrm{tr}(P\rho),$$

and for an observable A,

$$\mathbb{E}_\rho[A] = \mathrm{tr}(A\rho).$$

Upon a projective measurement with projectors P_j, two kinds of updates occur. If we condition on the outcome j, then

$$\rho \longmapsto \frac{P_j \rho P_j}{\mathrm{tr}(P_j \rho)}.$$

If the outcome is forgotten, then

$$\rho \longmapsto \sum_j P_j \rho P_j,$$

which removes coherences between the corresponding subspaces.

Joint states admit a notion of marginalization that mirrors our classical $\vec{1}^T$ trick. Given ρ_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$, the state of A alone is the **partial trace** over B:

$$\rho_A := \mathrm{tr}_B(\rho_{AB}) \in \mathcal{S}(\mathcal{H}_A).$$

In coordinates with respect to any orthonormal basis $\{|j\rangle_B\}$,

$$\mathrm{tr}_B(\rho_{AB}) = \sum_j \big(\mathbb{1}_A \otimes \langle j|\big)\, \rho_{AB}\, \big(\mathbb{1}_A \otimes |j\rangle\big). \tag{13}$$

The map tr_B is characterized by the identity

$$\mathrm{tr}\big[(X_A \otimes \mathbb{1}_B)\rho_{AB}\big] = \mathrm{tr}\big[X_A \,\mathrm{tr}_B(\rho_{AB})\big] \quad \text{for all } X_A\,,$$

so it really is the quantum version of taking a marginal. If ρ_{AB} is diagonal in the product basis, (13) reduces exactly to summing out the B index. The identity $\mathrm{tr}_B(\rho_{AB}) = \rho_A$ is the quantum sibling of marginalization by dotting probability vectors with with $\vec{1}^T$, as appeared in our earlier discussion.

Two corollaries are immediate from the Schmidt decomposition. First, if $|\Psi\rangle$ is a pure vector on AB and $\rho_{AB} = |\Psi\rangle\langle\Psi|$, then $\rho_A = \mathrm{tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{tr}_A(\rho_{AB})$ share the same nonzero eigenvalues. The state $|\Psi\rangle$ is entangled if and only if either reduced state is mixed, equivalently if and only if the Schmidt rank is strictly greater than 1. Second, every mixed state can be realized as the marginal of a pure state on a larger space. Given a decomposition $\rho_A = \sum_k \lambda_k |k\rangle\langle k|$, the vector

$$|\Phi\rangle_{AR} = \sum_k \sqrt{\lambda_k}\, |k\rangle_A \otimes |k\rangle_R$$

on an auxiliary space $\mathcal{H}_R$ satisfies $\mathrm{tr}_R(|\Phi\rangle\langle\Phi|) = \rho_A$. This construction is called a **purification**.

With the above notations at hand, we can finally give a proof of the Schmidt decomposition. As mentioned above, it is really just a repackaging of the singular value decomposition, but it is instructive to go through the argument in the quantum language above.

Proof of Theorem 32. Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a unit vector. Form the rank-one projector

$$\rho_{AB} := |\Psi\rangle\langle\Psi|$$

and the reduced state on A

$$\rho_A := \mathrm{tr}_B(\rho_{AB}) \in \mathcal{S}(\mathcal{H}_A).$$

Then ρ_A is Hermitian, positive semidefinite, and satisfies $\mathrm{tr}(\rho_A) = 1$. By the spectral theorem there exist an orthonormal set $\{|k\rangle_A\}_{k=1}^r$ and numbers $\lambda_k \geq 0$ with $\sum_{k=1}^r \lambda_k = 1$ such that

$$\rho_A = \sum_{k=1}^r \lambda_k\, |k\rangle_A\langle k|\,,$$

where $r = \mathrm{rank}(\rho_A) \leq N_A$.

For each k with $\lambda_k > 0$ define a vector in $\mathcal{H}_B$ by

$$|\tilde{k}\rangle_B := \frac{1}{\sqrt{\lambda_k}} (\langle k|_A \otimes \mathbb{1}_B)\, |\Psi\rangle\,.$$

We first check orthonormality. For k, ℓ with $\lambda_k, \lambda_\ell > 0$ we compute

$$\begin{aligned}\langle\tilde{k}|\tilde{\ell}\rangle &= \frac{1}{\sqrt{\lambda_k\lambda_\ell}} \langle\Psi|\big(|k\rangle\langle\ell|_A \otimes \mathbb{1}_B\big)|\Psi\rangle \\ &= \frac{1}{\sqrt{\lambda_k\lambda_\ell}}\, \langle k|\, \rho_A\, |\ell\rangle = \frac{1}{\sqrt{\lambda_k\lambda_\ell}}\, \lambda_\ell\, \delta_{k\ell} = \delta_{k\ell}\,,\end{aligned}$$

so $\{|\tilde{k}\rangle_B\}_{k=1}^r$ is an orthonormal set in $\mathcal{H}_B$. Hence $r \leq N_B$ as well.

Next we claim that $|\Psi\rangle = \sum_{k=1}^r \sqrt{\lambda_k}\, |k\rangle_A \otimes |\tilde{k}\rangle_B$. Let us define

$$|\Phi\rangle := \sum_{k=1}^r \sqrt{\lambda_k}\, |k\rangle_A \otimes |\tilde{k}\rangle_B\,,$$

and compare the two vectors by projecting onto A. For any m in an orthonormal basis of $\mathcal{H}_A$ that extends $\{|k\rangle_A\}_{k=1}^r$ we have

$$(\langle m|_A \otimes \mathbb{1}_B)\, |\Psi\rangle = \begin{cases} \sqrt{\lambda_m}\, |\tilde{m}\rangle_B & \text{if } \lambda_m > 0 \\ 0 & \text{if } \lambda_m = 0 \end{cases}$$

by construction. The same identities hold with $|\Psi\rangle$ replaced by $|\Phi\rangle$. Therefore

$$(\langle m|_A \otimes \langle \phi|_B)\left(|\Psi\rangle - |\Phi\rangle\right) = 0$$

for every m and every $|\phi\rangle \in \mathcal{H}_B$. Since such product bras span $(\mathcal{H}_A \otimes \mathcal{H}_B)^*$, it follows that $|\Psi\rangle = |\Phi\rangle$.

Finally observe the reduced state on B,

$$\rho_B := \mathrm{tr}_A(\rho_{AB}) = \sum_{k=1}^{r} \lambda_k \, |\tilde{k}\rangle_B \langle \tilde{k}| \,,$$

so the nonzero spectra of ρ_A and ρ_B agree and equal $\{\lambda_k\}$. The number r is therefore the common rank of ρ_A and ρ_B, which gives $r \leq \min\{N_A, N_B\}$.

We have produced orthonormal sets $\{|k\rangle_A\}$ and $\{|\tilde{k}\rangle_B\}$ and nonnegative numbers $\{\lambda_k\}$ that sum to one such that

$$|\Psi\rangle = \sum_{k=1}^{r} \sqrt{\lambda_k} \, |k\rangle_A \otimes |\tilde{k}\rangle_B \,,$$

which is the desired form. □

Remark 34 (Uniqueness and degeneracies)**.** *The multiset of nonzero coefficients $\{\lambda_k\}$ is uniquely determined by $|\Psi\rangle$ since it is the spectrum of ρ_A and also of ρ_B. The orthonormal families $\{|k\rangle_A\}$ and $\{|\tilde{k}\rangle_B\}$ are unique up to phases when the λ_k are distinct. Within a degenerate eigenspace one may apply a unitary rotation on A and the same conjugate rotation on the corresponding span on B without changing the state $|\Psi\rangle$.*

Now we turn to some examples.

Example 11 (Embedding classical probability into quantum states). Fix the computational basis $\{|i\rangle\}_{i=1}^{N}$ of $\mathbb{C}^N$. A classical distribution $\vec{p} = (p_1, \ldots, p_N) \in \Delta_N$ is encoded as the diagonal density matrix

$$\rho_{\mathrm{cl}}(\vec{p}) = \sum_{i=1}^{N} p_i \, |i\rangle\langle i|.$$

A measurement in this basis with projectors $P_i = |i\rangle\langle i|$ returns outcome i with probability $\mathrm{tr}(P_i \rho_{\mathrm{cl}}) = p_i$, matching the classical rule.

Example 12 (Bell state, reduced states, and entanglement). Consider two qubits with computational basis $|0\rangle, |1\rangle$. We will write $|00\rangle$ as a shorthand for $|0\rangle \otimes |0\rangle$, and similarly for $|11\rangle$. The maximally entangled vector

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle\right) \,, \quad \rho_{AB} = |\Phi^+\rangle\langle\Phi^+| \,,$$

has reduced states

$$\rho_A = \mathrm{tr}_B(\rho_{AB}) = \frac{1}{2}\,\mathbb{1}, \qquad \rho_B = \mathrm{tr}_A(\rho_{AB}) = \frac{1}{2}\,\mathbb{1} \,.$$

Each qubit by itself looks completely random, yet the pair together sits in a definite pure state. Local mixedness together with global purity is a signature of entanglement and has no classical analogue.

In summary, tensor products allow us to assemble composite systems, while density matrices enable us to represent both quantum superposition and classical randomization within a single calculus. The partial trace serves as the quantum marginalization operator, mirroring our earlier $\vec{1}^T$ trick. Together, these tools provide a unified linear-algebraic framework for handling open systems, correlations, and measurements on subsystems.

2.3.2. *POVMs and channels*

We now broaden the two pillars introduced so far, namely unitary time evolution and projective (yes/no) measurements, into the general language of **quantum channels** and **POVMs** (positive operator-valued measures). This framework cleanly captures **open-system** dynamics (interaction with an environment) and the most general measurement statistics allowed by quantum mechanics. The picture to keep in mind is simple: attach an ancilla (the "apparatus" or "environment"), evolve unitarily on the larger space, and then either (i) forget the ancilla (a channel), or (ii) read the ancilla (a measurement). Everything that happens to a system can be modeled this way.

First we consider dynamics in the form of quantum channels. Fix a system Hilbert space $\mathcal{H}_S \simeq \mathbb{C}^d$ (here the subscript 'S' stands for 'system'). In practice a system rarely evolves in isolation; it can interact with an external register $\mathcal{H}_E \simeq \mathbb{C}^{d'}$ prepared in some state ρ_E (here the subscript 'E' stands for 'environment'). If the joint closed dynamics is unitary U_{SE}, then any initial system state ρ_S evolves as

$$\rho_S \longmapsto \mathcal{E}[\rho_S] := \mathrm{tr}_E\big(U_{SE}(\rho_S \otimes \rho_E)U_{SE}^\dagger\big).$$

From the cyclicity of trace and $\mathrm{tr}(\rho_E) = 1$, we immediately get $\mathrm{tr}(\mathcal{E}[\rho_S]) = \mathrm{tr}(\rho_S)$, i.e. *trace preservation*. Moreover, tensoring with an arbitrary ancilla and applying the above form shows *complete positivity*: $(\mathrm{Id}_A \otimes \mathcal{E})[X] \succeq 0$ for every positive X on $\mathcal{H}_A \otimes \mathcal{H}_S$.[9]

Definition 35 (Quantum channel)**.** *A* ***quantum channel*** *(or* ***quantum process****) on $\mathcal{H}_S$ is a linear map $\mathcal{E} : \mathcal{S}(\mathcal{H}_S) \to \mathcal{S}(\mathcal{H}_S)$ that is completely positive and trace-preserving (CPTP).*

The dilation form above is not just an example; it is universal:

Theorem 36 (Stinespring dilation)**.** *Every CPTP map $\mathcal{E}$ on $\mathcal{H}_S \simeq \mathbb{C}^d$ admits a representation of the above form for some environment dimension d', environment state ρ_E, and unitary U_{SE} that are fixed independently of the input ρ_S.*

We defer the proof of this since we need an additional structural result about quantum channels.

A convenient "matrix-element" form drops out when ρ_E is pure, say $\rho_E = |0\rangle\langle 0|$. Expanding U_{SE} in an orthonormal basis $\{|i\rangle_E\}$ and defining the **Kraus operators**

$$K_i := \langle i|U_{SE}|0\rangle \in \mathbb{C}^{d\times d},$$

[9]Positivity alone would require $\mathcal{E}[X] \succeq 0$ whenever $X \succeq 0$ on $\mathcal{H}_S$; *complete* positivity demands the same after adjoining *any* spectator system A. Physically, this guarantees the map never creates negative probabilities even on half of an entangled state.

we obtain the **operator-sum** (Kraus) representation

$$\mathcal{E}[\rho] = \sum_i K_i \, \rho \, K_i^\dagger, \qquad \sum_i K_i^\dagger K_i = \mathbb{1}.$$

Conversely, any family $\{K_i\}$ obeying the completeness relation defines a CPTP map. The Kraus representation is nonunique: $\{K_i\}$ and $\{\sum_j u_{ij} K_j\}$ (with u unitary) describe the same channel. These facts are formalized and proved in the following theorem:

Theorem 37 (Kraus decomposition)**.** *Let $\mathcal{E} : \mathcal{S}(\mathcal{H}_S) \to \mathcal{S}(\mathcal{H}_S)$ be CPTP on a d–dimensional Hilbert space $\mathcal{H}_S \simeq \mathbb{C}^d$. Then there exist operators $K_1, \dots, K_r$ on $\mathcal{H}_S$ with*

$$\mathcal{E}[X] = \sum_{i=1}^{r} K_i \, X \, K_i^\dagger \quad \textit{for all } X, \quad \sum_{i=1}^{r} K_i^\dagger K_i = \mathbb{1},$$

where $r \le d^2$. Conversely, any finite family $\{K_i\}$ obeying $\sum_i K_i^\dagger K_i = \mathbb{1}$ defines a CPTP map by the same formula. The representation is nonunique: if U is any unitary and $K_i' := \sum_j U_{ij} K_j$, then $\{K_i'\}$ yields the same channel.

Proof. To begin, recall the Choi-Jamiołkowski isomorphism. Fix an orthonormal basis $\{|j\rangle\}_{j=1}^d$ of $\mathcal{H}_S$ and define the (unnormalized) maximally entangled vector

$$|\Omega\rangle := \sum_{j=1}^{d} |j\rangle \otimes |j\rangle \;\in\; \mathcal{H}_S \otimes \mathcal{H}_S.$$

The **Choi matrix** of $\mathcal{E}$ is

$$J_{\mathcal{E}} := (\mathrm{Id} \otimes \mathcal{E})(|\Omega\rangle\langle\Omega|) = \sum_{j,k=1}^{d} |j\rangle\langle k| \otimes \mathcal{E}\big(|j\rangle\langle k|\big).$$

By complete positivity we know that $J_{\mathcal{E}} \succeq 0$. Moreover, one can check that for any X on $\mathcal{H}_S$ we have

$$\mathcal{E}[X] = \mathrm{tr}_1\Big[(X^T \otimes \mathbb{1})\, J_{\mathcal{E}}\Big], \tag{14}$$

where tr_1 is the partial trace over the first tensor factor. This "reconstruction identity" follows by expanding X in the basis $\{|j\rangle\langle k|\}$.

Next observe that since $J_{\mathcal{E}} \succeq 0$ it admits a decomposition into rank–one projectors,

$$J_{\mathcal{E}} = \sum_{i=1}^{r} |v_i\rangle\langle v_i|$$

where $|v_i\rangle \in \mathcal{H}_S \otimes \mathcal{H}_S$, and $r = \mathrm{rank}(J_{\mathcal{E}}) \le d^2$. Each vector $|v_i\rangle$ can be viewed as defining an operator $K_i : \mathcal{H}_S \to \mathcal{H}_S$ via the canonical "vectorization" correspondence: if $|v_i\rangle = \sum_{a,b} v_{ab}^{(i)} |a\rangle \otimes |b\rangle$, then

$$K_i = \sum_{a,b} v_{ab}^{(i)} \, |b\rangle\langle a|.$$

One can verify directly that for every X,

$$\mathrm{tr}_1\Big[(X^T \otimes \mathbb{1})\, |v_i\rangle\langle v_i|\Big] = K_i X K_i^\dagger.$$

Combining this with (14), we find

$$\mathcal{E}[X] = \sum_{i=1}^{r} K_i X K_i^\dagger,$$

which is precisely the operator–sum form.

It remains to check the normalization. Since $\mathcal{E}$ is trace–preserving, for all ρ we have

$$\mathrm{tr}(\rho) = \mathrm{tr}(\mathcal{E}[\rho]) = \sum_{i=1}^{r} \mathrm{tr}(K_i \rho K_i^\dagger) = \mathrm{tr}\Big(\rho \sum_{i=1}^{r} K_i^\dagger K_i\Big).$$

Because this holds for all density operators ρ, it follows that $\sum_i K_i^\dagger K_i = \mathbb{1}$.

Conversely, suppose we start with any collection of operators $\{K_i\}$ satisfying $\sum_i K_i^\dagger K_i = \mathbb{1}$. The map

$$\mathcal{E}[X] = \sum_i K_i X K_i^\dagger$$

is clearly linear. Trace preservation follows from the same computation above, and complete positivity is immediate: for any ancilla system A and any positive operator Z on $\mathcal{H}_A \otimes \mathcal{H}_S$, we have

$$(\mathrm{Id}_A \otimes \mathcal{E})[Z] = \sum_i (\mathbb{1}_A \otimes K_i)\, Z\, (\mathbb{1}_A \otimes K_i)^\dagger \;\succeq\; 0.$$

Finally, note that the Kraus representation is not unique. If $u = (u_{ij})$ is any unitary matrix and we define $K_i' = \sum_j u_{ij} K_j$, then

$$\sum_i K_i' X K_i'^\dagger = \sum_j K_j X K_j^\dagger, \qquad \sum_i K_i'^\dagger K_i' = \sum_j K_j^\dagger K_j = \mathbb{1},$$

so $\{K_i\}$ and $\{K_i'\}$ describe the same channel. □

Remark 38 (Minimal Kraus number)**.** *The number r of Kraus operators can always be chosen as $r = \mathrm{rank}(J_\mathcal{E}) \le d^2$. This number is minimal; any other representation can be obtained by enlarging the list with zero operators and applying a unitary rotation among them.*

Having established the Kraus decomposition, we can now establish Stinespring dilation:

Proof of Theorem 36. By the Kraus decomposition, choose operators $K_1, \dots, K_r$ on $\mathcal{H}_S$ with $r \le d^2$ such that

$$\mathcal{E}[\rho] = \sum_{i=1}^{r} K_i\, \rho\, K_i^\dagger \quad \text{and} \quad \sum_{i=1}^{r} K_i^\dagger K_i = \mathbb{1}.$$

Let us introduce an environment Hilbert space $\mathcal{H}_E \simeq \mathbb{C}^r$ with orthonormal basis $\{|i\rangle_E\}_{i=1}^r$ and define an isometry

$$V : \mathcal{H}_S \longrightarrow \mathcal{H}_S \otimes \mathcal{H}_E, \quad V|\psi\rangle \;:=\; \sum_{i=1}^{r} K_i|\psi\rangle \otimes |i\rangle_E.$$

Because $\sum_i K_i^\dagger K_i = \mathbb{1}$, we have $V^\dagger V = \mathbb{1}$; indeed, for all $|\phi\rangle, |\psi\rangle \in \mathcal{H}_S$,

$$\langle\phi|V^\dagger V|\psi\rangle = \sum_{i=1}^{r} \langle\phi|K_i^\dagger K_i|\psi\rangle = \langle\phi|\psi\rangle.$$

Taking the partial trace over E then recovers the channel:

$$\mathrm{tr}_E\big(V\rho V^\dagger\big) = \mathrm{tr}_E\left(\sum_{i,j} K_i \rho K_j^\dagger \otimes |i\rangle\langle j|\right) = \sum_i K_i \rho K_i^\dagger = \mathcal{E}[\rho].$$

To express V using a unitary on system plus environment with a fixed environment state, fix a distinguished vector $|0\rangle_E \in \mathcal{H}_E$ and identify $\mathcal{H}_S$ with the d–dimensional subspace $\mathcal{H}_S \otimes |0\rangle_E \subset \mathcal{H}_S \otimes \mathcal{H}_E$. Define U_{SE} on this subspace by

$$U_{SE}\big(|\psi\rangle \otimes |0\rangle_E\big) := V|\psi\rangle$$

for all $|\psi\rangle$ in $\mathcal{H}_S$. Since V is an isometry, this prescription maps an orthonormal basis of $\mathcal{H}_S \otimes |0\rangle_E$ to an orthonormal set in $\mathcal{H}_S \otimes \mathcal{H}_E$. Extend that partial isometry to a unitary U_{SE} on all of $\mathcal{H}_S \otimes \mathcal{H}_E$ by completing orthonormal bases on the domain and codomain and defining U_{SE} to map one basis to the other. Consequently,

$$\mathcal{E}[\rho] = \mathrm{tr}_E\big(V\rho V^\dagger\big) = \mathrm{tr}_E\Big(U_{SE}\,(\rho \otimes |0\rangle\langle 0|)\,U_{SE}^\dagger\Big),$$

which is precisely the stated dilation with environment state $\rho_E = |0\rangle\langle 0|$ and environment dimension $d' = r$. The unitary U_{SE} and the state ρ_E are determined by the chosen Kraus family for $\mathcal{E}$ and therefore are fixed independently of the input ρ. This completes the proof. □

Remark 39 (Minimal and nonunique dilations)**.** *If the Kraus family is chosen to be minimal (with $r = \mathrm{rank}(J_\mathcal{E})$), then $d' = r$ is the minimal environment dimension. Any two Kraus representations $\{K_i\}$ and $\{K_i'\}$ related by a unitary mixing $K_i' = \sum_j u_{ij} K_j$ yield dilations whose isometries differ by a unitary on the environment: $V' = (\mathbb{1} \otimes u)\,V$. Allowing a mixed ρ_E entails no extra generality, since any mixed state can be purified by enlarging the environment.*

Next we make some additional remarks about quantum channels.

Remark 40 (Composition and randomized control)**.** *Channels are closed under composition and convex combination. If $\mathcal{E}$ and $\mathcal{F}$ are channels, then so is $\mathcal{F} \circ \mathcal{E}$. And if with classical probabilities r_j you apply $\mathcal{E}_j$, the average map $\sum_j r_j\,\mathcal{E}_j$ is again a channel. Thus the set of channels is a convex monoid under composition.*

Remark 41 (Heisenberg picture)**.** *The adjoint map $\mathcal{E}^*$ acts on observables and satisfies*

$$\mathrm{tr}\big(\mathcal{E}[\rho]\,A\big) = \mathrm{tr}\big(\rho\,\mathcal{E}^*[A]\big), \quad \mathcal{E}^*[\mathbb{1}] = \mathbb{1}\,.$$

In Kraus form, $\mathcal{E}^*[A] = \sum_i K_i^\dagger A K_i$. We will use this duality to shuttle between "state evolution" and "observable evolution."

Having discussed general dynamics, we now turn out attension to general measurements. Projective measurements are special cases of more general procedures obtained by attaching an apparatus, evolving unitarily, and reading an outcome on the apparatus. Let $\{|i\rangle_A\}_{i=1}^N$ be an orthonormal basis for the apparatus and let U

act on system+apparatus. If the apparatus is initialized in $|0\rangle_A$ and we measure it in the $\{|i\rangle_A\}$ basis, the probability of outcome i on input ρ is

$$p(i) = \mathrm{tr}(F_i\,\rho)\,, \qquad F_i := M_i^\dagger M_i, \qquad M_i := \langle i|U|0\rangle,$$

with $\sum_i F_i = \mathbb{1}$ by unitarity.

Definition 42 (POVM)**.** *A **positive operator-valued measure** (POVM) on $\mathcal{H}_S$ is a finite collection of positive semidefinite operators $\{F_i\}_{i=1}^N$ obeying $\sum_i F_i = \mathbb{1}$. Given a state ρ, the Born rule assigns outcome probabilities $p(i) = tr(F_i\rho)$.*

The operators F_i are sometimes called **effects**. When $F_i = P_i$ are orthogonal projectors that sum to $\mathbb{1}$ we recover the projective measurements from the axioms. In general, many distinct physical procedures can realize the same POVM statistics. One convenient realization chooses **measurement operators** (one set among many)

$$M_i \quad \text{with} \quad M_i^\dagger M_i = F_i,$$

and then the post-measurement state conditioned on outcome i is

$$\rho \;\longmapsto\; \frac{M_i\,\rho\,M_i^\dagger}{\mathrm{tr}(F_i\rho)}.$$

The family $\{\mathcal{I}_i\}_i$ with $\mathcal{I}_i[\rho] := M_i\rho M_i^\dagger$ is called a **quantum instrument**; it records both the probabilities and the (normalized) output states. Forgetting the outcome yields the average channel $\sum_i \mathcal{I}_i$.

As with channels, there is a universal dilation theorem for POVMs:

Theorem 43 (Naimark dilation)**.** *Every POVM $\{F_i\}$ on $\mathcal{H}_S$ can be realized as a projective measurement on a larger space: there exist an auxiliary Hilbert space $\mathcal{H}_A$, an isometry $V : \mathcal{H}_S \to \mathcal{H}_S \otimes \mathcal{H}_A$, and orthogonal projections $\{\Pi_i\}$ on $\mathcal{H}_A$ such that*

$$F_i = V^\dagger(\mathbb{1}\otimes\Pi_i)V \quad \textit{and} \quad p(i) = \mathrm{tr}\big(F_i\rho\big) = \mathrm{tr}\big[(\mathbb{1}\otimes\Pi_i)\,V\rho V^\dagger\big].$$

Remark 44 (Rank-one refinement)**.** *Every POVM admits a refinement to rank-one effects. Diagonalize each $F_i = \sum_j \lambda_{ij}\,|v_{ij}\rangle\langle v_{ij}|$ and regard the collection $\{F_{i,j} := \lambda_{ij}|v_{ij}\rangle\langle v_{ij}|\}_{i,j}$ as a new POVM. Coarse-graining its outcomes by summing over j reproduces the original statistics:*

$$\sum_j \mathrm{tr}(F_{i,j}\rho) = \mathrm{tr}(F_i\rho).$$

Thus, without loss of generality, one may work with rank-one POVMs when convenient.

To concretize the formalism, we record two examples.

Example 13 (Unsharp qubit measurement). For a qubit with Pauli vector $\vec{\sigma} = (X, Y, Z)$ and a unit vector $\hat{n} \in \mathbb{R}^3$, the two-outcome effects

$$F_\pm^{(\eta,\hat{n})} = \frac{\mathbb{1} \pm \eta\,\hat{n}\cdot\vec{\sigma}}{2}, \qquad 0 \le \eta \le 1,$$

form a POVM. The parameter η is a *sharpness*: $\eta = 1$ gives the projective measurement along $\hat{n}$, while smaller η yields noisy readout with probabilities

$$p(\pm) = \mathrm{tr}\big(F_\pm^{(\eta,\hat{n})}\rho\big) = \tfrac{1}{2}\big(1 \pm \eta\,\hat{n}\cdot\vec{r}\big),$$

where $\vec{r} = (\langle X\rangle, \langle Y\rangle, \langle Z\rangle)$ is the Bloch vector of ρ.

Example 14 (Embedding classical dynamics into a channel). Classical column-stochastic matrices are naturally realized as quantum channels that act classically on the computational basis and erase coherence. Fix an orthonormal basis $\{|i\rangle\}_{i=1}^{N}$ and let $M = (M_{ij})$ be column-stochastic ($M_{ij} \geq 0$ and $\sum_i M_{ij} = 1$ for each j). Define Kraus operators

$$K_{i|j} = \sqrt{M_{ij}}\, |i\rangle\langle j|.$$

Then

$$\mathcal{E}_M[\rho] = \sum_{i,j} K_{i|j}\, \rho\, K_{i|j}^\dagger, \qquad \sum_{i,j} K_{i|j}^\dagger K_{i|j} = \sum_j \Big(\sum_i M_{ij}\Big)|j\rangle\langle j| = \mathbb{1},$$

so $\mathcal{E}_M$ is CPTP. On diagonal inputs $\rho_{\mathrm{cl}}(\vec{p}) = \sum_j p_j |j\rangle\langle j|$ we recover the classical update

$$\mathcal{E}_M\big[\rho_{\mathrm{cl}}(\vec{p})\big] = \sum_{i,j} M_{ij} p_j\, |i\rangle\langle i| = \rho_{\mathrm{cl}}(M \cdot \vec{p}),$$

while for $j \neq k$ the coherence $|j\rangle\langle k|$ is sent to 0 because each Kraus term carries the same input label on both sides. Thus $\mathcal{E}_M$ is a "classicalizing" channel: it dephases in the computational basis and then applies the Markov update to the resulting distribution.

We have seen that the familiar tools of unitary evolution and projective measurements represent only the simplest quantum operations. Real quantum systems demand a richer framework: we enlarge the Hilbert space with ancillary systems, apply unitary evolution to the combined system, then either trace out the ancilla (yielding quantum channels) or measure it (yielding POVMs). This procedure generates the most general dynamics and measurement statistics that quantum mechanics allows. We have explained that quantum channels are completely positive trace-preserving (CPTP) linear maps, characterized by the Kraus representation or Stinespring dilation. POVMs are sets of positive operators that sum to the identity, understood through Naimark's theorem. But the conceptual heart is simple: we compose systems, evolve them unitarily, and then selectively forget or record information.

This unified framework will prove essential for understanding real quantum devices; indeed, in the real world, noise is inevitable, information is incomplete, and systems interact with environments beyond our control. Rather than limitations to work around, these general operations become the natural language for describing quantum processes in practice.

3. A taste of quantum many-body physics

We now turn to many–body systems built from n qubits. The ambient Hilbert space is the n–fold tensor product

$$\mathcal{H} := (\mathbb{C}^2)^{\otimes n} \simeq \mathbb{C}^{2^n}.$$

It is convenient to fix the computational basis $\{|0\rangle, |1\rangle\}$ on each site and to use the Pauli operators X, Y, Z discussed above. To streamline notation, we introduce a

shorthand: for $1 \le i \le n$ we write

$$X_i := \mathbb{1}^{\otimes(i-1)} \otimes X \otimes \mathbb{1}^{\otimes(n-i)},$$

and similarly for Y_i and Z_i. Products such as Z_iZ_j are understood to mean $Z_i \otimes Z_j$ with identities on all other sites, which we will not display explicitly.

More generally, a **Pauli string** on n qubits is a tensor product

$$P = \sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n}, \qquad \sigma_{a_k} \in \{\, \mathbb{1}, X, Y, Z \,\},$$

and its **weight** is the number of non-identity factors,

$$w(P) := |\{\, k : \sigma_{a_k} \neq \mathbb{1} \,\}|,$$

while its **support** is the set $\mathrm{supp}(P)$ of sites where $\sigma_{a_k} \neq \mathbb{1}$. Two elementary commutation facts will be used repeatedly: Pauli matrices on different sites commute, while distinct Pauli matrices on the same site anticommute. Equivalently, Pauli strings P and Q either commute or anticommute, with

$$PQ = (-1)^{N_{\mathrm{anti}}(P,Q)}\, QP,$$

where $N_{\mathrm{anti}}(P, Q)$ counts the number of sites where both act nontrivially with different Pauli matrices.

With this notation in hand, we can define Hamiltonians. A **Hamiltonian** on $\mathcal{H}$ is a Hermitian operator $H = H^\dagger$. In units $\hbar \equiv 1$, the closed–system time evolution is

$$U(t) = e^{-\,iHt}, \qquad |\Psi(t)\rangle = U(t)\,|\Psi(0)\rangle.$$

Since H is Hermitian, its spectrum is real. We denote its smallest eigenvalue by E_0 (the **ground energy**) and the corresponding eigenspace by the **ground space**. A Hamiltonian is called k–**local** if it decomposes as

$$H = \sum_a H_a\,, \quad w(H_a) \le k \;\text{ for every term } H_a,$$

i.e. each interaction acts nontrivially on at most k sites. In the qubit setting one often expands H in the Pauli–string basis,

$$H = \sum_\alpha h_\alpha\, P_\alpha\,, \quad w(P_\alpha) \le k,$$

with real coefficients h_α. To express **geometric locality**, we can place the n qubits on the vertices V of a graph $G = (V, E)$. A geometrically k–local Hamiltonian has each H_a supported on a connected region of at most k vertices (for $k = 2$, typically on edges $(i, j) \in E$). For example, on a line $G = \{1, \ldots, n\}$ with edges $(i, i+1)$, a nearest–neighbor two–local Hamiltonian has the form

$$H = \sum_{i=1}^{n-1} H_{i,i+1} + \sum_{i=1}^{n} H_i,$$

with $H_{i,i+1}$ acting only on sites $i, i+1$ and H_i acting on site i.

The canonical playground for these ideas is the (ferromagnetic) **transverse–field Ising model** (TFIM) on a graph $G = (V, E)$:

$$H_{\mathrm{TFIM}}(J, h) = -\,J \sum_{(i,j)\in E} Z_iZ_j \;-\; h \sum_{i\in V} X_i, \quad J \ge 0,\, h \ge 0.$$

The first term lowers the energy when neighboring Z–spins align, while the second term lowers the energy for qubits pointing in the x–direction (the $|+\rangle$ eigenstate of

X). Thus the two terms compete, and since Z and X do not commute, the model is genuinely quantum. The model is 2–local and geometrically local on G.

Two limiting regimes are exactly solvable and already illustrative. In the classical limit $h = 0$, all terms commute. Ground states maximize each Z_iZ_j, so for $J > 0$ they are the two fully aligned product states $|0\cdots0\rangle$ and $|1\cdots1\rangle$, with two–fold degeneracy. Excitations are domain walls: a bond with anti–aligned neighbors costs energy $2J$ (on an open chain; with periodic boundary conditions, domain walls come in pairs costing $4J$ total). In the opposite paramagnetic limit $J = 0$, each site independently minimizes $-hX_i$, with a unique ground state $|+\rangle^{\otimes n}$ where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. A single spin flip to $|-\rangle$ costs energy $2h$.

Between these limits, the terms fail to commute, which is the source of genuinely quantum behavior. The model enjoys a $\mathbb{Z}_2$ symmetry generated by the global "spin–flip" operator

$$\mathcal{P} := \prod_{i\in V} X_i,$$

under which $Z_i \mapsto -Z_i$ while $X_i \mapsto X_i$. Since $[\mathcal{P}, H_{\mathrm{TFIM}}] = 0$, the Hamiltonian preserves this symmetry. For small h/J the ground space on large graphs approximately breaks the symmetry, exhibiting long–range Z–order. For large h/J the unique ground state is the symmetric paramagnet. On a one–dimensional chain the model is exactly solvable (via Jordan–Wigner fermionization), and at zero temperature there is a quantum phase transition in the thermodynamic limit ($n \to \infty$) at $h = J$ where the energy gap between the lowest and second lowest eigenvalues of H go to zero. While we will not derive this here, a two–site analysis already captures the competition of the two terms.

Example 15 (Two–site TFIM). On two qubits,

$$H_2(J, h) = -J\, Z_1Z_2 - h\,(X_1 + X_2).$$

Diagonalizing (for instance in the joint eigenbasis of the parity X_1X_2) yields four eigenvalues

$$E \in \left\{-\sqrt{J^2 + 4h^2},\ -J,\ +J,\ +\sqrt{J^2 + 4h^2}\right\}.$$

For $J \geq 0$ the ground energy is $E_0 = -\sqrt{J^2 + 4h^2}$, and the gap to the first excited level is

$$\Delta(J, h) = \sqrt{J^2 + 4h^2} - J.$$

We recover the limits discussed above: $\Delta(0, h) = 2h$ and $\Delta(J, 0) = 0$ (reflecting the two–fold degeneracy at $h = 0$). Already at two sites we see how the transverse field h lifts the classical degeneracy and stabilizes a unique paramagnet, while the interaction J favors ferromagnetic order.

On longer chains, the low–energy excitations can be understood in terms of order and disorder. In the $h = 0$ limit, excitations are domain walls that can move freely; turning on a small h allows them to hop and to be created or annihilated in pairs. In the opposite $J = 0$ limit, the excitations are independent spin flips. The $\mathbb{Z}_2$ symmetry generated by $\mathcal{P}$ forbids a nonzero $\langle Z_i\rangle$ expectation value in any exact eigenstate on a finite chain; nevertheless, in the ferromagnetic phase ($h/J \ll 1$) the ground space is nearly two–fold degenerate and exhibits robust long–range correlations $\langle Z_iZ_j\rangle \approx 1$ for distant i, j.

CHAPTER 3

Tensor Networks

Quantum learning involves systems with many degrees of freedom (e.g. many qubits), which are described by tensor products of Hilbert spaces. In some circumstances, the standard notation describing operators and states in tensor product Hilbert spaces can be unwieldy, and obscures certain structural intuitions. Here we develop a standardized *diagrammatic* notation for manipulating tensors on tensor product Hilbert spaces which illuminates various kinds of proofs. We will use this notation on occasion in this book.

Before diving into our review, we begin with an anecdote. One of the earlier usages of tensor diagrams is by Roger Penrose, which is why in some communities such diagrams are called 'Penrose graphical notation'. Penrose relayed to one of the authors the following story. When Penrose was a PhD student at Cambridge under the direction of Hodge, he developed his graphical notation to help him better visualize certain proofs in algebraic geometry. One day when he met with Hodge to report his progress, Penrose used these diagrams on Hodge's blackboard; Hodge was puzzled since he had never seen such diagrams before. Penrose said that he would go write up a note explaining the notation to Hodge, and did so in the ensuing week. He gave Hodge a 50 page manuscript with many diagrams, and by Penrose's account, Hodge thought that Penrose must have lost his mind, given that he was claiming tensor algebra could be performed by manipulating a bunch of squiggles. Penrose was of course correct, and so we commence with the squiggles.

1. Review of tensor network diagrams

As promised, the so-called 'tensor network' diagrams will render the index contraction of higher-rank tensors more transparent than standard notations. Our discussion here is based off of [**CCHL22**], and we also refer the interested reader to [**Lan11**, **BC17**] for a more comprehensive overview of tensor networks.

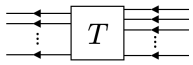
Diagrams for individual tensors

Throughout, a rank (m, n) tensor will mean a multilinear map $T : \mathcal{H}^{*\otimes m} \otimes \mathcal{H}^{\otimes n} \to \mathbb{C}$. If $\{|i\rangle\}$ is an orthonormal basis of $\mathcal{H}$, then in bra–ket notation T admits the expansion

$$T = \sum_{\substack{i_1,\dots,i_m \\ j_1,\dots,j_n}} T^{i_1\cdots i_m}_{j_1\cdots j_n} \left(|i_1\rangle \otimes \cdots \otimes |i_m\rangle\right)\left(\langle j_1| \otimes \cdots \otimes \langle j_n|\right).$$

for some $T^{i_1\cdots i_m}_{j_1\cdots j_n} \in \mathbb{C}$. A quantum state $|\Psi\rangle$ on $\mathcal{H}$ is thus a rank $(1, 0)$ tensor (a map $\mathcal{H}^* \to \mathbb{C}$), and its dual $\langle\Psi|$ is rank $(0, 1)$. Moreover, a matrix $M = \sum_{ij} M^i_j |i\rangle\langle j|$ is

a rank $(1,1)$ tensor. We will depict T diagrammatically as



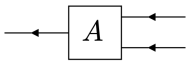
(15)

which carries m outgoing legs on the left and n incoming legs on the right. Each leg corresponds to one index of $T^{i_1\cdots i_m}_{j_1\cdots j_n}$. Our convention is that outgoing legs are ordered counter-clockwise, while incoming legs are ordered clockwise. Concretely, in (15) the upper left outgoing leg is i_1, the one below is i_2, etc.; symmetrically on the right, the top incoming leg is j_1, the next is j_2, and so forth.

Tensor contraction

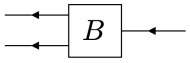
We now describe how to indicate tensor-network contractions. For illustration, consider a rank $(2,1)$ tensor

$$A = \sum_{ijk} A^i_{jk} \, |i\rangle \left(\langle j| \otimes \langle k| \right) \; =$$



and a rank $(1,2)$ tensor

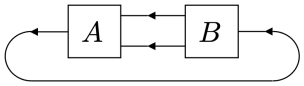
$$B = \sum_{\ell mn} B^{mn}_{\ell} \left(|m\rangle \otimes |n\rangle \right) \langle \ell| \; =$$



Suppose we wish to evaluate

$$\sum_{ijk} A^i_{jk} B^{jk}_i \, . \tag{16}$$

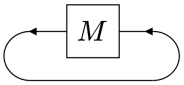
Here lower indices pair with upper indices, reflecting vector–covector contraction. The corresponding diagram is



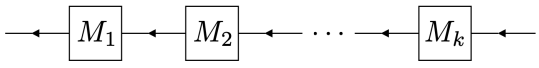
Reading this against (16), the contracted indices are precisely those whose incoming and outgoing legs are glued. Only legs with compatible orientations may be joined, encoding the rule that vectors contract with covectors.

As another instance, for a matrix $M = \sum_{ij} M^i_j \, |i\rangle\langle j|$, the trace is drawn as

$$\mathrm{tr}(M) = \sum_i M^i_i =$$



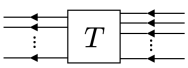
If $M_1, M_2, ..., M_k$ are matrices, then their product $M_1 M_2 \cdots M_k$ appears as



Multiplication by a scalar

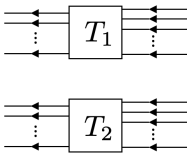
For a tensor T and scalar α, we notate $\alpha\, T$. In diagrams we simply write

α

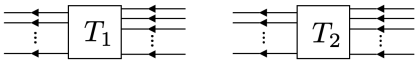


Tensor products

Given tensors T_1 and T_2, their tensor product $T_1 \otimes T_2$ is represented by



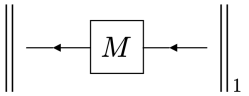
or equivalently by



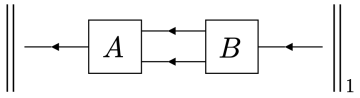
The ordering (e.g. $T_1 \otimes T_2$ versus $T_2 \otimes T_1$) will be evident from context.

Taking norms

Matrix norms are often conveniently expressed in this notation. If M is a matrix, its 1-norm $\|M\|_1$ is indicated by



Here the diagram for M acts as a placeholder inside $\|M\|_1$. This is especially useful when M itself arises from a contraction whose structure we wish to emphasize; for example, if $M = \sum_{ijk\ell} A^i_{k\ell} B^{k\ell}_j \, |i\rangle\langle j|$, then



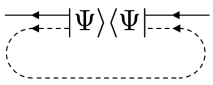
Tensors with legs of different dimensions

Thus far we have treated rank (m, n) tensors as maps $T : \mathcal{H}^{*\otimes m} \otimes \mathcal{H}^{\otimes n} \to \mathbb{C}$. More generally, consider

$$T : (\mathcal{H}_1^* \otimes \cdots \otimes \mathcal{H}_m^*) \otimes (\mathcal{H}_{m+1} \otimes \cdots \otimes \mathcal{H}_{m+n}) \to \mathbb{C},$$

where the Hilbert spaces need not be isomorphic. The same diagrammatic rules apply, with the additional restriction that two legs may be contracted only if they correspond to a Hilbert space and its dual of the same dimension.

As an example, take a state $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^3$ and its density operator $|\Psi\rangle\langle\Psi|$. We will draw the $\mathbb{C}^2$ (qubit) legs as solid and the $\mathbb{C}^3$ (qutrit) legs as dotted. A partial trace over the qutrit subsystem reads



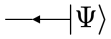
(17)

We return to partial traces in greater detail below.

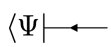
Identity operator

The identity on $\mathcal{H}$ is represented by a single oriented line:

Thus for a state in $\mathcal{H}$,



left-multiplication by the identity leaves the diagram (and therefore the state) unchanged. Likewise, for the dual state



right-multiplying by the identity returns the same diagram.

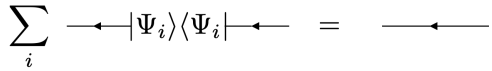
For k copies, the identity on $\mathcal{H}^{\otimes k}$ is

If instead the overall Hilbert space is $\mathcal{H} \otimes \mathcal{H}'$ with different factor dimensions, we take $\mathcal{H}$-legs to be solid and $\mathcal{H}'$-legs dotted; then

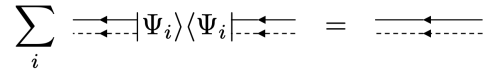
and the evident generalization covers more than two distinct factors.

Resolutions of the identity

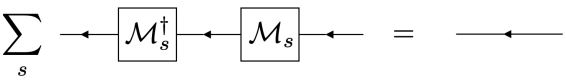
If $\{|\Psi_i\rangle\}_i$ is an orthonormal basis of $\mathcal{H}$, then $\sum_i |\Psi_i\rangle\langle\Psi_i| = \mathbb{1}$ is depicted by



If instead $\{|\Psi_i\rangle\}_i$ resolves the identity on $\mathcal{H} \otimes \mathcal{H}'$ with non-identical factor dimensions, we analogously draw



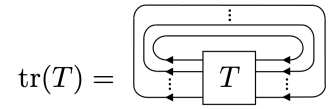
Similarly, if $\{M_s^\dagger M_s\}_s$ is a POVM on $\mathcal{H}$ with $\sum_s M_s^\dagger M_s = \mathbb{1}$, we write



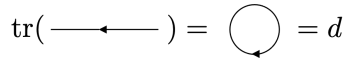
and the same idea extends to $\mathcal{H} \otimes \mathcal{H}'$ and larger tensor products.

Taking traces and partial traces

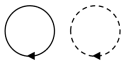
For a rank (n, n) tensor $T : \mathcal{H}^{*\otimes n} \otimes \mathcal{H}^{\otimes n} \to \mathbb{C}$, the trace is $\mathrm{tr}(T) = \sum_{i_1,\ldots,i_n} T^{i_1 \cdots i_n}_{i_1 \cdots i_n}$, drawn as



A particularly useful identity is the trace of $\mathbb{1} = \sum_i |i\rangle\langle i|$, viewed as a rank $(1, 1)$ tensor:



Thus a closed loop equals the dimension of the Hilbert space associated to that curve. For $\mathbb{1}_{d\times d}\otimes\mathbb{1}_{d'\times d'}$ on $\mathcal{H}\otimes\mathcal{H}'$, where $\dim(\mathcal{H})=d$ and $\dim(\mathcal{H}')=d'$, we have

$\mathrm{tr}($  $) =$  $= d\,d'$

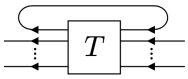
with solid denoting $\mathcal{H}$ and dotted denoting $\mathcal{H}'$.

Partial traces are handled analogously. Define the partial trace over the 'kth subsystem' by

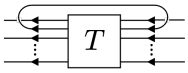
$$\mathrm{tr}_k(T)=\sum_{\substack{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n\\ j_1,\ldots,j_{k-1},j_{k+1},\ldots,j_n}}\left(\sum_{i_k}T^{i_1\cdots i_n}_{j_1\cdots j_n}\right)|i_1\rangle\langle j_1|\otimes\cdots\otimes|i_{k-1}\rangle\langle j_{k-1}|\otimes|i_{k+1}\rangle\langle j_{k+1}|\otimes\cdots\otimes|i_n\rangle\langle j_n|\,.$$

Note that $\mathrm{tr}_\ell(\mathrm{tr}_k(T))=\mathrm{tr}_k(\mathrm{tr}_\ell(T))$, so we may write $\mathrm{tr}_{k,\ell}(T)$, and $\mathrm{tr}_{1,\ldots,n}(T)=\mathrm{tr}(T)$.

Diagrammatically, the partial trace over the first subsystem is

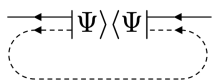
$\mathrm{tr}_1(T) =$ 

Over the second subsystem:

$\mathrm{tr}_2(T) =$ 

and so on.

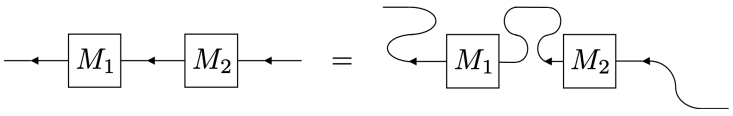
If the legs of a tensor correspond to Hilbert spaces of differing dimensions, traces and partial traces are still available whenever the paired spaces match. For example, if $T:(\mathcal{H}_1^*\otimes\cdots\otimes\mathcal{H}_n^*)\otimes(\mathcal{H}'_1\otimes\cdots\otimes\mathcal{H}'_m)\to\mathbb{C}$ and $\mathcal{H}_k=\mathcal{H}'_k$, we may compute $\mathrm{tr}_k(T)$. As a simple instance, for $|\Psi\rangle\in\mathcal{H}\otimes\mathcal{H}'$, the density operator $|\Psi\rangle\langle\Psi|$ is a $(2,2)$ tensor $(\mathcal{H}^*\otimes\mathcal{H}'^*)\otimes(\mathcal{H}\otimes\mathcal{H}')\to\mathbb{C}$, and

$\mathrm{tr}_2(|\Psi\rangle\langle\Psi|) =$ 

which matches the example in (17); a similar figure represents $\mathrm{tr}_1(|\Psi\rangle\langle\Psi|)$.

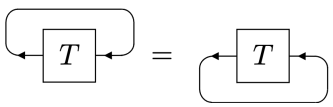
Isotopies

Tensor-network diagrams are interpreted up to isotopy of the legs: bending or smoothly deforming them does not change the meaning. For instance, for a product M_1M_2 we may equally draw

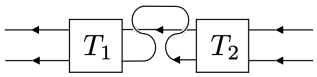


and the same holds in other cases.

Isotopies need not be planar; e.g.

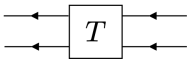


Leg crossings are also allowed:

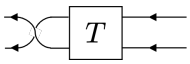


and we will not distinguish over- from under-crossings.

However, we keep fixed the relative ordering of the incoming and outgoing endpoints. Reordering them would permute the tensor factors on which the tensor acts. For example, let $T : (\mathcal{H}_1^* \otimes \mathcal{H}_2^*) \otimes (\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathbb{C}$ be drawn as



Then



corresponds to a tensor on $(\mathcal{H}_2^* \otimes \mathcal{H}_1^*) \otimes (\mathcal{H}_1 \otimes \mathcal{H}_2)$, where the dual factors have been swapped. See also the discussion of permutation operators below.

Permutation operators

Let S_k denote the permutation group on k elements, and let $\tau \in S_k$. Define $\mathrm{Perm}(\tau)$ acting on $\mathcal{H}^{\otimes k}$ by

$$\mathrm{Perm}(\tau)|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle = |\psi_{\tau^{-1}(1)}\rangle \otimes |\psi_{\tau^{-1}(2)}\rangle \otimes \cdots \otimes |\psi_{\tau^{-1}(n)}\rangle$$

and extend linearly. With this convention we have

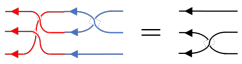
$$\mathrm{Perm}(\tau) \cdot \mathrm{Perm}(\sigma) = \mathrm{Perm}(\tau\sigma)$$

where $\tau\sigma$ denotes the group product (composition $\tau \circ \sigma$).

These representations admit an especially transparent diagrammatics. For S_3 and $\tau = (123)$, we draw

which becomes clear upon labeling the endpoints:

The group product is just as visible; e.g., $\mathrm{Perm}((123)) \cdot \mathrm{Perm}((12))$ is



with $\mathrm{Perm}((123))$ drawn in red and $\mathrm{Perm}((12))$ in blue for emphasis; allowable isotopies (without reordering endpoints) show the result is $\mathrm{Perm}((23))$. Note also that horizontally flipping the diagram for $\mathrm{Perm}(\tau)$ yields that for $\mathrm{Perm}(\tau^{-1})$.

As another example, acting with $\mathrm{Perm}((123))$ on a state $|\Psi\rangle \in \mathcal{H}^{\otimes 3}$ gives

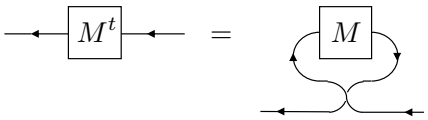
making it evident that the tensor factors are permuted according to $(123)^{-1} = (132)$.

In later arguments, when no confusion can arise, we will abbreviate $\mathrm{Perm}(\tau)$ simply as τ.

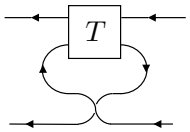
Transposes and partial transposes

Let $M = \sum_{i,j} M^i_j \, |i\rangle\langle j|$ be viewed as a rank $(1,1)$ tensor. Its transpose $M^t = \sum_{i,j} M^i_j \, |j\rangle\langle i|$ can be indicated diagrammatically as follows.

Here we dualize each leg by reversing its arrow, then use isotopy to reorient so the in-arrow enters from the right and the out-arrow exits to the left; this is done to match the orientation of the diagram on the left.



For a higher-rank tensor, e.g. a rank $(2,2)$ tensor $T = \sum_{ijk\ell} T^{ij}_{k\ell} \, |i\rangle\langle k| \otimes |j\rangle\langle \ell|$, we may transpose only one subsystem; the partial transpose on the second subsystem, $\sum_{ijk\ell} T^{ij}_{k\ell} \, |i\rangle\langle k| \otimes |\ell\rangle\langle j|$, is shown as



and the same notation extends in the obvious way to higher rank.

Maximally entangled state

The maximally entangled state is $|\Omega\rangle = \sum_i |i\rangle|i\rangle$ in the computational basis, taken unnormalized. We depict $|\Omega\rangle$ and its Hermitian conjugate by

$|\Omega\rangle =$ 

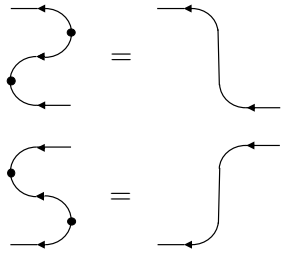
$\langle\Omega| =$ 

Let $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathcal{H}_C$. Then

$$(\mathbb{1}_A \otimes \langle\Omega|_{BC}) \, (|\Omega\rangle_{AB} \otimes \mathbb{1}_C) = \sum_i |i\rangle_A \langle i|_C$$

$$(\langle\Omega|_{AB} \otimes \mathbb{1}_C) \, (\mathbb{1}_A \otimes |\Omega\rangle_{BC}) = \sum_i |i\rangle_C \langle i|_A$$

which we draw as



One can view the black dot as implementing a transpose, since it flips the leg's orientation; two such dots cancel, reflecting that a double transpose is the identity.

2. Some applications

We give three initial applications of tensor network diagrams to illustrate how they illuminate certain kinds of mathematical relationships and proofs in multilinear algebra.

Example 1: A property of the maximally entangled state. First, consider a Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$ with an orthonormal basis $\{|i\rangle\}_{i=1}^d$. As discussed before, the identity matrix can be written as $\mathbb{1} = \sum_{i=1}^d |i\rangle\langle i|$. Then we have the following definition:

Definition 45 (Maximally entangled state)**.** *The (normalized)* ***maximally entangled state*** *on $\mathbb{C}^d \otimes \mathbb{C}^d$ is given by*

$$|\Phi^+\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|i\rangle \,.$$

This is related to our previous notation above by $|\Phi^+\rangle = \frac{1}{\sqrt{d}} |\Omega\rangle$.

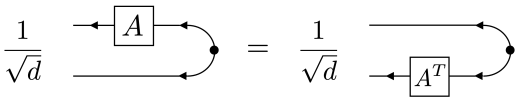
We observe that the maximally entangled state is proportional to the identity matrix if we take the 'transpose' of the subsystems, namely $|\Phi^+\rangle \propto \sum_{i=1}^d |i\rangle(\langle i|)^T = \sum_{i=1}^d |i\rangle|i\rangle$.

As an aside, the reason we use transpose here, and not Hermitian transposition, is as follows. Consider a more general state $\sum_{i,j=1}^d c_{ij}|i\rangle|j\rangle$ where the c_{ij} are complex. Upon transposing the second subsystem, we find the linear operator $\sum_{i,j=1}^d c_{ij}|i\rangle(|j\rangle)^T = \sum_{i,j=1}^d c_{ij}|i\rangle\langle j|$, and conversely we can go from an operator back to a state via a transposition. Note that the transpose is inert if we group the c_{ij}'s with the $|j\rangle$'s, namely $\sum_{i,j=1}^d |i\rangle(c_{ij}\,|j\rangle)^T = \sum_{i,j=1}^d c_{ij}|i\rangle\langle j|$. But if instead we considered Hermitian conjugation then, we would have $\sum_{i,j=1}^d c_{ij}|i\rangle(|j\rangle)^\dagger = \sum_{i,j=1}^d c_{ij}|i\rangle\langle j|$ and $\sum_{i,j=1}^d |i\rangle(c_{ij}\,|j\rangle)^\dagger = \sum_{i,j=1}^d c_{ij}^*|i\rangle\langle j|$, which are not equal. That is, it would matter if we 'grouped' the c_{ij}'s with the $|j\rangle$'s or not. Said a different way, taking the 'partial Hermitian conjugation' of a state, operator, or tensor *violates* multilinearity, whereas taking a 'partial transpose' maintains multilinearity. This is why partial transposition is a valid operation to do.

With the above considerations in mind, we can represent the maximally entangled state by the tensor

$|\Phi^+\rangle = \frac{1}{\sqrt{d}}$ 

which is proportional to the identity tensor with a transpose inserted in. Now consider an operator $A \otimes \mathbb{1}$ on $\mathbb{C}^d \otimes \mathbb{C}^d$. Then applying this operator to $|\Phi^+\rangle$ and applying elementary tensor network manipulations, we find



Thus we see that

$$(A \otimes \mathbb{1})|\Phi^+\rangle = (\mathbb{1} \otimes A^T)|\Phi^+\rangle \,,$$

which is a useful property of the maximally entangled state.

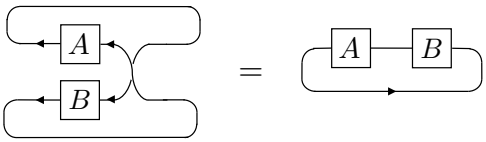
Example 2: SWAP trick and SWAP test. Let us define that **swap operator** $\mathsf{SWAP} : \mathbb{C}^d \otimes \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^d$ by its action on basis states as

$$\mathsf{SWAP}|i\rangle|j\rangle = |j\rangle|i\rangle$$

for all i, j. The action of SWAP extends to other states by multilinearity, and is a permutation operator on two tensor factors. We observe that $\mathsf{SWAP}^\dagger = \mathsf{SWAP}$, and $\mathsf{SWAP}^2 = \mathbb{1}$, so it is both Hermitian and unitary. In line with our notation above, SWAP is expressed diagrammatically as

$\mathsf{SWAP} =$ 

Now let A and B be linear operators acting on $\mathbb{C}^d$. We have



and so we have shown that $\mathrm{tr}(\mathsf{SWAP} \cdot A \otimes B) = \mathrm{tr}(AB)$.

The above identity is the foundation for the so-called **swap test**. Suppose we are given two states $|\Psi\rangle$ and $|\Phi\rangle$ and want to test if they are the same or not. Since the states in question are pure, their corresponding density matrices are $|\Psi\rangle\langle\Psi|$ and $|\Phi\rangle\langle\Phi|$. Since SWAP is a Hermitian operator, it is an observable, and so we are welcome to compute the 'observable' expectation value

$$\mathrm{tr}(\mathsf{SWAP}\,|\Psi\rangle\langle\Psi| \otimes |\Phi\rangle\langle\Phi|) = |\langle\Psi|\Phi\rangle|^2\,,$$

which gives the overlap of the two states. Thus, the overlap between two given states is observable; if it is close to one then states are close to being parallel; if it is close to being zero then the states are close to being orthogonal.

The swap test can also be used for the task of **purity testing**. We will return to this in more detail later, but informally, if we are given copies of a density matrix ρ, we would like to ascertain how close it is to being 'rank one' or 'pure'. Diagonalizing ρ as $\rho = \sum_{i=1}^d p_i |v_i\rangle\langle v_i|$ where $p_i \geq 0$ and $\sum_{i=1}^d p_i = 1$, we see that performing the swap test on two copies of ρ we obtain

$$\mathrm{tr}(\mathsf{SWAP}\,\rho \otimes \rho) = \mathrm{tr}(\rho^2) = \sum_{i=1}^d p_i^2\,.$$

If ρ is pure (so that one $p_i = 1$ and all the rest are zero), then the right-hand side of the above is one; if ρ is impure, then the right-hand side is less than one. Indeed, the smallness of $\sum_{i=1}^d p_i^2$ is a measure of the *impurity* of ρ.

As a special case of purity testing, consider a pure state density matrix $|\Psi\rangle\langle\Psi|$ and a maximally mixed density matrix $\frac{1}{d}\,\mathbb{1}$. Then if we perform the swap test on two copies of $|\Psi\rangle\langle\Psi|$, we find

$$\mathrm{tr}(\mathsf{SWAP}\,|\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi|) = 1\,,$$

whereas if we perform the swap test on $\frac{1}{d}\mathbb{1}$ we find

$$\operatorname{tr}(\mathsf{SWAP}\,\frac{\mathbb{1}}{d}\otimes\frac{\mathbb{1}}{d}) = \frac{1}{d}\,.$$

If d is large, then the difference between the 'pure state' and 'maximally mixed state' swap tests is stark; the former is one, and the latter is $1/d$ which is close to zero.

Example 3: A completeness identity for orthonormal operator bases. For our final example, we derive a rather interesting (and useful) identity. First we require a definition.

Definition 46 (Hilbert-Schmidt inner product)**.** *Consider $\mathbb{C}^{d\times d}$ as a vector space of $d\times d$ matrices. We can turn it into a Hilbert space in its own right via the* ***Hilbert-Schmidt inner product***

$$\langle A,B\rangle_{\mathrm{HS}} := \operatorname{tr}(A^\dagger B) = \sum_{i,j=1}^{d} A^*_{ij}B_{ij}\,.$$

Let $\{M_i\}_{i=1}^{d^2}$ be a complete orthonormal basis of linear operators on $\mathbb{C}^d$. (We note that there must be d^2 such basis elements since the dimension of the space of $d\times d$ matrices is d^2.) Here we mean orthonormal with respect to the Hilbert-Schmidt inner product. Orthonormality means that

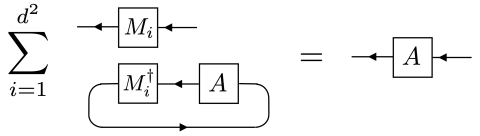
$$\langle M_i,M_j\rangle_{\mathrm{HS}} = \operatorname{tr}(M_i^\dagger M_j) = \delta_{ij}\,, \tag{18}$$

and completeness means that any operator A can be written as $A=\sum_{i=1}^{d^2} c_i\,M_i$ for some coefficients c_i. In fact, using (18) fixes the c_i's to be

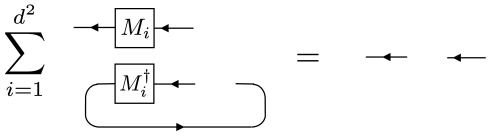
$$A = \sum_{i=1}^{d^2} \operatorname{tr}(M_i^\dagger A)\,M_i\,, \tag{19}$$

namely $c_i = \langle M_i,A\rangle_{\mathrm{HS}} = \operatorname{tr}(M_i^\dagger A)$.

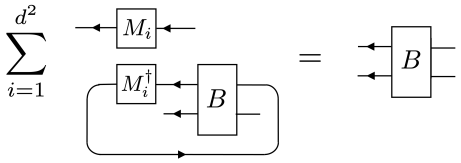
We can write (19) diagrammatically as



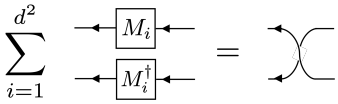
and since it holds for all A, we can remove the A to find the tensor identity



Since the above is a tensor identity, we are welcome to stick in an operator B on $\mathbb{C}^d \otimes \mathbb{C}^d$; this gives



Now if we happen to choose $B = \mathsf{SWAP}$, we find



where we have labeled the ends of the tensor legs with indices for clarity. The above can be written in an algebraic form as the identity

$$\sum_{i=1}^{d} M_i \otimes M_i^\dagger = \mathsf{SWAP}\,.$$

This is a striking identity. As an example if $d = 2^n$, we can let $\{M_i\}_{i=1}^{4^n}$ be the set of n-qubit normalized Pauli strings $\{\frac{1}{2^{n/2}} P_i\}_{i=1}^{4^n}$ which form an orthonormal basis of $\mathbb{C}^{2^n \otimes 2^n}$ with respect to the Hilbert-Schmidt inner product. Thus we find

$$\frac{1}{2^n} \sum_{i=1}^{4^n} P_i \otimes P_i = \mathsf{SWAP}\,,$$

where we have used that $P_i^\dagger = P_i$ for Pauli strings. Since the Paulis include the identity operator, sometimes the above is rearranged as

$$\sum_{P_i \neq \mathbb{1}} P_i \otimes P_i = 2^n\, \mathsf{SWAP} - \mathbb{1} \otimes \mathbb{1}\,.$$

Part 2

Learning General States

CHAPTER 4

Algorithms for State Tomography

One of the most fundamental objects in quantum theory is the quantum state, or density matrix. Therefore, one of the most fundamental problems in quantum learning theory is that of learning a density matrix ρ. More formally, suppose we are given access to a device that prepares some unknown state ρ. We can make measurements on ρ, and then ask for a new copy of ρ. How many copies of ρ do we need to measure such that we can learn ρ to some fixed precision? This is the problem of **quantum state tomography**. The word tomography comes from the Greek 'tomos', meaning 'slice' or 'section', and 'grapho', meaning 'to write' or 'to describe'. In this sense, quantum state tomography results in a description of an unknown quantum state, attained by one measurement or 'slice' at a time.

At first, we will be concerned with the problem of **query complexity**; that is, *how many* ρ's do we need to learn a description of the unknown state to fixed precision? This will depend on exactly what we mean by *fixed precision* (we will see that there are multiple plausible definitions with different features), as well as what kinds of measurements we allow ourselves. As for the latter, we will start by considering **single-copy measurements**, namely where we can only measure one ρ at a time. More generally, we can make entangled measurements on multiple copies of ρ; this will give more optimal query complexity bounds, and we will pursue this later.

It is worth emphasizing that query complexity is distinct from **gate complexity**, which quantifies the number of operations required to perform each measurement. That is, by initially only quantifying query complexity, we are only counting the number of total measurements required, and will not initially be attentive to the difficulty of making each measurement. (We will, however, comment on the difficulties of performing certain kinds of measurements as we go along.) We remark that the gate complexity of a protocol is always lower bounded by the query complexity. Roughly speaking, if we need to query a state k times, then we require *at least* k gate operations. Thus the query complexity at the very least gives us a lower bound on the absolute difficulty of implementing a given protocol in practice.

1. Basic State Tomography

We will consider density matrices ρ on a Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$, as usual. The standard orthonormal basis of $\mathbb{C}^d$ is $\{|i\rangle\}_{i=0}^{d-1}$, and we can consider the matrices $E_{ij} := |i\rangle\langle j|$ where $\{E_{ij}\}_{i,j=0}^{d-1}$ is a complete orthonormal (with respect to Hilbert-Schmidt) basis for $\mathbb{C}^{d\times d}$. Indeed $\mathrm{tr}(E_{ij}^\dagger E_{k\ell}) = \delta_{ik}\delta_{j\ell}$, and we can write any density

matrix ρ as

$$\rho = \sum_{i,j=0}^{d-1} \rho_{ij}\, E_{ij}\,.$$

Our first goal is to provide a protocol for measuring all of the ρ_{ij}'s. Now the E_{ij}'s for $i \neq j$ are not Hermitian operators, we cannot measure $\rho_{ij} = \mathrm{tr}(E_{ij}^\dagger \rho) = \langle i|\rho|j\rangle$ directly. So it is natural to separate our analysis into measuring the diagonal components of ρ, namely ρ_{ii}, and measuring the off-diagonal components of ρ, namely ρ_{ij} for $i \neq j$.

Measuring the diagonal components. To measure the diagonal components of ρ, we need to get good estimates for $\rho_{ii} = \mathrm{tr}(E_{ii}\rho) = \langle i|\rho|i\rangle$ for $i = 0, ..., d-1$. To do so, we can simply measure ρ in the $\{|i\rangle\}_{i=0}^{d-1}$ basis. Upon making a measurement, our apparatus will output 'i' with probability $\langle i|\rho|i\rangle$. So suppose we make N_{diag} total measurements, requiring as many copies of ρ. If N_i is the number of times our measurement apparatus outputted 'i' during those trials, then we can estimate ρ_{ii} by

$$\widehat{\rho}_{ii} := \frac{N_i}{N_{\mathrm{diag}}}\,,$$

which is just the fraction of the times that the apparatus outputted 'i'. We will return soon to estimating the size of N_{diag} such that we can guarantee that e.g. $|\rho_{ii} - \widehat{\rho}_{ii}| \leq \varepsilon$ for all i.

Measuring the off-diagonal components. While the E_{ij} are not Hermitian for $i \neq j$, we can instead consider

$$E_{ij}^+ = |i\rangle\langle j| + |j\rangle\langle i|$$
$$E_{ij}^- = -\mathrm{i}\,(|i\rangle\langle j| - |j\rangle\langle i|)$$

which are Hermitian (in E_{ij}^- the i out front is the imaginary number, not an index label), and satisfy

$$E_{ij} = \frac{1}{2}\left(E_{ij}^+ + \mathrm{i}\, E_{ij}^-\right).$$

Thus if we can estimate $\mathrm{Re}(\rho_{ij}) = \frac{1}{2}\,\mathrm{tr}(E_{ij}^+\rho)$ and $\mathrm{Im}(\rho_{ij}) = -\frac{1}{2}\,\mathrm{tr}(E_{ij}^-\rho)$, then we can estimate $\mathrm{tr}(E_{ij}\rho)$.

To organize these two–outcome blocks into a small number of measurement settings, we group disjoint pairs of indices so that many (i,j)'s are probed simultaneously. To this end, let us fix a decomposition of the complete graph on vertices $\{0, 1, \ldots, d-1\}$ into disjoint perfect matchings $M_1, \ldots, M_T$. For even d one may take $T = d-1$; for odd d one can take $T = d$ where in each round exactly one index is left unpaired (the unpaired index varies from round to round). For each matching M_t we use two measurement settings:

$$\mathsf{R}_t := \left\{\, |\psi_{ij,\pm}^{(R)}\rangle = \tfrac{|i\rangle \pm |j\rangle}{\sqrt{2}} : \{i,j\} \in M_t \,\right\}, \quad \mathsf{I}_t := \left\{\, |\psi_{ij,\pm}^{(I)}\rangle = \tfrac{|i\rangle \pm \mathrm{i}|j\rangle}{\sqrt{2}} : \{i,j\} \in M_t \,\right\},$$

implemented as a block-diagonal projective measurement whose 2×2 blocks are the two-outcome projectors on each pair in M_t (and, if d is odd, singleton projectors for the unpaired index). Now let $N_{\mathrm{off},t}^{\mathrm{R}}$ (respectively $N_{\mathrm{off},t}^{\mathrm{I}}$) denote the number of

copies of ρ we measure in setting R_t (respectively I_t). The total number of copies used for off-diagonals is therefore

$$N_{\text{off}} = \sum_{t=1}^{T} \big(N^{\text{R}}_{\text{off},t} + N^{\text{I}}_{\text{off},t}\big).$$

A single outcome in a given setting contributes information to *all* pairs in that matching. Note that we do not dedicate separate copies to each pair; instead, the copies from round t are shared statistically across the pairs $\{i,j\} \in M_t$.

Fix a round t of the R ("real") setting. For each pair $\{i,j\}$ that appears in round t, write $m^{\text{R},t}_{ij,+}$ and $m^{\text{R},t}_{ij,-}$ for the observed counts of the $+$ and $-$ outcomes in the $\{i,j\}$ two-dimensional block. Then define

$$\widehat{\text{Re}(\rho_{ij})} := \frac{m^{\text{R},t}_{ij,+} - m^{\text{R},t}_{ij,-}}{2\, N^{\text{R}}_{\text{off},t}}\,, \tag{20}$$

which is our estimator for $\text{Re}(\rho_{ij})$.

Analogously, in the I ("imaginary") setting for the same round t, with total shots $N^{\text{I}}_{\text{off},t}$ and counts $m^{\text{I},t}_{ij,\pm}$, we set

$$\widehat{\text{Im}(\rho_{ij})} := \frac{m^{\text{I},t}_{ij,-} - m^{\text{I},t}_{ij,+}}{2\, N^{\text{I}}_{\text{off},t}}. \tag{21}$$

Then combining the above two estimators, we have

$$\widehat{\rho}_{ij} := \frac{m^{\text{R},t}_{ij,+} - m^{\text{R},t}_{ij,-}}{2\, N^{\text{R}}_{\text{off},t}} + \text{i}\,\frac{m^{\text{I},t}_{ij,-} - m^{\text{I},t}_{ij,+}}{2\, N^{\text{I}}_{\text{off},t}}\,.$$

This amounts to an unbiased estimator for ρ_{ij}.

So far we have explained the procedure for estimating the ρ_{ij}'s, but have not specified how many measurements we need to perform so that $|\rho_{ij} - \widehat{\rho}_{ij}| \le \varepsilon$ for all i,j. For this, we require a standard but highly useful concentration inequality:

Theorem 47 (Hoeffding inequality)**.** *Let $Z_1, \ldots, Z_n$ be independent random variables with $Z_k \in [a_k, b_k]$ almost surely and mean $\mu = \mathbb{E}[\frac{1}{n}\sum_{k=1}^n Z_k]$. Then for any $\eta > 0$,*

$$\Pr\left[\left|\frac{1}{n}\sum_{k=1}^{n} Z_k - \mu\right| \ge \eta\right] \le 2\,\exp\!\left(-\frac{2n^2\eta^2}{\sum_{k=1}^n (b_k - a_k)^2}\right).$$

The Hoeffding inequality says that the empirical mean of independent, bounded trials is sharply concentrated around its expectation, with a tail that decays like $\exp(-\text{const} \times n\,\eta^2)$. In our setting each measurement outcome is in $\{+1,-1\}$ or $\{0,1\}$, so a simple instance of the Hoeffding inequality applies directly to each estimated quantity.

We can use our estimators in tandem with the Hoeffding inequality to get our first bound on the sample complexity of quantum state tomography:

Theorem 48 (Basic tomography)**.** *Fix accuracy $\varepsilon \in (0,1)$ and confidence $\delta \in (0,1)$. Then with probability at least $1-\delta$, we can obtain $\widehat{\rho}_{ij}$ such that*

$$|\rho_{ij} - \widehat{\rho}_{ij}| \le \varepsilon \quad \textit{for all } i,j = 0, ..., d-1\,,$$

with at most $O(\frac{d}{\varepsilon^2}\log\frac{d^2}{\delta})$ copies of ρ and as many measurements.

PROOF. We use the Hoeffding inequality stated above and a union bound over all matrix entries.

First we consider obtaining the diagonal entries of ρ. Measuring in the computational basis yields indicators $X_s^{(i)} \in \{0,1\}$ for the event "outcome i," with $\mathbb{E}[X_s^{(i)}] = \rho_{ii}$. The estimator $\widehat{\rho}_{ii} = N_i/N_{\mathrm{diag}} = \frac{1}{N_{\mathrm{diag}}} \sum_{s=1}^{N_{\mathrm{diag}}} X_s^{(i)}$ is therefore an empirical mean of $[0,1]$–bounded variables. Hoeffding's inequality gives us

$$\Pr\big[\, |\widehat{\rho}_{ii} - \rho_{ii}| \geq \eta \,\big] \leq 2\exp\big(-2N_{\mathrm{diag}}\eta^2\big).$$

Choosing, for example, $N_{\mathrm{diag}} \geq \frac{1}{2\varepsilon^2} \log \frac{2d^2}{\delta}$ ensures $\Pr[\, |\widehat{\rho}_{ii} - \rho_{ii}| \geq \varepsilon \,] \leq \delta/d^2$ for each i.

Next we turn to obtaining the off-diagonal entries of ρ. Fix a round t and a pair $\{i,j\} \in M_t$. To obtain the real part of ρ_{ij} in the R_t setting, define per–shot variables

$$X_s^{\mathrm{R},t,(ij)} = \begin{cases} +\frac{1}{2}, & \text{if the outcome is the } + \text{ projector in the } \{i,j\} \text{ block,} \\ -\frac{1}{2}, & \text{if the outcome is the } - \text{ projector in the } \{i,j\} \text{ block,} \\ 0, & \text{if the outcome lies in a different block,} \end{cases}$$

where $X_s^{\mathrm{R},t,(ij)} \in \big[-\frac{1}{2}, \frac{1}{2}\big]$. Then $\widehat{\mathrm{Re}(\rho_{ij})} = \frac{1}{N_{\mathrm{off},t}^{\mathrm{R}}} \sum_{s=1}^{N_{\mathrm{off},t}^{\mathrm{R}}} X_s^{\mathrm{R},t,(ij)}$ (compare with (20)) and thus

$$\mathbb{E}\Big[\widehat{\mathrm{Re}(\rho_{ij})}\Big] = \tfrac{1}{2}\Big(\Pr[+] - \Pr[-]\Big) = \tfrac{1}{2}\,\mathrm{tr}(E_{ij}^{+}\rho) = \mathrm{Re}(\rho_{ij}),$$

and so the estimator is unbiased. Hoeffding's inequality with range length 1 implies

$$\Pr\Big[\, \big|\widehat{\mathrm{Re}(\rho_{ij})} - \mathrm{Re}\,\rho_{ij}\big| \geq \eta \Big] \leq 2\exp\big(-2N_{\mathrm{off},t}^{\mathrm{R}}\eta^2\big).$$

To obtain the imaginary part of ρ_{ij} in the I_t setting, define

$$X_s^{\mathrm{I},t,(ij)} = \begin{cases} +\frac{1}{2}, & \text{if the outcome is the } - \text{ projector in the } \{i,j\} \text{ block,} \\ -\frac{1}{2}, & \text{if the outcome is the } + \text{ projector in the } \{i,j\} \text{ block,} \\ 0, & \text{otherwise,} \end{cases}$$

so that $\widehat{\mathrm{Im}(\rho_{ij})} = \frac{1}{N_{\mathrm{off},t}^{\mathrm{I}}} \sum_{s=1}^{N_{\mathrm{off},t}^{\mathrm{I}}} X_s^{\mathrm{I},t,(ij)}$ (compare with (21)). Using $\mathrm{tr}(E_{ij}^{-}\rho) = -2\,\mathrm{Im}\,\rho_{ij}$, we get

$$\mathbb{E}\Big[\widehat{\mathrm{Im}(\rho_{ij})}\Big] = \tfrac{1}{2}\Big(\Pr[-] - \Pr[+]\Big) = -\frac{1}{2}\,\mathrm{tr}(E_{ij}^{-}\rho) = \mathrm{Im}\,\rho_{ij},$$

which is evidently an unbiased estimator, and Hoeffding's inequality gives

$$\Pr\Big[\, \big|\widehat{\mathrm{Im}(\rho_{ij})} - \mathrm{Im}\,\rho_{ij}\big| \geq \eta \Big] \leq 2\exp\big(-2N_{\mathrm{off},t}^{\mathrm{I}}\eta^2\big).$$

Now impose $N_{\mathrm{off},t}^{\mathrm{R}} = N_{\mathrm{off},t}^{\mathrm{I}} =: N_{\mathrm{off},t}$ for all t and set $\eta = \varepsilon/\sqrt{2}$. Then for each $i \neq j$,

$$\Pr\Big[\, \big|\widehat{\mathrm{Re}(\rho_{ij})} - \mathrm{Re}\,\rho_{ij}\big| \geq \tfrac{\varepsilon}{\sqrt{2}} \,\Big] \leq 2e^{-N_{\mathrm{off},t}\varepsilon^2}, \quad \Pr\Big[\, \big|\widehat{\mathrm{Im}(\rho_{ij})} - \mathrm{Im}\,\rho_{ij}\big| \geq \tfrac{\varepsilon}{\sqrt{2}} \,\Big] \leq 2e^{-N_{\mathrm{off},t}\varepsilon^2}.$$

By a union bound,

$$\Pr[\,|\widehat{\rho}_{ij}-\rho_{ij}|\geq\varepsilon\,]\leq\Pr\Big[\,\big|\widehat{\mathrm{Re}(\rho_{ij})}-\mathrm{Re}\,\rho_{ij}\big|\geq\tfrac{\varepsilon}{\sqrt{2}}\,\Big]+\Pr\Big[\,\big|\widehat{\mathrm{Im}(\rho_{ij})}-\mathrm{Im}\,\rho_{ij}\big|\geq\tfrac{\varepsilon}{\sqrt{2}}\,\Big]$$

$$\leq\;4\,e^{-N_{\mathrm{off},t}\varepsilon^2}.$$

Now there are d diagonal events and $2\binom{d}{2}$ off-diagonal (real or imaginary) events, giving in total d^2 events. Choosing, for example,

$$N_{\mathrm{diag}}\;\geq\;\frac{1}{2\varepsilon^2}\log\frac{2d}{\delta},\quad N_{\mathrm{off},t}\;\geq\;\frac{1}{\varepsilon^2}\log\frac{4d^2}{\delta},$$

makes each event fail with probability at most δ/d^2, whence by a union bound all hold simultaneously with probability at least $1-\delta$. This means that with probability at least $1-\delta$ we have

$$|\widehat{\rho}_{ii}-\rho_{ii}|\leq\varepsilon\quad\text{for all } i=0,\ldots,d-1\,,$$

and with probability at least $1-\delta$ we have

$$|\widehat{\rho}_{ij}-\rho_{ij}|\leq\varepsilon\quad\text{for all }\{i,j\}\in M_t\,.$$

Note that each shot consumes one fresh copy of ρ. We use one diagonal setting with N_{diag} shots and two settings per matching round (R_t and I_t) with $N^{\mathrm{R}}_{\mathrm{off},t}$ and $N^{\mathrm{I}}_{\mathrm{off},t}$ shots per each per round. With $T=d-1$ if d is even and $T=d$ if d is odd, the total is

$$N_{\mathrm{tot}}=N_{\mathrm{diag}}+\sum_{t=1}^{T}\big(N^{\mathrm{R}}_{\mathrm{off},t}+N^{\mathrm{I}}_{\mathrm{off},t}\big)=N_{\mathrm{diag}}+2\sum_{t=1}^{T}N_{\mathrm{off},t}=O\!\left(\frac{d}{\varepsilon^2}\log\frac{d^2}{\delta}\right).$$

This equals the number of measurements performed. The theorem follows. □

There are several ways to quantify the error in approximating ρ. To articulate another way, consider the following definition:

Definition 49 (Frobenius norm)**.** *The* ***Frobenius norm*** *on* $\mathbb{C}^{d\times d}$ *is defined by*

$$\|A\|_F:=\sqrt{\langle A,A\rangle_{\mathrm{HS}}}=\sqrt{\mathrm{tr}(A^\dagger A)}\,,$$

for all $A\in\mathbb{C}^{d\times d}$*.*

Then we have the following corollary of our basic tomography theorem:

Corollary 50 (Frobenius error)**.** *Under the conditions of Theorem 48, with probability at least* $1-\delta$*,*

$$\|\rho-\widehat{\rho}\|_F\leq\sqrt{\sum_{i,j}|\widehat{\rho}_{ij}-\rho_{ij}|^2}\leq\sqrt{d^2\,\varepsilon^2}=d\,\varepsilon.$$

Above, if we choose $\varepsilon=\varepsilon'/d$, then we can get $\|\rho-\widehat{\rho}\|_F\leq\varepsilon'$ with probability at least $1-\delta$. But then this will require $O(\frac{d^3}{\varepsilon'^2}\log\frac{d^2}{\delta})$ copies of ρ and as many measurements.

Remark 51 (Projection to the density-matrix cone preserves and possibly improves Frobenius error)**.** *The empirical matrix* $\widehat{\rho}$ *constructed entrywise need not be positive semidefinite nor have unit trace. Let* $\mathcal{D}:=\{X\succeq 0:\mathrm{tr}(X)=1\}$ *denote the set of*

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:*

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

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

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

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 $\rho, \widehat{\rho}$,*

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

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

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

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

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 $\|O\|_\infty \le 1$ (a very common normalization), (23) gives

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

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

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

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,*

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

Under our entrywise scheme, a uniform per-entry accuracy $\max_{i,j}|X_{ij}| \le \varepsilon_{\max}$ yields $\|X\|_{\mathrm{F}} \le d\,\varepsilon_{\max}$ (and therefore $\|X\|_\infty \le d\,\varepsilon_{\max}$). Consequently, to guarantee an operator-norm target $\|\rho-\widehat{\rho}\,\|_\infty \le \varepsilon$ via entrywise control, one must take $\varepsilon_{\max} = \varepsilon/d$, which (through the Hoeffding scaling $1/\varepsilon_{\max}^2$) 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 $\|\rho-\hat{\rho}\|_\infty \le \varepsilon$ 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

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

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 $D(\theta) = \mathrm{diag}(e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d})$ 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*

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

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

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

so

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

since $\varepsilon \le 1$. Maximality of $\mathcal{M}$ implies it is an ε-net for the Euclidean metric on $\mathbb{S}^{2d-1}$; since $d_{\mathrm{E}} \le \|\cdot\|_2$ 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 $X_1, \dots, X_n$ be independent, mean-zero real random variables. Assume $|X_i| \le b$ almost surely and set $\sigma^2 := \sum_{i=1}^n \mathbb{E}[X_i^2]$. Then for all $t > 0$,*

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

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

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

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 $\mathbb{E}[\exp(\lambda X)] \le \exp(\lambda^2\nu^2/2)$ for all $\lambda \in \mathbb{R}$. This implies Gaussian-type concentration $\Pr(|X| \ge t) \le 2\exp(-t^2/(2\nu^2))$. Bounded variables are sub-Gaussian (with $\nu \lesssim b$ when $|X| \le b$), 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 $\mathbb{E}[\exp(\lambda X)] \le \exp(\frac{\alpha^2\lambda^2}{2})$ for $|\lambda| \le 1/\beta$. 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 $v_1, \dots, v_N$ 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 $1-\delta$,

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

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

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

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 $\sup_{\|u\|=1}\left|\frac{1}{N}\sum_i Y_i\right| = \|H_N(\rho)-\rho\|_\infty$. 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. □

CHAPTER 5

Sample-Optimal Algorithm for State Tomography

In the previous lecture we saw the most basic algorithms for learning quantum states. While their analysis was simple, the algorithm incurs a dependence on the Hilbert space dimension d which is far from optimal. Indeed, the total number of parameters describing an arbitrary mixed state is d^2, so intuitively we expect that the correct sample complexity for state tomography is $\Theta(d^2/\epsilon)$, where ϵ is the target infidelity.[1]

In this lecture, we show using more sophisticated algebraic tools that with a different measurement scheme, one can indeed achieve this optimal sample complexity, up to logarithmic factors. Notably, the algorithm will measure *all copies of the unknown state ρ at once*, rather than simply measuring every qubit of every copy one at a time. So the question becomes: what is the POVM $\{M_\sigma\}$ that one can perform on $\rho^{\otimes N}$ that maximally extracts information about ρ?

1. Some Forced Moves

There are two symmetries we can exploit in the problem. Firstly, there is the trivial permutation symmetry: the "dataset" of copies of ρ being measured is invariant under swapping different copies around. Additionally, the measurement should be agnostic to the eigenbasis of ρ, because we are not making any assumptions about it: if the algorithm achieves some level of statistical efficient for states in some eigenbasis, they had better be equally statistically efficient in any other eigenbasis. Taken together, these two points imply two things about the POVM $\{M_\sigma\}$ we perform:

- **Permutation invariance**: Elements of the POVM should be invariant under conjugation by any permutation operator.
- **Rotation *equi*variance**: If M_σ is the POVM element corresponding to outputting σ, then $M_{U\sigma U^\dagger} = (U^\dagger)^{\otimes N} M_\sigma U^{\otimes N}$ for any $U \in U(d)$.

We will implement such a POVM $\{M_\sigma\}$ in two stages:

(1) Measure $\rho^{\otimes N}$ with a POVM $\{\tilde{M}_\sigma\}$ which is both permutationally and rotationally *in*variant in order to learn the rotationally invariant information about ρ, i.e., its eigenvalues
(2) Apply a suitable rotationally *equi*variant POVM to the post-measurement state to learn the eigenbasis of ρ

[1] The *fidelity* between two states ρ, σ is defined to be $F(\rho, \sigma) \triangleq \mathrm{tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})$, and the *infidelity* is $1 - F(\rho, \sigma)$. Sometimes fidelity is defined to be the square of our definition, but in the regime where infidelity is small, these only affect the definition of infidelity by a constant factor.

There are different ways of implementing the latter step, but the former step is rather generic and useful in other quantum learning tasks as well. It goes by the name of **weak Schur sampling**, and as we will see, its construction is entirely predetermined by the symmetries above.

Remark 61. *We will assume for convenience throughout this lecture that ρ is full rank. This is just for convenience of prose, as we will be talking about representations of the general linear group, but the reasoning below can be extended to states of degenerate rank by taking appropriate limits.*

2. Representation Theory Toolkit

The key ingredient behind weak Schur sampling is **Schur-Weyl duality**, a fundamental algebraic result that, very roughly speaking, ensures that there is a unitary U_{schur} for which

$$U_{\mathsf{schur}}^{\dagger} \rho^{\otimes N} U_{\mathsf{schur}}$$

is block-diagonal with very particular structure in each diagonal block. As such, we may assume without loss of generality that our algorithm first performs a projective measurement to project to the subspace corresponding to one of these diagonal blocks. Weak Schur sampling is precisely this initial projective measurement, which we spell out in detail below.

2.1. Basic Notions

Let $\mathcal{H} \triangleq (\mathbb{C}^d)^{\otimes N}$. Here we introduce just enough representation theory to be able to present Schur-Weyl duality and the full learning algorithm. Representation theory is the study of groups by associating their elements with linear transformations. Throughout this lecture, we work exclusively with representations over complex vector spaces. Given a vector space V, let $\mathrm{GL}(V)$ denote the group of invertible linear transformations on V.

Definition 62. *Given a group G, a (finite-dimensional) **representation** is given by a vector space V and a group homomorphism $\mu : G \to \mathrm{GL}(V)$, i.e., a map satisfying $\mu(gh) = \mu(g)\mu(h)$ for all $g, h \in G$. We say that μ is a G-*representation over V*, or a* representation over V *if G is clear from context. The* dimension *of the representation, denoted* $\dim(\mu)$*, is the dimension of V.*

We will often refer to a representation by the vector space on which G acts, with the homomorphism μ being implicitly understood from context. Similarly, we may write $g \cdot v$ to denote $\mu(g)v$.

In any introductory text on representation theory, one can find a laundry list of examples, for instance the trivial representation, the standard representation of GL_d, the regular representation of any finite group, the Fourier character representation of $\mathbb{Z}_2^n$, etc. The following two representations are most relevant to this lecture:

Example 63. *$\mathcal{S}_N$ and GL_d admit the following representations over $(\mathbb{C}^d)^{\otimes N}$, call them $P : \mathcal{S}_N \to \mathrm{GL}((\mathbb{C}^d)^{\otimes N})$ and $Q : \mathrm{GL}_d \to \mathrm{GL}((\mathbb{C}^d)^{\otimes N})$ respectively. $P(\pi)$ is the permutation operator on N qudits associated to $\pi \in \mathcal{S}_N$, and $Q(M) = M^{\otimes N}$ for $M \in \mathrm{GL}_d$.*

Note that these clearly commute with each other, so we can also define a representation of $\mathcal{S}_N \times \mathrm{GL}_d$, which we will denote by μ_{SW}, over $\mathcal{H}$ via

$$\mu_{\mathsf{SW}}(\pi, M) = P(\pi)Q(M)\,.$$

Definition 64. *A representation (μ, V) is **irreducible** if there does not exist an* invariant subspace*, i.e. a proper subspace $U \subsetneq V$ for which $\mu(g)U = U$ for all $g \in G$. We sometimes call (μ, V) an **irrep** for short.*

Like the vector spaces over which they act, representations can be stitched together using standard operations like direct sum and tensor product.

Definition 65. *The* direct sum *of representations $(\mu_1, V_1), \ldots, (\mu_m, V_m)$ of G with multiplicities $a_1, \ldots, a_m$ is the representation $(\mu, V_1^{\oplus a_1} \oplus \cdots \oplus V_m^{\oplus a_m})$ for which $\mu(g)$ is given by the block diagonal matrix whose diagonal blocks consist of a_i copies of $\mu_i(g)$ for all i. We will sometimes write $\mu = \bigotimes_{i=1}^m a_i \cdot \mu_i$ and $V \cong \sum_{i=1}^m a_i \cdot V_i$.*

The tensor product *of representations (μ, V) and (ν, W) of G is the representation $(\mu \otimes \nu, V \otimes W)$ for which*

$$\mu \otimes \nu(g \cdot h) = \mu(g) \otimes \nu(h)\,.$$

Somewhat confusingly, representations which are reducible need not decompose into a direct sum of irreps in general. Those that do are said to be **semisimple**. While not all representations are semisimple, all of the ones we will care about in this lecture are. For instance, it can be shown that any representation of $\mathcal{S}_N$ or U_d is semisimple – the former is **Maschke's theorem** (see e.g. [**FH13**, Corollary 1.6]), and the latter is immediate from the fact that if U_d preserves some subspace W, then it also preserves the orthogonal complement $W^\perp$. For the rest of this lecture, we will work with the implicit understanding that all representations discussed are semisimple.

Definition 66 (Hom spaces)**.** *Given two G-representations μ, ν over spaces V, W, a linear map $f : V \to W$ is a **G-linear map** if it commutes with the action of G, that is, if $\nu(g)f(v) = f(\mu(g)v)$ for all $v \in V$. We denote by $\mathrm{Hom}_G(V, W)$ the space of all G-linear maps $V \to W$. We say that representations μ and ν are **isomorphic**, which we denote by $V \cong W$ when μ and ν are clear from context, if there is a G-linear map $f : V \to W$ which is an isomorphism of vector spaces.*

The following is an elementary but extremely useful fact about G-linear maps:

Lemma 67 (Schur's lemma)**.** *Let V and W be irreps of G.*

(1) If $V \ncong W$, then $\mathrm{Hom}_G(V, W)$ consists of only the zero map.
(2) If $V = W$, then $\mathrm{Hom}_G(V, W)$ consists of all scalar multiples of the identity map.

Proof. For the first part, suppose to the contrary that there is a nonzero G-linear map $f : V \to W$. If it has a nontrivial kernel $V' \subsetneq W$, then note that $f(g \cdot v') = g \cdot f(v') = 0$, so the kernel is stable under the action of G, contradicting the assumption that V is an irrep. So f is injective. An identical argument for the image of f shows that f is surjective, completing the proof of the first part.

For the second part, we need to prove that apart from scalar multiples of the identity, there are no other elements in $\mathrm{Hom}_G(V, V)$. Let $f \in \mathrm{Hom}_G(V, V)$, and consider $f' \triangleq f - \lambda \mathrm{Id}_V$ for any eigenvalue λ of f. As the sum of G-linear maps is G-linear, f' is G-linear, and furthermore it has nontrivial kernel as it vanishes on

the eigenvector of f associated to λ. But its kernel cannot be a proper subspace of V or, as in the proof of the first part, this would contradict the fact that V is an irrep. So the kernel of f' must be all of V, implying that $f' \equiv 0$ and thus that $f = \lambda \mathrm{Id}_V$ as claimed. □

Finally, an important object in the study of representations is their associated *characters*, which basic information about the representation like their dimension.

Definition 68 (Characters)**.** *Given a G-representation over V, its **character** is the map $\chi : G \to \mathbb{C}$ given by $\chi(g) = \mathrm{tr}(\mu(g))$. Note that if g is the identity element* id*, then $\chi(\mathrm{id}) = \dim(V)$.*

2.2. Representation Theory of the Symmetric Group

Here we give a complete classification of all irreps of the symmetric group.

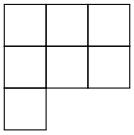
We begin with a useful shift in perspective. An equivalent way to think about representations of finite groups G like the symmetric group is in terms of *modules* over the *group algebra* associated to G.

Definition 69 (Group algebras)**.** *Given a finite group G, the associated **group algebra** $\mathbb{C}[G]$ is the vector space over $\mathbb{C}$ of formal linear combinations $\sum_{g \in G} a_g g$ for $a_g \in \mathbb{C}$, additionally equipped with the* multiplication *operation $(\sum_{g \in G} a_g g)(\sum_{h \in G} b_h h) = \sum_{g,h \in G} a_g b_h \cdot gh$.*

*Any G-representation μ over V naturally gives rise to an algebra homomorphism $\mathbb{C}[G] \to \mathrm{GL}(V)$ by extending linearly; the latter equips the vector space V with the structure of a $\mathbb{C}[G]$-*module*, and sub-representations of μ then correspond to* sub-modules *of V. This identification also goes in the reverse direction: any such algebra homomorphism gives rise to a G-representation by restricting to the elements $G \subset \mathbb{C}[G]$.*

We now define the central objects in the representation theory of the symmetric group.

Definition 70 (Young tableaux)**.** *A **Young diagram** is a sequence of rows of boxes of nonincreasing length, like the following:*



Its shape *is the tuple of row lengths; for example, the shape of the above is $(3, 3, 1)$.*

*A **Young tableau** T with entries in $[d]$ is a Young diagram where each box is labeled with a number from $[d]$, e.g.*

1	5	3
4	2	2
5		

*The Young tableau with **canonical labeling** is given by labeling the entries from left to right and top to bottom with the numbers $1, 2, \cdots$ in increasing order.*

*A Young tableau is said to be **semi-standard** if every row consists of entries in non-decreasing order, and if every column consists of entries in strictly increasing order, e.g.,*

1	1	3
2	3	4
5		

It is further said to be ***standard*** *if the rows are also strictly increasing, though we will not use this notion in this lecture.*

Any Young diagram is naturally associated with a ***partition*** *via its shape: for instance $\lambda = (3,3,1)$ is a partition of $N = 7$. We denote such a partition with the notation $\lambda \vdash [N]$. If the partition/shape consists of m entries and we refer to the j-th entry of λ for $j > m$, by default we set $\lambda_j = 0$.*

Given a partition $\lambda = (\lambda_1, \dots, \lambda_m) \vdash [N]$, we can associate a probability distribution over $[m]$ which places mass λ_i/N on element i. We will refer to the vector of probabilities $(\lambda_1/N, \dots, \lambda_m/N)$ by $\overline{\lambda}$.

Definition 71 (Young symmetrizer)**.** *Let T be a Young tableau. Define $P_T \subseteq \mathcal{S}_N$ (resp. $Q_T \subseteq \mathcal{S}_N$) to consist of permutations which preserve the rows (resp. columns) of T. Define the group algebra elements $a_T, b_T \in \mathbb{C}[\mathcal{S}_N]$ by $a_T \triangleq \sum_{p \in P_T} p$ and $b_T \triangleq \sum_{q \in Q_T} \operatorname{sgn}(q) q$, and define the corresponding* ***Young symmetrizer*** *to be $c_T \triangleq a_T b_T$.*

If T is the canonical labeling for shape λ, we use $a_\lambda, b_\lambda, c_\lambda$ to denote a_T, b_T, c_T.

The Young symmetrizer is ultimately just some cleverly chosen linear combination of permutation operators, but it will play a central role not just in the classification of the irreps of the symmetric group, but also in our characterization of the representation μ_{SW} from Example 63.

Definition 72 (Specht module)**.** *Define the* ***Specht module***

$$V_\lambda \triangleq \mathbb{C}[\mathcal{S}_N] c_\lambda \,.$$

In this section, our goal is to show that the irreps of $\mathcal{S}_N$ are precisely the Specht modules for different $\lambda \vdash [N]$. First, let us get some intuition for what they look like:

Example 73. *Consider the case of $N = 3$, for which there are three possible partitions: $(3), (2,1), (1,1,1)$. When the partition is (3), the corresponding Young tableau with canonical labeling is*

1	2	3

and a_T is a sum over all permutations in $\mathcal{S}_3$, while b_T is the identity. The Young symmetrizer in this case is the sum over all permutations, and the Specht module is simply the line spanned by $\sum_\pi \pi$. One can see that this is simply the 1-dimensional trivial representation *that maps all group elements to the identity.*

When the partition is $(1,1,1)$, the corresponding Young tableau with canonical labeling

1
2
3

and the Young symmetrizer is given by $\sum_\pi \operatorname{sgn}(\pi)\pi$. The Specht module is again just the line spanned by this element, and one can see that this is the 1-dimensional sign representation *that maps all group elements to their sign.*

When the partition is $(2,1)$*, the corresponding Young diagram is*

1	2
3	

and a_T *is the sum of the identity permutation* id *and* τ_{12}*, the transposition of elements* 1 *and* 2*, while* b_T *is the difference between the identity and* τ_{13}*, the transposition of elements* 1 *and* 3*. The Young symmetrizer is given by*

$$c_\lambda = (\mathrm{id} + \tau_{12})(\mathrm{id} - \tau_{13}) = \mathrm{id} + \tau_{12} - \tau_{13} - \pi\,,$$

where π *is the permutation* $(1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 1)$*. A calculation shows that the Specht module in this case is two-dimensional, spanned by* c_λ *and* $\tau_{13}c_\lambda$*, and is in fact isomorphic to the following representation, sometimes called the* standard representation*. Consider the two-dimensional subspace of* $\mathbb{R}^3$ *given by vectors with coordinates summing to zero. There is a natural action of* $\mathcal{S}_3$ *on this space, i.e., permuting the coordinates of such a vector keeps it in that subspace.*

One would be hard pressed to come up with other irreps of $\mathcal{S}_3$*, and indeed there are none: as we will show, the Specht modules make up all irreps of* $\mathcal{S}_3$ *up to isomorphism!*

Lemma 74. *Let* T *be a Young tableau. For any* $g \in \mathbb{C}[\mathcal{S}_N]$*,* $a_T g b_T$ *is a multiple of* c_T*. In particular,* $c_T\mathbb{C}[\mathcal{S}_N]c_T \subseteq \mathbb{C}c_T$*.*

Proof. It suffices to show this for $g = \pi$ for $\pi \in \mathcal{S}_N$. We have

$$a_T \pi b_T = \sum_{p \in P_T, q \in Q_T} \mathrm{sgn}(q) p\pi q\,.$$

If $\pi = pq$ for $p \in P_T, q \in Q_T$, then $a_T \pi b_T = \mathrm{sgn}(q) c_T$ as desired.

We will show that for $\pi \notin P_T Q_T$, $a_T \pi b_T = 0$. To show this, we show there is a *transposition* $p' \in P_T$ for which $q' \triangleq \pi^{-1} p' \pi \in Q_T$. In this case, $\pi = p'\pi q'$, so

$$a_T \pi b_T = \sum_{p \in P_T, q \in Q_T} \mathrm{sgn}(q) p p' \pi q' q = -\sum_{p \in P_T, q \in Q_T} \mathrm{sgn}(q) p \pi q = -a_T \pi b_T = 0\,,$$

and we would be done (note that in the penultimate step we used the fact that $\mathrm{sgn}(qq') = \mathrm{sgn}(q)\mathrm{sgn}(q') = \mathrm{sgn}(q)\mathrm{sgn}(p') = -\mathrm{sgn}(q)$ as p' is a transposition).

To show the existence of the transposition p', define T' by mapping every entry x in T to $\pi(x)$. We want to show there are two numbers x, y which appear in the same row of T and the same column of T'. If to the contrary such x, y do not exist, then we can shuffle the entries in each column of T' so that the first row of T' is the same as the first row of T up to permutation. Then we can proceed to shuffle the entries in each column of T', keeping the first row fixed, so that the second row agrees with that of T up to permutation. Continuing in this fashion, we end up with permutations $p \in P_T$ and $q' \in Q_{T'} = \pi Q_T \pi^{-1}$ for which $p \cdot T = q' \cdot T'$. Writing $q' = \pi q \pi^{-1}$ for $q \in Q_T$, we find that $p \cdot T = \pi q \pi^{-1} T' = \pi q \cdot T$, so $\pi = pq^{-1} \in P_T Q_T$, a contradiction. □

Next, we re-use the proof strategy in the last two paragraphs of the proof of Lemma 74 to show the following:

Lemma 75. *Let* λ, μ *be partitions for which* $\lambda > \mu$ *in lexicographic ordering. Then* $a_\lambda g b_\mu = 0$ *for all* $g \in \mathbb{C}[\mathcal{S}_N]$*. In particular,* $c_\lambda \mathbb{C}[\mathcal{S}_N] c_\mu = 0$*.*

PROOF. As before, it suffices to show this for $g = \pi$. Let T and $\tilde{T}$ denote the tableaux corresponding to λ, μ. By the reasoning in Lemma 74, letting T' denote the result of mapping every entry x in $\tilde{T}$ to $\pi(x)$. As in the proof of Lemma 74, we want to show there are two numbers x, y which appear in the same row of T and the same column of T'. Suppose to the contrary.

If $\lambda_1 > \mu_1$, then because there are strictly fewer than λ_1 columns in T', some column must contain two numbers from the first row of T, a contradiction. If $\lambda_1 = \mu_1$, then as in the proof of Lemma 74, we can shuffle the entries in the columns of T' and in the first row of T so that T and T' have identical first row, and we can recurse on the subsequent rows until we reach a row i for which $\lambda_i > \mu_i$, inducing the desired contradiction. □

This yields the following characterization for the irreps of $\mathcal{S}_N$.

Lemma 76. *(i) V_λ is an irrep of $\mathcal{S}_N$ for any $\lambda \vdash [N]$.*
(ii) For any distinct partitions $\lambda, \mu \vdash [N]$, V_λ and V_μ are not isomorphic.
(iii) Any irrep of $\mathcal{S}_N$ is isomorphic to some V_λ.

PROOF. See pset 2 for the proofs of (i) and (ii).

The proof of (iii) follows from the standard fact that the number of irreps of any finite group is equal to the number of conjugacy classes (see e.g. [**FH13**, Proposition 2.30]), and the fact that the conjugacy classes of $\mathcal{S}_N$ are in one-to-one correspondence with the partitions $\lambda \vdash [N]$. □

Lemma 77 (Hook length formula)**.** *Given an entry in a Young tableau T of shape λ, let its **hook length** denote the number of boxes either directly below it or directly to its right. Then $\dim(V_\lambda)$ is $n!$ divided by the product of the hook lengths of all entries T. In particular,*

$$\dim(V_\lambda) \leq e^{NH(\overline{\lambda})}.$$

PROOF. The first part is a standard fact whose proof would take us too far afield. Recall from Lemma 74 that c_T is idempotent up to scaling, that is, $c_T^2 = n_T c_T$ for some constant n_T. The hook length formula is equivalent to the claim that n_T is given by the product of hook lengths. A proof of this can be found, e.g., in [**Gri25**, Section 5.11.2].

As for the inequality in the second part of the lemma, note that the hook length of the entry in the i-th row and j-th column is at least $\lambda_i - j + 1$, so the product of hook lengths in row i is at least $\lambda_i!$. We thus have

$$\dim(V_\lambda) \leq \frac{N!}{\prod_i \lambda_i!} \leq e^{NH(\overline{\lambda})},$$

where in the last step we used the elementary fact that the logarithm of a multinomial coefficient $\binom{N}{\lambda_1 \cdots \lambda_m}$ is upper bounded by N times the entropy of the distribution with probability mass function given by $\overline{\lambda}$, see e.g. [**CS+04**, Lemma 2.2]. □

In short, the irreps of the symmetric group are in one-to-one correspondence with the Young diagrams $\lambda \vdash [N]$. Remarkably, this is also the case in a certain sense for $\mathrm{GL}(\mathbb{C}^d)$, and furthermore these irreps of $\mathrm{GL}(\mathbb{C}^d)$ are intimately tied to those of a $\mathcal{S}_N$. This is the content of Schur-Weyl duality, which we discuss in the next section.

2.3. Schur-Weyl Duality

Recall the representation μ_{SW} defined in Example 63. The following decomposition result is the key ingredient behind the learning algorithm we will describe in the next section.

Lemma 78 (Isotypic decomposition)**.** *The representation μ_{SW} of $\mathcal{S}_N \times \mathrm{GL}_d$ over $\mathcal{H}$ decomposes as*

$$\mathcal{H} \cong \bigoplus_{\lambda \vdash [N]} V_\lambda \otimes \mathbb{S}_\lambda V_\lambda \,,$$

*where the **Schur functor** $\mathbb{S}_\lambda V_\lambda \triangleq \mathrm{Hom}_{\mathcal{S}_N}(V_\lambda, \mathcal{H})$ is equipped with the natural GL_d-action of composition.*

Proof. The proof is essentially just symbol pushing, but the intuition is that every irrep V_λ appears with a certain multiplicity in the decomposition of μ_{SW} as a $\mathcal{S}_N$-representation, and the spaces $\mathbb{S}_\lambda V_\lambda$ are simply there to track these multiplicities. This argument is not specific to $\mathcal{S}_N, \mathrm{GL}_d$ and only uses the fact that their actions commute.

First note that $\mathbb{S}_\lambda V_\lambda$ is a valid representation of GL_d: given $f \in \mathrm{Hom}_{\mathcal{S}_N}(V_\lambda, \mathcal{H})$ and $g \in \mathrm{GL}_d$, we have $g \cdot f = g \circ f \in \mathrm{Hom}_{\mathcal{S}_N}(V_\lambda, \mathcal{H})$ as

$$(g \circ f)(\pi \cdot v) = g(\pi \cdot f(v)) = \pi \cdot (g \circ f)(v)$$

for any $v \in V_\lambda, \pi \in \mathcal{S}_N$, where the first step follows by the fact that f is $\mathcal{S}_N$-linear, and the second step follows by the fact that the actions of $\mathcal{S}_N$ and GL_d on $\mathcal{H}$ commute.

Now consider the map $\Phi : \oplus_\lambda V_\lambda \otimes \mathbb{S}_\lambda V_\lambda \to \mathcal{H}$ given by linearly extending the maps $v \otimes f \mapsto f(v)$ for $v \in V_\lambda, f \in \mathbb{S}_\lambda V_\lambda$. One can readily check that this map is $\mathcal{S}_N \otimes \mathrm{GL}_d$-linear. It remains to show it is bijective.

To show surjectivity: if we consider the decomposition of $\mathcal{S}_N$-representations $\mathcal{H} \cong \oplus_\lambda V_\lambda^{\oplus m_\lambda}$, then by Schur's lemma, $\mathrm{Hom}_{\mathcal{S}_N}(V_\lambda, \mathcal{H})$ is spanned by the embeddings $\iota_{\lambda_1}, \ldots, \iota_{\lambda_{m_\lambda}} : V_\lambda \to \mathcal{H}$ of V_λ into the m_λ copies of V_λ within $\mathcal{H}$. Now observe that $\Phi(V_\lambda \otimes \{\iota_{\lambda_j}\}) = \iota_{\lambda_j}(V_\lambda) = V_\lambda$, so the image of Φ contains every component in the decomposition $\mathcal{H} \cong \oplus_\lambda V_\lambda^{\oplus m_\lambda}$, establishing surjectivity.

Finally, Φ is bijective because the domain and image have the same dimension $\sum_\lambda m_\lambda \dim(V_\lambda)$. □

The following gives a simple characterization of the Schur functors in terms of Young symmetrizers, using the fact that Young symmetrizers are idempotent up to a scaling.

Lemma 79. $\mathbb{S}_\lambda V_\lambda \cong c_\lambda \cdot \mathcal{H}$

Proof. See pset 2. □

Remark 80 (Schur-Weyl duality)**.** *A beautiful fact is that these GL_d-representations $\mathbb{S}_\lambda(V_\lambda)$ are actually* irreps*; this is the content of the famous **Schur-Weyl duality**. This is a consequence of two facts that we will not prove:*

- *Not only do the actions of $\mathcal{S}_N$ and GL_d on $\mathcal{H}$ commute with each other, but any operator that commutes with $Q(M)$ for all $M \in \mathrm{GL}_d$ must be a linear combination of the operators $\{P(\pi)\}_{\pi \in \mathcal{S}_N}$.*

- *The **double commutant theorem**: in situations where one has an isotypic decomposition like in Lemma 78, this states that one action being the* commutant *of the other in the sense of the above bullet point allows one to further deduce that the Schur functors in the isotypic decomposition are irreps.*

The first fact has a remarkably elementary proof, see https://math.univ-lyon1.fr/~aubrun/recherche/schur-weyl.pdf*. The second fact can be found in any standard representation theory text, e.g.* [EGH+09, Theorem 5.18.1].

At this juncture, what we really need is a user-friendly description of the *characters* of the representations $\mathbb{S}_\lambda(V_\lambda)$. As we will see, these are given by the **Schur polynomials** s_λ. This will follow from the following lemmas:

Lemma 81. *Let $|\vec{i}\rangle = |i_1\rangle \otimes \cdots \otimes |i_N\rangle$ be a tensor product of basis states, i.e. $i_1, \ldots, i_N \in [d]$. For every $j \in [d]$, suppose it appears ν_j times among $i_1, \ldots, i_N$. Then for any $\lambda \vdash [N]$ and any diagonal matrix $D = \mathrm{diag}(x_1, \ldots, x_d)$, we have*

$$D^{\otimes N} c_\lambda |\vec{i}\rangle = x_1^{\nu_1} \cdots x_d^{\nu_d} c_\lambda |\vec{i}\rangle \ .$$

Proof. For any $\pi \in \mathcal{S}_N$, note that $D^{\otimes N} P(\pi) |\vec{i}\rangle = x_1^{\nu_1} \cdots x_d^{\nu_d}$. As c_λ is a linear combination of $P(\pi)$'s, the claim immediately follows. □

Definition 82. *Given $\lambda, \lambda' \vdash [N]$, we say that λ **majorizes** λ', denoted $\lambda' \prec \lambda$, if $\sum_{i \le j} \lambda'_i \le \sum_{i \le j} \lambda_i$ for all j.*

Lemma 83. *Let ν_j denote the number of times j appears among $i_1, \ldots, i_N$. Let $\nu^{\mathsf{sorted}} \vdash [N]$ denote the partition given by sorting the entries of ν in decreasing order.*

Then $c_\lambda |\vec{i}\rangle = 0$ if $\nu^{\mathsf{sorted}} \not\prec \lambda$. As a special case, this implies that $c_\lambda |\vec{i}\rangle = 0$ if λ has more than d rows.

Proof. We can associate to $|\vec{i}\rangle$ a Young tableau T with shape λ by filling in the entries of $\vec{i}$ from left to right and top to bottom. If any column of T has a repeated entry, then $c_\lambda |\vec{i}\rangle = 0$.

The condition that $\nu^{\mathsf{sorted}} \not\prec \lambda$ implies that there is some j for which $\sum_{i \le j} \nu_i^{\mathsf{sorted}} > \sum_{i \le j} \lambda_i$, which implies that some column of T has a repeated entry by pigeonhole principle. □

Corollary 84. *Let $q_\lambda : \mathrm{GL}_d \to \mathrm{GL}(\mathbb{S}_\lambda V_\lambda)$ denote the GL_d-representation associated to the Schur functor $\mathbb{S}_\lambda V_\lambda$. For any $M \in \mathrm{GL}_d$ with eigenvalues $x_1, \ldots, x_d$, we have*

$$\mathrm{tr}(q_\lambda(M)) = \sum_{\mathrm{SSYT}\ T} x^T \triangleq s_\lambda(\vec{x}) \,,$$

where the sum ranges over all semi-standard Young tableaux with shape λ over alphabet $[d]$, and x^T denotes the monomial $x_1^{\nu_1} \cdots x_d^{\nu_d}$ where ν_j is the number of occurrences of entry j in T. The polynomials s_λ are called the Schur polynomials*.*

Proof. A fact we will need but will not prove is that $c_\lambda |\vec{i}\rangle$ for any $|\vec{i}\rangle$ which does not correspond to a semi-standard Young tableau can be expressed as a linear combination of $c_\lambda |\vec{j}\rangle$ for $|\vec{j}\rangle$'s which do correspond to semi-standard Young tableau. This is a nontrivial fact, a consequence of the so-called *Garnir relations*, whose proof is out of the scope of these notes. The upshot of this fact is that the collection of $\{c_\lambda |\vec{i}\rangle\}$ for all $|\vec{i}\rangle$ which do correspond to semi-standard Young tableaux spans $\mathbb{S}_\lambda V_\lambda$, in fact, forms a basis.

Next, note that any character χ of a representation μ of GL_d only depends on the eigenvalues of the input: if $M \in \mathrm{GL}_d$ has diagonalization $M = U^{-1}DU$, then $\chi(M) = \chi(U^{-1}DU) = \mathrm{tr}(\mu(U)^{-1}\mu(D)\mu(U)) = \mathrm{tr}(\mu(D)) = \chi(D)$.

So we may assume that M in the corollary is diagonal, in which case Lemmas 81 and 83 imply that

$$\mathrm{tr}(q_\lambda(M)) = \sum_{\nu : \nu^{\mathsf{sorted}} \preceq \lambda} K_{\lambda\nu} x_1^{\nu_1} \cdots x_d^{\nu_d} ,$$

where ν ranges over all *ordered* partitions of $[N]$ and $K_{\lambda\nu}$ denotes the number of semi-standard Young tableaux with shape λ over alphabet $[d]$ such that each $j \in [d]$ occurs ν_j times.[2] The result then follows from the fact that $K_{\lambda\nu} = 0$ if $\nu^{\mathsf{sorted}} \not\preceq \lambda$, which we do not prove here. □

2.4. Schur Polynomial Facts

We will need two simple facts about the Schur polynomials.

Lemma 85. *For any partition $\lambda \vdash [N]$ with at most d rows,*

$$\dim(\mathbb{S}_\lambda V_\lambda) = s_\lambda(1^d) \leq N^{O(d^2)} .$$

Proof. The first equality follows from the fact that $s_\lambda(1^d) = \mathrm{tr}(q_\lambda(\mathrm{Id}))$ by Corollary 84, together with the fact that the character of a representation evaluated at the identity is the dimension of the representation.

The inequality is typically proved by invoking the identity

$$s_\lambda(1^d) = \prod_{1 \leq i < j \leq d} \frac{\lambda_i - \lambda_j + j - i}{j - i}$$

together with the fact that $0 \leq \lambda_i - \lambda_j \leq N$, but the proof for this identity is nontrivial. Instead, we can prove the claimed bound by the following simple combinatorial argument. Recall that $s_\lambda(1^d)$ counts the number of SSYT of shape λ with entries from $[d]$. Given any such SSYT T, we can consider the sequence of SSYT's $(T^{\leq j})$ given by removing all blocks with an entry larger than j. Note that the shapes $\lambda^{\leq j}$ of these SSYT's can be used to uniquely recover the original T, because the difference between $T^{\leq j}$ and $T^{\leq j-1}$ specifies the locations of the j entries withint T. The number of such sequences of shapes $(\lambda^{\leq j})$ can be naively bounded by $N^{O(d^2)}$. □

Lemma 86. *Given two d-dimensional density matrices σ, ρ, for any partition $\lambda \vdash [N]$ with at most d rows,*

$$\mathrm{tr}(q_\lambda(\rho\sigma)) \leq \dim(\mathbb{S}_\lambda V_\lambda) \cdot e^{-2NH(\overline{\lambda})} F(\rho, \sigma)^{2N} .$$

Here $\overline{\lambda}$ denotes the distribution over $[d]$ with probability mass function given by $(\lambda_i/N)_{i \in [d]}$, and $H(\cdot)$ denotes Shannon entropy.

Proof. Let X be a psd matrix with eigenvalues $x_1 \geq \cdots \geq x_d$ sorted in non-increasing order, and define $\overline{x}_i = x_i/\mathrm{tr}(X)$.

By Fact 87 and the fact that there are $\dim(\mathbb{S}_\lambda V_\lambda)$ monomials in the definition of s_λ, we can bound

$$\mathrm{tr}(q_\lambda(X^2)) = s_\lambda(x_1^2, \ldots, x_d^2) \leq \dim(\mathbb{S}_\lambda V_\lambda) \cdot x_1^{2\lambda_1} \cdots x_d^{2\lambda_d} .$$

[2] Note that if ν and ν' are equal up to permutation of entries, $K_{\lambda\nu} = K_{\lambda\nu'}$, though this is not *a priori* clear.

Note that

$$\begin{aligned} x_1^{2\lambda_1} \cdots x_d^{2\lambda_d} &= \overline{x}_1^{2\lambda_1} \cdots \overline{x}_d^{2\lambda_d} \cdot \operatorname{tr}(X)^{2N} \\ &= \exp(2N \mathbb{E}_{i \sim \overline{\lambda}} \log \overline{x}_i) \cdot \operatorname{tr}(X)^{2N} \\ &\leq \exp(-2NH(\overline{\lambda})) \cdot \operatorname{tr}(X)^{2N} , \end{aligned}$$

where in the last step we used Gibbs' inequality.

To conclude the proof, take $X = \sqrt{\sigma^{1/2} \rho \sigma^{1/2}}$ so that $\operatorname{tr}(X) = F(\rho, \sigma)$. As the Schur polynomials are characters, $\operatorname{tr}(q_\lambda(\rho\sigma)) = \operatorname{tr}(q_\lambda(\sigma^{1/2}\rho\sigma^{1/2})) = \operatorname{tr}(q_\lambda(X^2))$, concluding the proof. □

The above argument uses the following elementary fact:

Fact 87. *Given any partitions $\nu \preceq \lambda$ with at most d rows, and any nonnegative reals $x_1 \geq \cdots \geq x_d$,*

$$x_1^{\nu_1} \cdots x_d^{\nu_d} \leq x_1^{\lambda_1} \cdots x_d^{\lambda_d} .$$

Finally, we need the following simple lower bound on the Schur polynomials:

Lemma 88. *For any $\lambda \vdash [N]$ $s_\lambda(\overline{\lambda}) \geq e^{-NH(\overline{\lambda})}$.*

Proof. Every Schur polynomial is a sum of monomials and thus lower bounded by any single monomial. Consider the SSYT T with shape λ that has λ_j j's in the j-th row. Then

$$s_\lambda(\overline{\lambda}) \geq \overline{\lambda}^T = \overline{\lambda}_1^{N\overline{\lambda}_1} \cdots \overline{\lambda}_d^{N\overline{\lambda}_d} = e^{-NH(\overline{\lambda})}$$

as claimed. □

3. Weak Schur Sampling

As discussed at the beginning of the previous section, the upshot of the isotypic decomposition in Lemma 78 is that there is a certain unitary U_{schur} over $\mathcal{H}$ that block-diagonalizes all states of the form $\rho^{\otimes N}$. More generally, for any $\pi \in \mathcal{S}_N$ and $M \in \mathrm{GL}_d$, we have

$$U_{\mathsf{schur}} P(\pi) Q(\rho) U_{\mathsf{schur}}^\dagger = \bigoplus_{\lambda \vdash [N]} p_\lambda(\pi) \otimes q_\lambda(M) ,$$

where p_λ is the $\mathcal{S}_N$-representation of π acting on the Specht module $V_\lambda \cong \mathbb{C}[\mathcal{S}_N] \cdot c_\lambda$, and q_λ is the GL_d-representation whose character is given by the Schur polynomial s_λ. In particular, when $\pi = \mathrm{Id}$ and $M = \rho$, then we obtain the block decomposition

$$U_{\mathsf{schur}} \rho^{\otimes N} U_{\mathsf{schur}}^\dagger = \bigotimes_{\lambda \vdash [N]} \mathrm{Id}_{V_\lambda} \otimes q_\lambda(\rho) .$$

The reason this block decomposition is helpful is that we have a good understanding of the blocks $q_\lambda(\rho)$ and how they behave when the copies of ρ get rotated.

The first step of our learning algorithm is thus to rotate the input $\rho^{\otimes N}$ into the basis prescribed by U_{schur} and perform a projective measurement onto one of these blocks, a procedure called *weak Schur sampling*.

Definition 89 (Weak Schur Sampling)**.** *Let ρ be a density matrix with eigenvalues $\overline{\lambda}_1^* \geq \cdots \geq \overline{\lambda}_d^* \geq 0$.*

Let Π_λ denote the projector to the isotypic component $V_\lambda \otimes \mathbb{S}_\lambda V_\lambda$. Weak Schur sampling *is the following procedure:*

(1) *Perform a projective measurement* $\{\Pi_\lambda\}_\lambda$ *on* $\rho^{\otimes N}$ *to obtain a state in* $V_\lambda \otimes \mathbb{S}_\lambda V_\lambda$ *with probability*

$$\operatorname{tr}(\mathrm{Id}_{V_\lambda} \otimes q_\lambda(\rho)) = \dim(V_\lambda) \cdot s_\lambda(\overline{\lambda}^*)\,.$$

The distribution over partitions $\lambda \vdash [N]$ *with this probability mass function is called the* ***Schur-Weyl distribution****, denoted* $\mathrm{SW}^N(\overline{\lambda}^*)$.

(2) *Trace out the* V_λ *register, resulting in the state with unnormalized density matrix* $q_\lambda(\rho)$. *Note that the trace of this is* $s_\lambda(\overline{\lambda}^*)$ *by Corollary 84, so the normalized density matrix is*

$$\tilde{\rho} \triangleq q_\lambda(\rho)/s_\lambda(\overline{\lambda}^*)\,.$$

Intuitively, the partition λ obtained by weak Schur sampling gives us a rough estimate $\overline{\lambda} = (\lambda_1/N, \ldots, \lambda_N/N)$ of the spectrum $\overline{\lambda}^*$ of ρ. This should be thought of as the quantum analogue of the classical algorithm for learning discrete distributions: given N samples from a distribution over $[d]$ which places mass $\overline{\lambda}^*_i$ on element i, the optimal estimator for $\overline{\lambda}^*$ is to output the **empirical histogram** $\overline{\lambda} = (\lambda_1/N, \ldots, \lambda_N/N)$, where λ_i is the number of samples in the dataset that are equal to i.

When learning quantum states however, there is a crucial missing piece even after we have estimated the spectrum of ρ: estimating the *eigenvectors* of ρ. This is where we will leverage the Schur polynomial estimates from the previous section.

4. Pretty Good Measurement

With the spectrum estimate $\overline{\lambda}$ and the post-measurement state $\tilde{\rho} = q_\lambda(\rho)/s_\lambda(\overline{\lambda}^*)$ in hand, a natural approach for learning the eigenbasis for ρ would be to sample a random unitary U and "measure $\tilde{\rho}$" with the operator $U\mathrm{diag}(\overline{\lambda})U^\dagger$ in each copy in our dataset. Of course, there is a type mismatch: our dataset is no longer an element of $\mathcal{H}$, so measuring $\tilde{\rho}$ with $Q(U\mathrm{diag}(\overline{\lambda})U^\dagger)$ doesn't quite make sense. But because q_λ is an irrep inside the representation $Q(\cdot)$, we know how $Q(U\mathrm{diag}(\overline{\lambda})U^\dagger)$ acts on $\tilde{\rho}$, namely via $q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)$. So by "measuring $\tilde{\rho}$ with $U\mathrm{diag}(\overline{\lambda})U^\dagger$," we really mean measuring with the operator $q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)$. Up to a normalizing constant, this is now entirely well-defined and provides the desired rotationally *equi*variant POVM needed for the second stage of our algorithm.

Lemma 90. *The POVM with elements*

$$\frac{\dim(\mathbb{S}_\lambda V_\lambda)}{s_\lambda(\overline{\lambda})} \cdot q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)\,\mathrm{d}U \tag{24}$$

is a valid POVM.

Proof. Note that this ensemble is invariant under conjugation by any unitary: for any $W \in U_d$ we have $q_\lambda(W) \cdot q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)q_\lambda(W^\dagger) = q_\lambda(WU\mathrm{diag}(\overline{\lambda})U^\dagger W^\dagger)$, and the Haar measure over U_d is invariant under left-multiplication by W by definition. So by irreducibility of q_λ, to show that the POVM elements integrate to $\mathrm{Id}_{\mathbb{S}_\lambda V_\lambda}$, it suffices to verify that their trace integrates to $\dim(\mathbb{S}_\lambda V_\lambda)$. This follows because

$$\operatorname{tr}(q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger)) = \operatorname{tr}(q_\lambda(\mathrm{diag}(\overline{\lambda}))) = s_\lambda(\overline{\lambda})\,. \qquad \square$$

The pseudocode for our final learning algorithm is as follows:

Algorithm 1: OPTIMALTOMOGRAPHY(ρ)

Input: N copies of unknown d-dimensional state ρ
Output: Estimate $\hat{\rho}$
1 Perform weak Schur sampling on $\rho^{\otimes N}$ (Definition 89) to obtain $\lambda \vdash [N]$ and post-measurement state $\tilde{\rho}$.
2 Measure $\tilde{\rho}$ with the POVM in Eq. (24) to obtain U.
3 **return** $\hat{\rho} = U\mathrm{diag}(\overline{\lambda})U^\dagger$

With all of the machinery from Section 2, the proof that this works ends up being remarkably simple. The following shows that the further $\hat{\rho} = U\mathrm{diag}(\overline{\lambda})U^\dagger$ is from ρ, the less likely it is to output $\hat{\rho}$.

Lemma 91. *The infinitesimal probability of obtaining λ in the first step of* OPTIMALTOMOGRAPHY *and U in the second step is at most $N^{O(d^2)} \cdot F(U\mathrm{diag}(\overline{\lambda})U^\dagger, \rho)^{2N} \, \mathrm{d}U$.*

PROOF. Let us first compute the infinitesimal probability of observing U in the second stage of the algorithm, conditioned on measuring the post-measurement state $\tilde{\rho}$:

$$\frac{\dim(\mathbb{S}_\lambda V_\lambda)}{s_\lambda(\overline{\lambda}) s_\lambda(\overline{\lambda}^*)} \cdot \mathrm{tr}(q_\lambda(U\mathrm{diag}(\overline{\lambda})U^\dagger) \cdot q_\lambda(\rho)) \, \mathrm{d}U = \frac{\dim(\mathbb{S}_\lambda V_\lambda)}{s_\lambda(\overline{\lambda}) s_\lambda(\overline{\lambda}^*)} \cdot \mathrm{tr}(q_\lambda(\rho U\mathrm{diag}(\overline{\lambda})U^\dagger)) \, \mathrm{d}U \,,$$

where we used the fact that q_λ is a GL_d-representation. Recall that the probability of getting λ and $\tilde{\rho}$ from weak Schur sampling is $\dim(V_\lambda) \cdot s_\lambda(\overline{\lambda}^*)$, so the infinitesimal probability that the algorithm outputs a particular $\hat{\rho} = U\mathrm{diag}(\overline{\lambda})U^\dagger$ is

$$\dim(\mathbb{S}_\lambda V_\lambda) \dim(V_\lambda) \cdot \frac{\mathrm{tr}(q_\lambda(\rho U\mathrm{diag}(\overline{\lambda})U^\dagger))}{s_\lambda(\overline{\lambda})} \, \mathrm{d}U \,.$$

By Lemmas 77, 85, 86, and 88, this is at most

$$N^{O(d^2)} F(U\mathrm{diag}(\overline{\lambda})U^\dagger, \rho)^{2N} \, \mathrm{d}U$$

as claimed. Note the fortuitous cancellation of the entropy terms. □

We are now ready to prove the main result.

Theorem 92. *For any $\epsilon > 0$, there is $N = \tilde{O}((d^2 + \log 1/\delta)/\epsilon)$ such that given N copies of ρ,* OPTIMALTOMOGRAPHY*(ρ) outputs an estimate $\hat{\rho}$ satisfying $F(\rho, \hat{\rho}) \geq 1 - \epsilon$ with probability at least $1 - \delta$.*

PROOF. There are $\leq N^{O(d)}$ partitions $\lambda \vdash [N]$ with at most d rows, and $\int \mathrm{d}U = 1$, so the total probability contributed by (λ, U) for which $F(U\mathrm{diag}(\overline{\lambda})U^\dagger, \rho) < 1 - \epsilon$ is at most $N^{O(d^2)}(1 - \epsilon)^{2N}$. Provided $N = \Omega((d^2 \log N + \log 1/\delta)/\epsilon)$, this is upper bounded by δ as desired. □

The bound is still off by a $\log(d/\epsilon)$ factor, and it is still open whether this can be tightened to match the best known lower bound of $\Omega((d^2 + \log 1/\delta)/\epsilon)$.

CHAPTER 6

Predicting properties using classical shadows

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

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

Remarkably, there is an efficient method for this task that relies only on simple, single-copy measurements. The number of measurements required will depend not on the Hilbert space dimension d, but only on the number of properties M we wish to predict. Furthermore, the dependence on M will be very favorable. An initial version of the protocol for k-local Pauli observables called **quantum overlapping tomography** was developed in [**CW20**] and refined in [**EHF19**]. At the same time, an independent protocol called **classical shadow** for predicting observables with bounded Frobenius norm was developed in [**HK19**] based on the randomized measurement schemes in [**KRT14**, **Wri16**]. They were absorbed into the more general classical shadow formalism in [**HKP20**], which enables the prediction of a large class of observables with a more flexible protocol. Classical shadows are now widely studied and used, and is the subject of this section.

1. How to Predict Properties?

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

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

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

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

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

estimate for $\mathrm{tr}(O_i\rho)$. To estimate all M properties, we repeat this process for each observable. A pseudocode for this algorithm is given in Algorithm 2.

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

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

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

quantum measurements.

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

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

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

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

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

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

2. Classical Shadow Formalism

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

2.1. Measurement Channel and Classical Snapshot

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

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

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

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

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

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

Proof. By linearity of expectation,

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

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

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

Plugging this in gives

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

which concludes the proof. □

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

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

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

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

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

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

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

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

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

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

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

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

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

This completes the proof of (i).

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

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

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

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

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

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

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

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

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

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

(iv) The trace of the snapshot is:

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

This completes the proof of (iv). □

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

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

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

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

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

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

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

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

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

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

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

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

The first term is

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

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

2.2. Algorithm

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

2.3. Performance Guarantee

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3. Instantiations of the Random Unitary Ensemble

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

3.1. A Useful Tool: Averaging over Unitary Group

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

From the 3-design property, we also have:

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

for any traceless operators A_0, B_0.

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

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

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

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

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

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

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

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

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

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

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

This concludes the proof of the first identity.

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

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

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

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

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

- 1 identity permutation id:

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

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

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

3.2. Random Clifford Measurements

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Plugging this back into the norm expression:

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

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

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

which establishes the shadow norm bound. □

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

3.3. Random Pauli Measurements

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The shadow norm is simply 3^k:

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

This completes the proof. □

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

CHAPTER 7

Shadow Tomography via Online Learning of Quantum States

The previous lecture covered an elegant framework for shadow tomography using only very simple randomized measurements. One drawback of this approach however is that in order for the sample complexity to be not too large, the observables being estimated have to be bounded in shadow norm, e.g., properties that are local or bounded in rank.

In this lecture, we present an alternative approach to shadow tomography which achieves efficient sample complexity for *general* observables, albeit at the cost of not being computationally efficient.

Throughout we assume that the Hermitian observables O_i are positive semidefinite and satisfy $\|O_i\|_{\mathsf{op}} \leq 1$; in other words, their eigenvalues all lie in the interval $[0, 1]$. The assumption of psd-ness is without loss of generality as we can write any observable as a difference of psd operators $O^+ - O^-$ and estimate O^+ and O^- separately.

1. Online Learning

We begin by considering the following abstract setting wherein a "student" trying to learn the observable values for an unknown d-dimensional quantum state ρ interacts with a "teacher." The teacher shows the student a sequence of observables $O_1, O_2, \ldots$ to the student, where $\|O_t\|_{\mathsf{op}} \leq 1$ for all i. Every time the student gets a new O_t, she needs to form a prediction $\hat{o}_t$ for $\mathrm{tr}(O_t\rho)$, at which point the teacher will declare either

- "Pass" if $|\hat{o}_t - \mathrm{tr}(O_t\rho)| \leq \frac{3}{4}\epsilon$, or
- "Fail" if $|\hat{o}_t - \mathrm{tr}(O_t\rho)| > \epsilon$.

If the student is in the "gray zone" where $|\hat{o}_t - \mathrm{tr}(O_t\rho)| \in (\frac{3}{4}\epsilon, \epsilon]$, the teacher's response can be either "pass" or "fail."

If however the teacher ever outputs "fail," then the teacher must tell the student whether $\hat{o}_t \geq \mathrm{tr}(O_t\rho)$ or $\hat{o}_t < \mathrm{tr}(O_t\rho)$.

This problem is called **online state learning**, first studied by [**ACH+18**]. Remarkably, as we will show in this section, there is an algorithm that the student can run to limit the number of "fails" they receive to a number independent of the number of rounds of interactions with the teacher!

Theorem 109. *In quantum state learning, there is an algorithm that the student can run which ensures that she only mistakes at most $O(\log(d)/\epsilon^2)$ mistakes in total, regardless of the length of the sequence of observables that the teacher provides, and regardless of the (potentially adversarial) strategy by which the teacher selects observables to give the student.*

Algorithm 5: MatrixMultiplicativeWeights($\{M_t\}_{t=1}^T, \eta$)

Input: Sequence of "cost" matrices $M_1, \dots, M_T$, learning rate $\eta > 0$
Output: Sequence of responses $\rho_1, \dots, \rho_T$

$H_1 \leftarrow 0$;
for $t = 1, \dots, T$ **do**
  Receive cost matrix M_t;
  Respond with $\rho_t = \exp(-\eta H_t)/\operatorname{tr}\exp(-\eta H_t)$;
  $H_{t+1} \leftarrow H_t + M_t = \sum_{s=1}^t M_t$;
end

1.1. Matrix multiplicative weights

To prove Theorem 109, we consider an even more abstract setting. Suppose the student is given a sequence of matrices $M_1, M_2, \dots, M_T$, again with $\|M_t\|_{\mathsf{op}} \leq 1$ for all t, and every time the student receives some M_t, she must respond with a density matrix ρ_t. Here we will assume that for each t, either M_t or $-M_t$ is psd, so that the eigenvalues of M_t either all lie in $[0, 1]$ or all lies in $[-1, 0]$.

Over T rounds of interaction, the student's goal is to minimize the **regret**

$$\sum_{t=1}^T \operatorname{tr}(M_t \rho_t) - \min_{\rho} \sum_{t=1}^T \operatorname{tr}(M_t \rho)\,,$$

where the minimum is taken over all density matrices ρ. Intuitively, $\operatorname{tr}(M_t\rho_t)$ corresponds to a "cost" that the student incurs by predicting with ρ_t in round t, and $\sum_{t=1}^T \operatorname{tr}(M_t\rho)$ is the cost incurred by the "best-in-hindsight" strategy ρ.

This is a classic question within the classical literature on **online learning**, one of the crown jewels of which is the **matrix multiplicative weights** algorithm.

The specific update rule in Algorithm 5 may appear magical upon first glance, but it has a very simple interpretation in terms of the *maximum-entropy principle*.

Definition 110 (Entropy)**.** *Given a density matrix ρ, the* von Neumann entropy *of ρ is defined by $S(\rho) \triangleq \operatorname{tr}(-\rho \log \rho)$.*

Lemma 111. *Given a Hermitian operator H, the density matrix minimizing*

$$F(\rho) \triangleq \eta \operatorname{tr}(H\rho) - S(\rho)$$

is given by $\rho^* = \exp(-\eta H)/\operatorname{tr}\exp(-\eta H)$.

Proof. Define $Z = \operatorname{tr}\exp(-\eta H)$ and let $\rho' = \exp(-\eta H')/Z'$ for some Hermitian operator H', where $Z' \triangleq \operatorname{tr}\exp(-\eta H')$. Then the *relative entropy* $S(\rho'\|\rho^*) \triangleq \operatorname{tr}(\rho' \log \rho') - \operatorname{tr}(\rho' \log \rho^*)$ is given by

$$S(\rho'\|\rho^*) = -\langle \rho', \eta H' - \eta H + \log(Z'/Z) \cdot I\rangle = -\log(Z'/Z) + \eta\langle \rho', H - H'\rangle\,.$$

Note that

$$\begin{aligned} S(\rho') &= \langle \rho', \eta H' + (\log Z') \cdot I)\rangle \\ &= \eta \operatorname{tr}(H'\rho') + \log Z' \\ &= \eta \operatorname{tr}(H\rho') + \log Z' - \langle \rho', H - H'\rangle\,, \end{aligned}$$

so $-\log Z' + \langle \rho', H - H' \rangle = F(\rho')$. We conclude that $S(\rho' \| \rho^*) = F(\rho') - F(\rho^*)$. By (the quantum version of) Gibbs' inequality, $S(\rho' \| \rho^*) \geq 0$, so $F(\rho^*) \leq F(\rho')$ for all ρ' as claimed. □

Thus, the state ρ_t in matrix multiplicative weights is simply the state which optimally balances between minimizing correlation with all of the cost matrices seen so far, and maximizing entropy. This tradeoff is modulated by the learning rate η, which can be interpreted physically as an inverse temperature parameter. Larger η corresponds to overfitting more to previous observations and maintaining less entropy, which may make it more difficult to adapt to future cost matrices. Conversely, smaller η corresponds to adapting less to previous observations less and maintaining large entropy, which may be disadvantageous if similar cost matrices as those seen in the past also appear in the future. The following main guarantee for matrix multiplicative weights ensures that there is a way to balance between these two extremes and achieve *sublinear regret*, that is, regret which is of lower order than the length of the time horizon T.

Theorem 112. *Given a sequence of cost matrices $M_1, \ldots, M_T$, possibly adaptively chosen,* MatrixMultiplicativeWeights*$(\{O_t\}, \eta)$ produces a sequence of responses $\rho_1, \ldots, \rho_T$ such that the regret is bounded by*

$$\sum_{t=1}^{T} \operatorname{tr}(M_t \rho_t) - \min_{\rho} \sum_{t=1}^{T} \operatorname{tr}(M_t \rho) \leq \eta T + \frac{\log d}{\eta}.$$

In particular, taking $\eta = \sqrt{\log(d)/T}$ results in a regret bound of $2\sqrt{T \log d}$.

The proof is based on an elegant potential function argument.

Proof. Let $Z_t \triangleq \operatorname{tr} \exp(-H_t)$. First note that

$$Z_t \geq \exp(-\eta \sigma_{\min}(H_t)) \tag{29}$$

for all t. Let $\sum_{s=1;\succeq 0}^{t}$ (resp. $\sum_{s=1;\preceq 0}^{t}$) denote the sum over all $1 \leq s \leq t$ for which M_s (resp. $-M_s$) is psd. We will show that

$$Z_t \leq d \exp\Big(-\eta_1 \sum_{s=1;\succeq 0}^{t-1} \operatorname{tr}(M_s \rho_s) - \eta_2 \sum_{s=1;\preceq 0}^{t-1} \operatorname{tr}(M_s \rho_s)\Big), \tag{30}$$

for $\eta_1 \triangleq 1 - e^{-\eta}$ and $\eta_2 \triangleq e^{\eta} - 1$. Combining Eqs. (29) and (30) for $t = T + 1$ and taking logs on both sides and dividing by η, we conclude that

$$\begin{aligned}
\min_{\rho} \operatorname{tr}(H_T \rho) &= \sigma_{\min}(H_T) \\
&\geq \frac{\log d}{\eta} + \frac{\eta_1}{\eta} \sum_{t=1;\succeq 0}^{T} \operatorname{tr}(M_t \rho_t) - \frac{\eta_2}{\eta} \sum_{t=1;\preceq 0}^{T} \operatorname{tr}(M_t \rho_t) \\
&\geq \frac{\log d}{\eta} + (1 - \eta) \sum_{t=1;\succeq 0}^{T} \operatorname{tr}(M_t \rho_t) - (1 + \eta) \sum_{t=1;\preceq 0}^{T} \operatorname{tr}(M_t \rho_t) \\
&\geq \frac{\log d}{\eta} + \eta T + \sum_{t=1}^{T} \operatorname{tr}(M_t \rho_t),
\end{aligned}$$

where the third line follows by $\eta_1 \geq \eta(1 - \eta)$ and $\eta_2 \leq \eta(1 + \eta)$, and the fourth line follows by the fact that $|\operatorname{tr}(M_t \rho_t)| \leq 1$ for all t.

It remains to show Eq. (30). We proceed by induction on t. For the inductive step, we will use the following nice fact that says that products of matrix exponentials behave like products of scalar exponentials upon taking trace:

Fact 113 (Golden-Thompson inequality)**.** *For all Hermitian operators A and B,* $\operatorname{tr}\exp(A+B) \le \operatorname{tr}(\exp(A)\exp(B))$.

By this, we have

$$\begin{aligned} Z_{t+1} &= \operatorname{tr}\exp(-\eta\sum_{s=1}^{t} M_s) \\ &= \operatorname{tr}(\exp(-\eta\sum_{s=1}^{t-1} M_s)\cdot\exp(-\eta M_t)) \end{aligned}$$

If M_t is psd, then $\exp(-\eta M_t) \preceq \mathrm{Id} - \eta_1 M_t$, using the scalar inequality $e^{-zx} \le 1-(1-e^{-z})x$ for all $z \ge 0$ and $x \in [0,1]$, and the above is thus at most

$$\operatorname{tr}(\exp(-\eta\sum_{s=1}^{t-1} M_s)\cdot(\mathrm{Id}-\eta_1 M_t)) = Z_t(1-\eta_1\operatorname{tr}(M_t\rho_t))\,.$$

Similarly, if $-M_t$ is psd, then $\exp(-\eta M_t) \preceq \mathrm{Id} - \eta_2 M_t$, using the scalar inequality $e^{-zx} \le 1-(e^{z}-1)x$ for all $z \ge 0$ and $x \in [-1,0]$, and the above is thus instead at most

$$\operatorname{tr}(\exp(-\eta\sum_{s=1}^{t-1} M_s)\cdot(\mathrm{Id}-\eta_2 M_t)) = Z_{t-1}(1-\eta_2\operatorname{tr}(M_t\rho_t))\,.$$

As $Z_1 = \operatorname{tr}\exp(-\eta H_1) = \operatorname{tr}(\mathrm{Id}) = d$, Eq. (30) follows by induction. □

1.2. Proof of Theorem 109

Theorem 109 is now almost immediate from the regret bound for matrix multiplicative weights.

Proof. The student will run matrix multiplicative weights, maintaining an estimate of the state ρ_t, but she will only perform an update after rounds in which she "fails." In such a round t, let

$$M_t = \begin{cases} O_t & \text{if } \operatorname{tr}(O_t\rho_t) > \operatorname{tr}(O_t\rho) \\ -O_t & \text{otherwise} \end{cases}.$$

(Note that the student can determine whether $M_t = O_t$ or $M_t = -O_t$ using the teacher's feedback.) By design, $\operatorname{tr}(M_t\rho_t) - \operatorname{tr}(M_t\rho) = |\operatorname{tr}(M_t\rho_t) - \operatorname{tr}(M_t\rho)|$. Let $\hat{o}_t \triangleq \operatorname{tr}(M_t\rho_t)$.

Suppose she ends up getting "fail" at least T times. Then the total cost she incurs in those T rounds is the sum of $\operatorname{tr}(M_t\rho) + |\hat{o}_t - \operatorname{tr}(M_t\rho)|$ from those rounds, whereas the total cost she would have incurred if she had chosen ρ in every round instead of ρ_t is $\operatorname{tr}(M_t\rho)$. Her regret is at least $\frac{3}{4}\epsilon T$ because $|\hat{o}_t - \operatorname{tr}(M_t\rho)| > \frac{3}{4}\epsilon$ in each round where she fails, but by Theorem 112 her regret is also at most $2\sqrt{T\log d}$. From

$$\frac{3}{4}\epsilon T < 2\sqrt{T\log d}$$

we conclude the desired bound on the number of mistakes T. □

2. Quantum Threshold Search

To use the online learning protocol from the previous section, we need a way to implement the "teacher." Ideally, we want a teacher that can:

- Correctly identify when the student has correctly learned all of the observable values
- If the student has not correctly learned all observable values, pinpoint some observable on which the student is incorrect and show it to them in the next round of interaction.

This is formalized in the following:

Definition 114 (Quantum Threshold Search)**.** *The* quantum threshold search problem *is the following task:*

- ***Input****: access to copies of an unknown quantum state ρ; a description of observables $O_1, \dots, O_m$; and numbers $\hat{o}_1, \dots, \hat{o}_m$*
- ***Output****: Either "$|\hat{o}_i - \mathrm{tr}(O_i\rho)| \leq \epsilon$ for all $i \in [m]$," or an index $i \in [m]$ for which $|\hat{o}_i - \mathrm{tr}(O_i\rho)| > \frac{3}{4}\epsilon$ together with whether or not $\hat{o}_i \geq \mathrm{tr}(O_i\rho)$.*

We would like an algorithm that outputs an incorrect statement with probability at most δ

In this section, we give an algorithm for this task with the following guarantee:

Theorem 115. *There is an algorithm for quantum threshold search which uses $O(\frac{\log^2 m + \log 1/\delta}{\epsilon^2} \cdot \log 1/\delta)$ copies of ρ and outputs an incorrect statement with probability at most δ.*

2.1. Basic reductions

Here we perform a sequence of simplifications to the threshold search problem that show that it suffices to solve the following version of the problem:

Definition 116 (Weak Threshold Search)**.** *The* weak quantum threshold search problem *is the following task:*

- ***Input****: access to copies of an unknown quantum state ρ; and a description of* projectors $O_1, \dots, O_m$ *such that $\mathrm{tr}(O_i\rho) > 3/4$ for at least one $i \in [m]$*
- ***Output****: Any index $j \in [m]$ for which $\mathrm{tr}(O_j\rho) > 1/4$.*

We would like an algorithm that succeeds with any $\Omega(1)$ probability.

Note the four key differences: (1) the observables are projectors, (2) instead of testing whether some observable value is *far* from some threshold $\hat{o}_i$, we are testing whether it is *larger* than some threshold, (3) in place of $\frac{\epsilon}{4}$ and $\frac{3\epsilon}{4}$, the thresholds in question are fixed constants 3/4 and 1/4, (4) we are operating under the *promise* that there is some observable value which is above the threshold 3/4, and (5) we are only aiming for constant success probability.

The reductions we perform to get to this new task are relatively straightforward, and the reader may skip the proof upon first reading without losing much intuition.

Projector observables. The fact that we can assume WLOG that the observables are projectors follows immediately from the Naimark dilation theorem (Theorem 43). In fact, we could have assume this at the outset of our discussion on shadow tomography, but it wouldn't have noticeably simplified anything up to this point.

One-sided testing. The fact that we can move from testing whether $\mathrm{tr}(O_i\rho)$ is close to some value ("two-sided testing") versus testing whether it is *above* some value ("one-sided testing") arises from the fact that if we check that $\mathrm{tr}(O_i\rho)$ is above some threshold $\hat{o}_i - O(\epsilon)$, and additionally that $\mathrm{tr}((\mathrm{Id} - O_i)\rho)$ is above some threshold $1 - \hat{o}_i - O(\epsilon)$, then this is equivalent to checking that $|\mathrm{tr}(O_i\rho) - \hat{o}_i| \leq O(\epsilon)$ as in the original formulation of threshold search.

Specific constant thresholds.

With the above reasoning, we can reduce to distinguishing between whether all $\mathrm{tr}(O_i\rho)$ are below the threshold $\hat{o}_i + \frac{3}{4}\epsilon$, or whether there is some $\mathrm{tr}(O_i\rho)$ which is above the threshold $\hat{o}_i - \epsilon$. That it suffices to do this when these thresholds are replaced by $3/4$ and $1/4$ respectively is immediate from the following "boosting" result:

Lemma 117. *For any $\epsilon > 0$ and $\delta, \theta \in (0,1)$, there is an $n = O(1/\epsilon^2)$ such that for any projector Π acting on $\mathbb{C}^d$, there is a projector Π^* acting on $(\mathbb{C}^d)^{\otimes n}$ such that the following holds for any state ρ: if $\mathrm{tr}(\Pi\rho) > \theta + \frac{3}{4}\epsilon$ then $\mathrm{tr}(\Pi^*\rho^{\otimes n}) > 3/4$, and if $\mathrm{tr}(\Pi\rho) < \theta - \epsilon$ then $\mathrm{tr}(\Pi^*\rho^{\otimes n}) < 1/4$.*

Proof. For convenience let $\Pi_1 = \Pi$ and $\Pi_0 = \mathrm{Id} - \Pi$. Given $0 \leq k \leq n$, define the d^n-dimensional projector $\Pi_k^* \triangleq \sum_{x \in \{0,1\}^n : |x| = k} \bigotimes_{i=1}^n \Pi_{x_i}$, where $|x|$ denotes the Hamming weight of x. Finally define $\Pi^* \triangleq \sum_{k : k/n > \theta} \Pi_k^*$.

By Chernoff bound, if $\mathrm{tr}(\Pi\rho) > \theta + \epsilon$, then $\mathrm{tr}(\Pi^*\rho^{\otimes n}) \geq \Pr \mathrm{Bin}(n, \theta + \frac{3}{4}\epsilon) > \theta \geq 1 - \exp(-\Omega(n\epsilon^2))$, so if $n = \Omega(1/\epsilon^2)$ with sufficiently large leading constant, the latter probability is at least $3/4$. Likewise, if $\mathrm{tr}(\Pi\rho) < \theta - \epsilon$, then $\mathrm{tr}(\Pi^*\rho^{\otimes n}) \leq \Pr \mathrm{Bin}(n, \theta - \epsilon) > \theta \leq \exp(-\Omega(n\epsilon^2))$, which is bounded by $1/4$ if $n = \Omega(1/\epsilon^2)$. □

Note that the reduction in Lemma 117 blows up the dimension to d^n, but from the perspective of sample complexity this is fine as the sample complexity claimed in Theorem 115 is independent of dimension.

Promise version and constant success probability. Finally, we justify why we can assume without loss of generality that there exists $i \in [m]$ for which $\mathrm{tr}(O_i\rho) > 3/4$ in the formulation of weak threshold search, and why constant success probability suffices. Given an algorithm $\mathcal{A}$ that successfully solves weak threshold search with constant probability under this "promise," in the absence of this promise we can simply run the algorithm $\mathcal{A}$ $O(\log 1/\delta)$ times pretending the promise holds to get some indices $j_1, \ldots, j_{O(\log 1/\delta)} \in [m]$; validate whether $\mathrm{tr}(O_j\rho) > 1/4$ for any of these indices j by directly measuring the observable O_j on $\tilde{O}(\log 1/\delta)$ copies, where $\delta > 0$ is the target failure probability; and return "Success" if not. If the promise held, then the guarantee of $\mathcal{A}$ would apply and we would be done. If the promise did not hold and yet we validated that $\mathrm{tr}(O_j\rho) > 1/4$ for some j, we would still be done. Finally, if the promise did not hold and yet we returned "Success," we would also be done as $\mathrm{tr}(O_i\rho) < 3/4$ for all i by definition.

It finally remains to give an algorithm for weak threshold search (Definition 116). There are a couple known ways of doing this, and in these notes we opt for a recently proposed approach via so-called *blended measurements*, as this builds upon ideas from the first problem set [**WB24**]. The main objective will be to prove the following guarantee:

Theorem 118. *There is an algorithm for weak threshold search which uses $O(\log^2(m))$ copies of ρ and succeeds with probability $\Omega(1)$.*

2.2. Gentle and Blended Measurements

We will use a basic fact about how measurements can damage a state (this was already shown in the first problem set in the special case of pure states):

Lemma 119 (Gentle measurement lemma)**.** *If $\{M, I - M\}$ is a two-outcome measurement, and ρ' is the post-measurement state upon observing the outcome $I - M$ after measuring ρ, then $\|\rho - \rho'\|_{\mathsf{tr}} \leq 2\sqrt{\mathrm{tr}(M\rho)}$.*

Intuitively, this says that if the probability of acceptance for two-outcome measurement is small, then the post-measurement state is not too far from the original state.

Proof. It suffices to lower bound the fidelity by $\mathrm{tr}((I - M)\rho)$. This is in turn bounded by the fidelity between the purifications of ρ, ρ', noting that if $|\psi\rangle$ denotes the purification of ρ, then the purification of ρ' is given by

$$|\psi'\rangle := \frac{\sqrt{I - M} \otimes I \, |\psi\rangle}{\sqrt{\langle\psi| \, (I - M) \otimes I \, |\psi\rangle}}$$

Letting $\Lambda := (I - M) \otimes I$, we see that the fidelity is given by

$$\langle\psi| \left(\frac{\sqrt{\Lambda} \, |\psi\rangle \, \langle\psi| \sqrt{\Lambda}}{\langle\psi| \, \Lambda \, |\psi\rangle} \right) |\psi\rangle = \frac{|\langle\psi| \sqrt{\Lambda} \, |\psi\rangle|^2}{\langle\psi| \, \Lambda \, |\psi\rangle} \geq \langle\psi| \, \Lambda \, |\psi\rangle \, ,$$

where in the last step we used that $\sqrt{\Lambda} \succeq \Lambda$ because $\Lambda \preceq I$. The proof is complete upon noting that $\langle\psi| \, \Lambda \, |\psi\rangle = \mathrm{tr}((I - M)\rho)$. □

Remark 120. *The square root in the gentle measurements lemma is the key source of the **anti-Zeno effect** also explored in the first problem set: we could imagine repeatedly applying that lemma for measurements M_i such that the most recent post-measurement state ρ_{i-1} satisfies that $\mathrm{tr}(M_i \rho_{i-1})$ is small for all i, yet the total "damage" to the system as measured by the distance between the final state and the original state could be large. The gentle* sequential *measurements lemma doesn't fix this: even though the sum over acceptance probabilities is under the square root, note that those are the acceptance probabilities* with respect to the original state ρ*!*

Definition 121 (Blended measurements)**.** *Given a set of two-outcome projective measurements $M_1, \ldots, M_m$, define the* blended measurement *to be the $(m + 1)$-outcome POVM with given by $\{E_0^2, \ldots, E_m^2\}$, where*

$$E_i = \sqrt{M_i/m} \quad \text{for} \quad i = 1, \ldots, m$$

and

$$E_0 = \sqrt{I - \frac{1}{m} \sum_i M_i} \, .$$

We refer to the measurement outcome corresponding to E_0 as the "reject" outcome.

Define the state

$$\rho_{\mathsf{BM}}^{(k)} := \frac{E_0^k \rho E_0^k}{\mathrm{tr}(E_0^k \rho E_0^k)} \, ,$$

i.e., the result of applying the blended measurement k times and getting all rejects. Define the acceptance probability

$$\mathrm{Acc}_{\mathsf{BM}}(k) := 1 - \mathrm{tr}(E_0^k \rho E_0^k) \, ,$$

i.e., the probability that at least one measurement in the first k blended measurements accepts.

The gentle measurement lemma immediately implies that

$$\|\rho - \rho_{\mathsf{BM}}^{(k)}\|_{\mathsf{tr}} \leq 2\sqrt{\mathrm{Acc}_{\mathsf{BM}}(k)}\,. \tag{31}$$

2.3. Threshold Search with Blended Measurements

We now give our algorithm for weak threshold search. We begin by providing some intuition. Note that in the classical setting, what allows us to easily achieve sample complexity $O(\log(m)/\epsilon^2)$ is our ability to *reuse* samples, whereas in the quantum setting, measurement is inherently destructive and seems to preclude such reuse. Our saving grace is the gentle measurement lemma (Lemma 119), which intuitively says that measurements with very lopsided outcome probabilities are not very destructive and allow some level of data reuse.

To motivate how to leverage this, consider the following strategy: select a random observable O_i from the list and measure with the two-outcome POVM $\{O_i, I - O_i\}$ (note that this is equivalent to performing the *blended* measurement), and *post-select* on the O_i outcome. Conditioned on this, by Bayes' rule the posterior probability over getting a particular $i \in [m]$ is $\frac{\mathrm{tr}(O_i\rho)}{\sum_j \mathrm{tr}(O_j\rho)}$, which is higher for O_i such that $\mathrm{tr}(O_i\rho)$ is large. Of course, this doesn't quite work as simulating post-selection may require many samples if $\mathrm{tr}(O_i\rho)$ is small, and if we're unlucky this might be the case for all but a few, or even just one, of the O_i's. In that case, we might require $\Omega(m)$ samples just to simulate one draw from the posterior, which defeats the entire purpose of this approach.

Here is how data reuse can help: suppose after measuring the randomly chosen two-outcome POVM $\{O_i, I - O_i\}$, if we don't observe outcome O_i, we simply rerun the above experiment on the exact same copy. The hope is that if we don't observe outcome O_i, the state hasn't been damaged too much, and we can keep repeating this experiment until we get something that somewhat resembles the aforementioned posterior distribution.

This motivates the protocol in Algorithm 6 below.

Algorithm 6: BLENDEDTHRESHOLDSEARCH$(\sigma, \{M_i\})$

Input: Single copy of σ; observables $0 \preceq M_1, \ldots, M_m \preceq I$
Output: Index $i \in [m]$ or "Reject"

1 Repeatedly apply the blended measurement to σ for m times.
2 If at any point the measurement accepts, **return** the corresponding observable index $i \in [m]$
3 Otherwise, **return** "Reject"

Note that in Algorithm 6, we are using observables $M_1, \ldots, M_m$ of σ instead of observables $O_1, \ldots, O_m$ of ρ. Eventually we will take σ to be $\rho^{\otimes s}$ for $s = O(\log^2 m)$, and we will also specify how to construct $M_1, \ldots, M_m$ below.

Note that by definition, $\mathrm{Acc}_{\mathsf{BM}}(m)$ is the probability BLENDEDTHRESHOLDSEARCH outputs "Accept." Also define the quantities

$$\gamma := \frac{\sum_i \mathrm{tr}(M_i\sigma)^2}{\sum_i \mathrm{tr}(M_i\sigma)}$$

$$\gamma^* := \sum_{i=1}^{m} \sum_{j=0}^{m-1} (1 - \mathrm{Acc}_{\mathsf{BM}}(j)) \cdot \frac{\mathrm{tr}(M_i \sigma_{\mathsf{BM}}^{(j)})}{m} \cdot \mathrm{tr}(M_i \sigma)\,.$$

The interpretation is as follows:

- γ is the expected value of $\mathrm{tr}(M_i\sigma)$ under the posterior distribution from the above discussion, namely given by selecting a random M_i, measuring σ with $\{M_i, I - M_i\}$, and conditioning on the M_i outcome. This is the experiment that BlendedThresholdSearch is trying to simulate, albeit only approximately as the state gets somewhat damaged by each reuse.
- γ^* is the expected value of the observable value $\mathrm{tr}(M_i\sigma)$ where M_i is the measurement returned by BlendedThresholdSearch, where we define the observable value to be zero if the procedure does not output ACCEPT by the end.

We start by establishing a basic lower bound on the probability of returning some observable using BlendedThresholdSearch; this also motivates repeating the blended measurement up to m times.

Lemma 122. $\mathrm{Acc}_{\mathsf{BM}}(m) \geq \frac{1}{4} \max_i \mathrm{tr}(M_i\sigma)^2 \geq \gamma^2/4$.

Proof. Permute the M_i's so that WLOG, $\mathrm{tr}(M_1\sigma)^2 = \max_i \mathrm{tr}(M_i\sigma)^2$. Then $\mathrm{Acc}_{\mathsf{BM}}(m)$ is at least the sum over all m rounds of the probability that all measurements up to that round have rejected, and in that round we observe M_1, i.e.

$$\begin{aligned}
\mathrm{Acc}_{\mathsf{BM}}(m) &\geq \sum_{i=0}^{m-1} (1 - \mathrm{Acc}_{\mathsf{BM}}(i)) \cdot \frac{1}{m} \mathrm{tr}(M_1 \sigma_{\mathsf{BM}}^{(i)}) \\
&\geq \frac{1}{m} \sum_{i=0}^{m-1} (1 - \mathrm{Acc}_{\mathsf{BM}}(i)) \cdot \left(\mathrm{tr}(M_1\sigma) - \sqrt{\mathrm{Acc}_{\mathsf{BM}}(i)}\right) \\
&\geq (1 - \mathrm{Acc}_{\mathsf{BM}}(m)) \cdot (\mathrm{tr}(M_1\sigma) - \sqrt{\mathrm{Acc}_{\mathsf{BM}}(m)})\,.
\end{aligned}$$

We used the gentle measurement lemma (Eq. (31)) and the operational definition of trace distance in the second step. Rearranging the above inequality, the claim follows. □

Using the above Lemma, we can now relate γ to γ^*, showing that if $\mathrm{tr}(M_i\rho)$ has large expectation under the posterior distribution, then it has large expectation under BlendedThresholdSearch:

Lemma 123. $\gamma^* \geq \Omega(\gamma^3)$.

Proof. By the gentle measurement lemma,

$$\begin{aligned}
\gamma^* &\geq \sum_{i=1}^{m} \sum_{j=0}^{m-1} (1 - \mathrm{Acc}_{\mathsf{BM}}(j)) \cdot \frac{\mathrm{tr}(M_i\sigma)}{m} \cdot \left(\mathrm{tr}(M_i\sigma) - \sqrt{\mathrm{Acc}_{\mathsf{BM}}(j)}\right) \\
&= \sum_{j=0}^{m-1} (1 - \mathrm{Acc}_{\mathsf{BM}}(j)) \sum_{i=1}^{m} \frac{\mathrm{tr}(M_i\sigma)}{m} \left(\gamma - \sqrt{\mathrm{Acc}_{\mathsf{BM}}(j)}\right) \\
&= \sum_{j=0}^{m-1} (1 - \mathrm{Acc}_{\mathsf{BM}}(j)) \mathrm{Acc}_{\mathsf{BM}}(1) \left(\gamma - \sqrt{\mathrm{Acc}_{\mathsf{BM}}(j)}\right)
\end{aligned}$$

Recall from Lemma 122 that $\mathrm{Acc}_{\mathsf{BM}}(m) \geq \gamma^2/4$, and by monotonicity of $\mathrm{Acc}_{\mathsf{BM}}(k)$ in k, there is some $m^* \leq m$ such that $\mathrm{Acc}_{\mathsf{BM}}(k) \geq \gamma^2/4$ for all $k \geq m^*$ and $\mathrm{Acc}_{\mathsf{BM}}(k) < \frac{\gamma^2}{4}$ for all $k < m^*$. We can thus lower bound the above by

$$\gamma^* \geq m^* \left(1 - \frac{\gamma^2}{4}\right) \frac{\gamma}{2} \cdot \mathrm{Acc}_{\mathsf{BM}}(1) \geq \left(1 - \frac{\gamma^2}{4}\right) \frac{\gamma}{2} \cdot \mathrm{Acc}_{\mathsf{BM}}(m^*) \gtrsim \gamma^3 \,,$$

where in the penultimate step we again used monotonicity, specifically the fact that for every i, the probability that i is the first step where the blended measurement accepts is upper bounded by the probability that the blended measurement accepts in the first step. □

The reason we care about γ^* is that if it is large, then we expect BlendedThresholdSearch to output something with large observable value. If there were a large "gap" among the observable values, e.g. all the observables that we regard as "small" are much smaller in value (e.g. $\leq \tau$) than the ones we regard as "large", then using the fact that

$$\gamma^* \leq \tau \cdot p_b + p_g \,,$$

where p_b is the probability the protocol outputs a small observable and p_g is the probability it outputs a large observable, then we would conclude that the protocol succeeds with probability $p_g \geq \gamma^* - \tau \cdot p_b \geq \gamma^* - \tau$.

This now leads us to describe our explicit construction for the M_i's in terms of the original observables O_i. We will engineer such a gap by simply choosing a threshold and "boosting" O_i around this threshold using the same idea as Lemma 117. That is, for any threshold $\theta \in [0,1]$ and $n \in \mathbb{N}$, we can use the proof of Lemma 117 to design an n-copy observable M_i for every single-copy observable O_i such that for $\sigma \triangleq \rho^{\otimes n}$,

$$\mathrm{tr}(M_i \sigma) = \Pr[\mathrm{Bin}(n, \mathrm{tr}(O_i \rho)) \geq \theta n] \,.$$

So if we boost around a threshold $\theta \in (1/4, 3/4)$, then for i such that $\mathrm{tr}(O_i\rho) \leq 1/4$, we have $\mathrm{tr}(M_i\sigma) \leq \exp(-\Omega(n))$, so we can take τ above to be this. In other words, it is not hard to engineer the "gap" needed in the argument outlined above.

Instead, the tricky part is to ensure that we can take a threshold θ such that γ^* is sufficiently large. By the above Lemma, it suffices to show there exists a threshold θ such that the simpler quantity γ is sufficiently large. We carry this out in the next subsection.

2.4. Finding a Good Threshold

Given thresholds $0 \leq a \leq b \leq 1$, let $M[a,b]$ denote the set of indices i for which $\mathrm{tr}(O_i\rho) \in [a,b]$. Also let $n[a,b] := |M[a,b]|$.

We first show a sufficient condition for a threshold θ to yield large γ for the "boosted" observables.

Lemma 124. *For any threshold θ, the corresponding γ for the "boosted" n-copy observables satisfies*

$$\frac{1}{4\gamma} - 1 \leq \frac{1}{n[\theta,1]} \sum_{i \in M[0,\theta]} \exp(-n(\theta - \mathrm{tr}(O_i\rho))^2) \,. \tag{32}$$

PROOF. Let M_i^* denote the "boosted" n-copy observables associated to this choice of θ. Then by definition of γ, we have

$$\sum_{i\in M[\theta,1]} \operatorname{tr}(M_i^*\rho^{\otimes n})^2 \le \gamma\Big(\sum_{i\in M[0,\theta]} \operatorname{tr}(M_i^*\rho^{\otimes n}) + \sum_{i\in M[\theta,1]} \operatorname{tr}(M_i^*\rho^{\otimes n})\Big)$$

Note that for any $i \in M[\theta,1]$, the quantity $\operatorname{tr}(M_i^*\rho^{\otimes n})$ is given by $\Pr[\mathrm{Bin}(n,\theta') \ge \theta]$ for some $\theta' \ge \theta$ and thus lies in $[\frac{1}{2},1]$. Substituting this into the above and rearranging, we conclude that

$$\begin{aligned}\Big(\frac{1}{4\gamma}-1\Big)n[\theta,1] &\le \sum_{i\in M[0,\theta]} \operatorname{tr}(M_i^*\rho^{\otimes n}) \\ &\le \sum_{i\in M[0,\theta]} \exp(-n(\theta-\operatorname{tr}(O_i\rho))^2)\end{aligned}$$

as claimed. □

Henceforth, we will take $n \triangleq 100\log^2 m$. The above lemma implies that in order for the threshold to suffice for our protocol, we just need to ensure that the right-hand side of Eq. (32) is upper bounded by some constant, as this would imply γ is at least some constant.

Motivated by this, we say that a threshold is *bad* if

$$\frac{1}{n[\theta,1]} \sum_{i\in M[0,\theta]} \exp(-100\log^2 m(\theta-\operatorname{tr}(O_i\rho))^2) \ge 2\,.$$

We will show that a *random* threshold from $(1/4,3/4)$ is not bad with at least constant probability.

Lemma 125. *Suppose $n[\theta,1] \ge 1$ and that θ is bad. Then there is $\beta_\theta \le \theta$ such that*

$$n[\beta_\theta,\theta] \ge \exp(50\log^2 m(\theta-\beta_\theta)^2)\cdot n[\theta,1]\,.$$

PROOF. We will show that if, to the contrary, $n[\beta,\theta] < \exp(50\log^2 m(\theta-\beta)^2)\cdot n[\theta,1]$ for all $\beta \le \theta$, then θ is bad. Define $\eta(x) = n[\theta-x,\theta]$ and $\delta(x) \triangleq \exp(-100\log^2(m)x^2)$ so that

$$\sum_{i\in n[0,\theta]} \exp(-100\log^2 m(\theta-\operatorname{tr}(O_i\rho))^2) = \sum_{i\in n[0,\theta]} \delta(\theta-\operatorname{tr}(O_i\rho))\,.$$

Then because $-\int_z^\infty \delta'(x)\,\mathrm{d}x = \delta(z)$, we have

$$\begin{aligned}\sum_{i\in n[0,\theta]} \delta(\theta-\operatorname{tr}(O_i\rho)) &= -\int_0^\infty \eta(x)\delta'(x)\,\mathrm{d}x \\ &< \int_0^\infty 200x\log^2(m)\cdot\exp(-50\log^2(m)x^2)\,\mathrm{d}x \\ &\le 2 \le 2n[\theta,1]\,,\end{aligned}$$

where in the second step we used the assumed bound on $n[\theta-x,\theta] = \eta(x)$, so we conclude that θ is not bad. □

This lemma implies that for any bad threshold, there are exponentially many observable values below that threshold. We will use this to argue that the bad thresholds are confined to a small collection of highly concentrated "clumps," and any threshold outside of these clumps is good.

Define θ_0 to be the largest threshold within $[1/4, 3/4]$ which is bad (if no such threshold exists, we are done). By the above lemma, we know that the interval $[\beta_{\theta_0}, \theta_0]$ contains a lot of observable values. For every $i \geq 0$, let θ_{i+1} be the largest threshold within $[1/4, \beta_{\theta_i})$ which is bad, if one exists. By design, any threshold outside of the intervals $[\beta_{\theta_i}, \theta_i]$ is good, so we just need to upper bound the sum of the interval lengths $\ell_i \triangleq \theta_i - \beta_{\theta_i}$. Let n_{thres} denote the total number of θ_i's.

The following is immediate from Lemma 125:

Corollary 126. *Suppose $n[3/4, 1] \geq 1$. Then for every $j < n_{\mathsf{thres}}$,*

$$\sum_{i=0}^{j} n[\beta_{\theta_i}, \theta_i] \geq \max\left(2^j, \exp\left(50 \log^2(m) \sum_{i=0}^{j} \ell_i^2\right)\right).$$

In particular, $n_{\mathsf{thres}} \leq \log m$, and

$$\sum_{i=0}^{n_{\mathsf{thres}}-1} \ell_j^2 \leq \frac{1}{50 \log m}.$$

Proof. By Lemma 125, we have

$$n[\beta_{\theta_j}, \theta_j] \geq n[\theta_j, 1] \geq \exp(50 \log^2(m) \ell_j^2) \cdot \sum_{i=0}^{j-1} n[\beta_{\theta_{i-1}}, \theta_{i-1}],$$

so the partial sums $\sum_{i=0}^{j} n[\beta_{\theta_i}, \theta_i]$ are increasing at a rate of at least

$$1 + \exp(50 \log^2(m) \ell_j^2) \geq \max(2, \exp(50 \log^2(m) \ell_j^2)).$$

with each additional summand, as claimed. □

We can now complete the proof that the bad thresholds are concentrated in clumps whose total measure is small. This concludes our proof of Theorem 118, as it implies that we can boost the observables around a randomly chosen threshold from $[1/4, 3/4]$.

Lemma 127. *Suppose that $n[3/4, 1] \geq 1$. Then the set of bad thresholds in $[1/4, 3/4]$ has measure at most $1/6$.*

Proof. The lengths $\ell_0, \ldots, \ell_{n_{\mathsf{thres}}-1}$ are a collection of $\log m$ nonnegative numbers whose squares sum to $\frac{1}{50 \log m}$. By the fact that $\|\vec{\ell}\|_1 \leq \|\vec{\ell}\|_2 \cdot \sqrt{D}$ for any D-dimensional vector $\vec{\ell}$, we conclude that $\sum_i \ell_i \leq \sqrt{1/50} \leq 1/6$ as claimed. □

Part 3

Learning Structured States

CHAPTER 8

Learning Gibbs States: High Temperature

We have so far explored algorithms for learning completely general quantum states via quantum state tomography, as well as algorithms for efficiently learning the expectation values of many observables in a quantum state via shadow tomography and its variants. The quantum state tomography algorithms circumscribe our ability to learn about general quantum states, although even for modest system sizes the algorithms are completely impractical, and in fact fundamentally so. On the other hand, the comparatively efficient algorithms for learning quantum observables capitalize on the fact that there are often specific, structured sets of observables that are of interest to us. This relationship between structure and efficiency of learnability will continue to be a persistent theme.

In this chapter, we turn again to learning quantum states, but of a more structured kind. Specifically, we consider Gibbs states, which describe quantum systems at finite temperature. Indeed, essentially all material objects we commonly interact with are (approximately) at finite temperature, and so Gibbs states are particularly natural. Almost all condensed matter experiments operate at (approximately) finite temperature, so even in laboratory settings more removed from ordinary experience, finite temperature states are salient.

We begin with a discussion of Gibbs states and their properties, and then turn to an initial strategy for learning the parameters of a Gibbs state via quantum measurements.

1. Some history

While we do not intend to give a detailed account of thermodynamics and its quantum counterpart, we can at least marvel at its conceptual innovation and technical power. The subject of **thermodynamics**, and its cousin **statistical mechanics**, were developed over the course of the 19th century, in large part instigated to build a theory of steam engines that were essential to the industrial revolution. In this way the subjects were highly practical: there was a great need to make engines more efficient, and to understand what aspects of an engine's design were essential or extraneous.

Instead of focusing on engines, let us examine a slightly different thread of the history. In the 17th and 18th centuries, there was progress on understanding "ideal gas laws", namely how a gas' temperature, pressure, and volume are related in simple circumstances. The interrelations were empirically observed to be strikingly simple, which is surprising since we now know that a gas is a complex system of interacting particles. We should note, however, that the discoverers of the ideal gas laws did not subscribe (or at least did not fully subscribe) to the 'atomic hypothesis', and so had a rather different physical picture of gases than we now have. By the mid-19th century when the atomic hypothesis was back in vogue

and properties of gases were used industrially for designing engines, the founders of statistical mechanics articulated an interesting puzzle: if a gas is made of atoms, perhaps interacting in a complex manner, then why should gross properties (like temperature, pressure, and volume) be so simply related?

Here is where they made a surprising conceptual move. The standard practice of Newtonian physics is to write down the equations for every particle in a system and track their dynamics and interactions; this is simply too complicated to carry out for a realistically-sized gas, and especially in the 19th century. Instead of appealing to exact dynamical laws, the founders of statistical mechanics reasoned that it would instead be sensible to describe large systems in a *statistical* manner, that we will partially explicate shortly. Then, for appropriate physical observables, the microscopic details may wash out, giving accurate predictions for gross quantities. Amazingly, this works beautifully, and among myriad successes provides a first-principles derivation of the ideal gas laws.

An analogy may be helpful for the uninitiated. It is well-known, by the central limit theorem, that the statistical distribution of a sum of i.i.d. random variables converges to a Gaussian, characterized by its mean and variance; all of the other microscopic details wash away in the appropriate limit. What the founders of statistical mechanics did was figure out a type of 'central limit theorem' for *Hamiltonian dynamics*. To be sure, their arguments were not rigorous (and over the intervening century-and-a-half there has been much effort to make the arguments rigorous), but are empirically correct. That is to say, even if mathematics cannot yet verify all of their arguments in complete generality, Nature has definitively demonstrated their correctness.

The pioneering work of James Clerk Maxwell, Ludwig Boltzmann, and Josiah Willard Gibbs on statistical mechanics in the mid-to-late 19th century was ultimately synthesized into quantum mechanics in the 1930s, in large part by John von Neumann and Lev Landau. An information-theoretic articulation of statistical mechanics was later emphasized by Edwin Thompson Jaynes, based on the crucial and pioneering work of Claude Shannon in the mid-to-late 1940s. The perspective of our discussion below is most indebted to von Neumann and Jaynes.

2. Gibbs states and their properties

Before delving into the 'derivation' of a finite-temperature quantum state, it is first worthwhile to re-examine our understanding of information-theoretic entropy, due to Shannon. For a probability distribution $\vec{p} = (p_1, ..., p_n)$, its entropy is given by

$$S[\vec{p}] = -\sum_i p_i \log_2(p_i)\,,$$

where we have temporarily decided to use the base two logarithm; we will later turn back to the natural logarithm. Let us ask: what does the entropy of a probability distribution *mean*? While we have used the classical entropy, and many of its mathematical properties, so far in our analyses, we have not until now stared into its soul.

As a warm-up, suppose your friend flips a fair coin and hides it in their hand. You would surmise that the probability of heads is $\frac{1}{2}$ and the probability of tails is likewise $\frac{1}{2}$. Let us ask: how many bits of information do you expect to learn,

once the state of the coin is revealed to you? Well, in this case, with probability $\frac{1}{2}$ if heads is revealed you learn $-\log_2(\frac{1}{2}) = \log_2(2) = 1$ bit of information, and with probability 1/2 if the tail is revealed you likewise learn $-\log_2(\frac{1}{2}) = \log_2(2) = 1$ bit of information. Thus the expected number of bits you learn is 1. Indeed, the entropy of $\vec{p} = (\frac{1}{2}, \frac{1}{2})$ is $-\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1$. By a similar argument, you can convince yourself that if your friend flips n unbiased coins, then when the outcomes are revealed to you, you in expectation learn n bits of information; this is because the entropy of the uniform distribution $\vec{p} = (\frac{1}{2^n}, \frac{1}{2^n}, ..., \frac{1}{2^n})$ is n.

Let us try one more example. Suppose we have a three-outcome experiment, where the outcomes have probabilities $\vec{p} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. Then, upon learning the outcome, the expected number of bits we learn is $-\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{4}\log_2(\frac{1}{4}) - \frac{1}{4}\log_2(\frac{1}{4}) = \frac{3}{2}$, which is the entropy of $\vec{p}$. Even though we have been considering probability distributions which interface nicely with powers of 2, more general distributions have the same interpretation.

In summary, we have the interpretation:

If we sample i from $\vec{p}$, the number of bits we expect to learn when the sample is revealed to us is $S[\vec{p}]$.

If we use the natural logarithm for the entropy, where $S[\vec{p}] = -\sum_i p_i \log p_i = -\log(2)\sum_i p_i \log_2 p_i$, then the amount of information we learn is counted in 'nats', which is the natural log version of 'bits'. The entropies in different bases for the logarithm are evidently equal up to constants of proportionality.

Our ordinary-language interpretation of entropy is suggestively written: it tells us that (classical) entropy may be regarded as contingent on the observer and their knowledge of a system. For instance, if your friend knows how to toss a coin the same way every time, they can exactly predict that it will land on heads, and therefore when they see 'heads' they will learn 0 bits of information. On the other hand, if you are unaware that the toss is sure to be heads and instead surmise that the coin toss is fair, you will expect to learn 1 bit of information upon seeing the outcome. In this way, our formulation sneaks in a Bayesian viewpoint: your prior and your friend's prior may be different, and so entropy depends on the knowledge of the individual.

To this end, let us consider the following question. Suppose we have a classical system where each configuration i has an associated energy E_i, and we know that when we measure the energy at different times it is on average equal to E. If the system is composed of many interacting particles, we cannot in practice keep track of the configuration or detailed dynamics of the system; so in the spirit of statistical mechanics, let us describe the system by a probability distribution $\vec{p}$ over configurations i. That is, we say that the probability of configuration i is p_i. Then which probability distribution $\vec{p}$ is natural to choose? One desired property is that $\mathbb{E}_{i\sim\vec{p}}[E_i] = \sum_i p_i E_i = E$, since we would like the distribution to be consistent with our empirical observation that the energy of the system is E on average.[1] This property is very much not sufficient to uniquely pin down $\vec{p}$, so we need some other criteria as well. To this end, consider the following: suppose we let $\vec{p}$ be the probability distribution of *maximum entropy* with average energy E. In other words, we adopt a principle of *maximum ignorance* in which we stipulate that, if we

[1] This is actually a subtle point, which we will return to later.

were to learn the state of the system, then it would be maximally surprising to us (i.e. we would learn the maximum possible number of bits) provided that we already knew the average energy was E. Before taking a step back to decide if this is a good idea, let us pursue the stated maximization. We consider the maximization of the cost function

$$\mathsf{C}[\vec{p}, \beta, \lambda] = S[\vec{p}] + \beta \left(E - \sum_i p_i E_i \right) + \lambda \left(1 - \sum_i p_i \right) \tag{33}$$

where we enforce $\mathbb{E}_{i\sim\vec{p}}[E_i] = E$ via a Lagrange multiplier β and $\sum_i p_i = 1$ via a Lagrange multiplier λ. We will also look for maximizers for which $p_i \geq 0$ on the simplex, which we can pin down uniquely. Solving the saddle point equations $\frac{\partial \mathsf{C}}{\partial \beta} = 0$ and $\frac{\partial \mathsf{C}}{\partial \lambda} = 0$, as well as $\frac{\partial \mathsf{C}}{\partial p_i} = 0$ for all i, we find

$$p_i = \frac{e^{-\beta E_i}}{Z(\beta)}\,, \tag{34}$$

where $Z(\beta) = e^{\lambda+1} = \sum_i e^{-\beta E_i}$ is a constant such that $\sum_i p_i = 1$, and β is chosen such that $\sum_i \frac{e^{-\beta E_i}}{Z} E_i = E$. Because the Shannon entropy is a strictly concave function of $\vec{p}$ on the probability simplex, and because the constraints in (33) are linear, any stationary point of the cost function is automatically a global maximum. Moreover, the strict concavity ensures that this maximum is unique (except in degenerate cases where all E_i are equal). Provided that E lies within the feasible range $[\min_i E_i, \max_i E_i]$ and that the partition function $Z(\beta)$ converges, the **Gibbs distribution** (or Gibbs ensemble) $p_i = \frac{e^{-\beta E_i}}{Z(\beta)}$ in (34) is therefore the unique maximum-entropy distribution consistent with the given average energy constraint. The reciprocal of the parameter β, often denoted by $T := \frac{1}{\beta}$, is called the **temperature** of the system.

Crucially, the Gibbs distribution (34) does not just predict the average energy which we put in 'by hand'; additionally it makes predictions for any other observables we can measure. That is, given some observable which takes value O_i on configuration i, the Gibbs distribution makes the prediction that we will measure

$$\overline{O} := \sum_i \frac{e^{-\beta E_i}}{Z(\beta)} O_i\,.$$

Empirically, this distribution provides remarkably accurate predictions for observables O that are insensitive to microscopic details; that is, when $O_i \approx O_j$ for configurations i and j that are similar on medium or large scales. The empirical success of the Gibbs distribution is thus far from trivial: it reflects deep underlying principles of statistical mechanics and points to the existence of universality classes governing macroscopic behavior across diverse physical systems.

So why should the Gibbs distribution work so well? Let us revisit some of the ingredients in the optimization that led us to the distribution. First, suppose we call our system of interest $\mathcal{S}$, and also consider an environment $\mathcal{E}$ that couples to our system. If the system and the environment are weakly coupled and come to equilibrium, then their energies are stable on average: the average energy of our system is E and the average energy of our environment is $E_{\mathcal{E}}$, with total conserved energy $E_{\text{tot}} = E + E_{\mathcal{E}}$. Thus the energy of $\mathcal{S}$ can fluctuate about its mean E. Imagine, for instance, that $\mathcal{S}$ is described by particles in a box, and that the environment $\mathcal{E}$ is the world outside the box. Particles outside the box can hit the

box and confer energy to it; similarly, particles within the box can hit the box and confer energy to the environment. Importantly, if we characterize our knowledge of what is inside the box, it may initially have little entropy, e.g. if we knew very well the initial state of the system. But due to the chaotic dynamics inside the box, a small lack of knowledge about the system can balloon at later times into a *near-total* lack of knowledge, further facilitated by the interaction of our system with an environment of which we have little knowledge. Moreover, the chaos happens quickly, on time scales shorter than we can measure; thus our empirical average when we e.g. measure the average energy is secretly a time average as well. While it is not surprising, then, that we might *personally* have little knowledge of the system at later times (and thus would assign a large entropy if the state of the system were revealed to us), it is surprising that our *personal* ignorance suggests a distribution which is *predictive* of any coarse measurement.

To generalize the above arguments to quantum systems, we need to define a suitable notion of entropy. One criterion is if we picked a diagonal density matrix

$$\rho = \mathrm{diag}(p_1, p_2, ..., p_d)\,,$$

we would like for the quantum entropy to satisfy

$$S[\rho] = -\sum_i p_i \log(p_i)\,.$$

For a general state ρ, von Neumann stipulated the following quantum generalization of the classical entropy, which in his honor, we call the **von Neumann entropy**, or *quantum entropy* (or even merely the *entropy* when the context is clear):

$$S[\rho] := -\mathrm{tr}(\rho \log \rho)\,.$$

Here $\log \rho$ is the matrix logarithm, meaning that if $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ is the spectral decomposition of ρ, then $\log \rho = \sum_i \log(\lambda_i)|\psi_i\rangle\langle\psi_i|$. Accordingly,

$$S[\rho] = -\sum_i \lambda_i \log(\lambda_i)\,,$$

where we are taking $0 \log 0 := 0$. (Or if you want to be fussy, if ρ has zero eigenvalues let $\rho_\varepsilon := (1-\varepsilon)\,\rho + \varepsilon\,\frac{\mathbb{1}}{d}$ and take $\lim_{\varepsilon \to 0^+} S[\rho_\varepsilon]$, which will land you on the "$0 \log 0 := 0$" mnemonic.) The von Neumann entropy enjoys many nice mathematical properties, which we will ourselves enjoy later on.

With the von Neumann entropy in hand, the quantum analogue of the classical maximum-entropy construction is immediate. Let H be the Hamiltonian of a finite-dimensional quantum system, and let ρ be a density matrix. Among all density matrices ρ obeying $\mathrm{tr}(\rho H) = E$, we seek the state of maximal entropy $S[\rho] = -\mathrm{tr}(\rho \log \rho)$.

Let us introduce Lagrange multipliers $\beta, \lambda \in \mathbb{R}$ and consider

$$\mathsf{C}[\rho, \beta, \lambda] := -\mathrm{tr}(\rho \log \rho) + \beta(E - \mathrm{tr}(\rho H)) + \lambda(1 - \mathrm{tr}(\rho))\,,$$

where the second Lagrange multiplier enforces that ρ has unit trace; we will see that Hermiticity and positive semi-definiteness will be readily enforced. A first derivative

$$\frac{\delta \mathsf{C}}{\delta \rho} = -\log \rho - \mathbb{1} - \beta H - \lambda\,\mathbb{1}\,.$$

At an interior maximizer (which is full-rank whenever $E \in (E_{\min}, E_{\max})$), $\frac{\delta \mathsf{C}}{\delta \rho} = 0$, and so $\log \rho = -(\lambda+1)\,\mathbb{1} - \beta H$ which implies

$$\rho = e^{-(\lambda+1)}\, e^{-\beta H}.$$

Similar to the classical case, imposing the normalization condition $\operatorname{tr}(\rho) = 1$ identifies the partition function

$$Z(\beta) := \operatorname{tr}\big(e^{-\beta H}\big) = e^{\lambda+1},$$

and thus

$$\rho_\beta := \frac{e^{-\beta H}}{Z(\beta)}\,, \tag{35}$$

which is called a **quantum Gibbs state**. The multiplier β is then set by $\operatorname{tr}(\rho_\beta H) = E$, equivalently

$$E(\beta) = \operatorname{tr}(\rho_\beta H) = -\frac{\partial}{\partial \beta} \log Z(\beta).$$

Because $S[\rho]$ is strictly concave on the convex set of density operators and the constraints are linear, any stationary point is the global maximizer; strict concavity further implies uniqueness of ρ_β (the maximizer becomes rank-deficient only in the endpoint limits $\beta \to \pm\infty$). As before $\beta := \frac{1}{T}$ is the inverse temperature.

To see the connection with the classical Gibbs distribution, we can diagonalize the Hamiltonian as $H = \sum_a E_a\, \Pi_a$ with projectors Π_a onto energy-E_a subspaces (of dimensions $g_a := \operatorname{tr} \Pi_a$). Then

$$\rho_\beta = \frac{e^{-\beta H}}{Z(\beta)} = \sum_a \frac{e^{-\beta E_a}}{Z(\beta)}\, \Pi_a, \qquad Z(\beta) = \sum_a g_a\, e^{-\beta E_a}.$$

Thus ρ_β is block-diagonal in the energy basis and proportional to the identity on each degenerate eigenspace; in any orthonormal energy eigenbasis $\{|a,\mu\rangle\}_{\mu=1}^{g_a}$ one has diagonal entries $\langle a,\mu|\rho_\beta|a,\mu\rangle = e^{-\beta E_a}/Z(\beta)$, reproducing the classical Gibbs weights for energy eigenstates.

As in the classical case, predictions for observables follow by averaging with ρ_β:

$$\langle O \rangle_\beta := \operatorname{tr}(\rho_\beta O).$$

The quantum Gibbs state provides accurate results for a physical system when O is a coarse-grained, gross observable, such as those corresponding to energy and pressure. But unlike the classical case, the quantum Gibbs state also can provide accurate predictions when O is a local observable. The reason is subtle: in the classical setting, a system is in a definite configuration, not a probabilistic average; as such, measurements of the system only portray a probabilistic average when our measurements are coarse enough in space and time so as to blur the fast, chaotic dynamics of the system. In the quantum setting, if our system becomes entangled with its environment, then the density matrix describing our system can *genuinely* be a probabilistic mixture, i.e. a mixed state density matrix. Thus, even local observables, which would reveal the specific microscopic configuration in a classical system, are accurately described by thermal averages in the quantum case, as the system's reduced density matrix has genuinely lost information about quantum superpositions through its coupling to the environment.

All of the above said, we have ample motivation to design quantum learning algorithms for learning quantum Gibbs states. Doing so will require us to leverage

the geometric structure of physical Hamiltonians, and the particular exponential form of the quantum Gibbs state (35). We proceed with this task below.

3. A strategy for learning Gibbs states at high temperatures

We will present the quantum learning algorithm of [**HKT22**] for learning quantum Gibbs states at high temperature. For this we set up some notation and then sketch the proof before delving into details. Suppose we write our many-body Hamiltonian as

$$H = \sum_{a\in[M]} \lambda_a E_a \,,$$

where each $E_a \in \mathbb{C}^{d\times d}$ is a distinct, non-identity, traceless Hermitian operator with $\|E_a\| \leq 1$, and we take each Hamiltonian coefficient to be a real $\lambda_a \in [-1,1]$. The list of coefficients is $\lambda = (\lambda_1, ..., \lambda_M)$. We will sometimes denote the data of the Hamiltonian by (a, E_a, λ), indicating the index set $a \in [M]$, the set of operators E_a, and the list of coefficients λ. The following definition is useful.

Definition 128 (Dual interaction graph, using the notation of [**HKT22**])**.** *For any Hamiltonian in the set* $\{(a, E_a, \lambda) : a \in [M]\}$ *there is an* ***dual interaction graph*** $\mathfrak{G}$ *with vertex set* $[M]$ *where an edge connects vertices* a *and* b *if and only if* $a \neq b$ *and*

$$\mathrm{Supp}(E_a) \cap \mathrm{Supp}(E_b) \neq \emptyset \,.$$

We let $\mathfrak{d}$ *denote the maximum degree of the graph* $\mathfrak{G}$*.*

We will consider the setting in which $\mathfrak{d}$ is a constant independent of M, e.g. if each E_a acts on a constant number of qubits and each qubit participates in a constant number of terms. This covers a large class of physical Hamiltonians. Henceforth we will take, without loss of generality, each E_a to be a non-identity Pauli string acting on constant number of qubits.

The basic architecture of the proof of [**HKT22**] is as follows. We consider expectation values of each E_a in the thermal state of interest, and perform the high-temperature (i.e. small β) expansion

$$\langle E_a \rangle_\beta = \frac{\mathrm{tr}(e^{-\beta H} E_a)}{\mathrm{tr}(e^{-\beta H})} = \frac{\mathrm{tr}(E_a)}{d} + \sum_{m=1}^{\infty} \beta^m p_m^{(a)}(\lambda_1, ..., \lambda_M) \tag{36}$$

where the term $\frac{\mathrm{tr}(E_a)}{d} = 0$ since each E_a is a non-identity Pauli and thus traceless, and each $p_m^{(a)}$ is a degree m homogeneous polynomial in the Hamiltonian coefficients. Moreover, we can determine the form of any particular $p_m^{(a)}$ via an efficient classical computation. We first find a constant β_c below which the above series converges, i.e. a temperature above which our expansion makes sense. For this we find the radius of convergence of the series in the complex β-plane, which involves constraining the maximum 'sizes' of the polynomials p_m (recalling that λ_a's are at most magnitude 1) using the locality structure of the Hamiltonian and a so-called **cluster expansion**, to be explicated shortly.

Having argued that (36) makes sense above some constant temperature, the basic strategy is to argue that we can truncate the sum over m at some finite order;

then we have

$$\langle E_a \rangle_\beta \approx \sum_{m=1}^{m_{\max}} \beta^m p_m^{(a)}(\lambda_1, ..., \lambda_M) \tag{37}$$

for each a. (We will of course quantify how the approximation '$\approx$' depends on $m_{\max}$ later on.) The left-hand side is measurable, and since the p_m's on the right-hand side are efficiently classically computable, we can feasibly try to 'solve the polynomial system' given by (36) for $a \in [M]$, and obtain the coefficients $\lambda_1, ..., \lambda_M$. This last part seems tricky; for instance, there could be many spurious solutions to the equations which do not give the true Hamiltonian, or possibly many 'near'-solutions which are hard to distinguish from true solutions. Remarkably, using some nice properties of a generating function for correlation functions of the Gibbs state, we can formulate the solution of the system in (37) as a *minimization problem* which is guaranteed to be *convex* in our high-temperature regime of interest. As such, we can efficiently land on the correct λ_a's within a small approximation error.

We will segment our description of the proof into four parts accordingly. First we will explain the high-temperature cluster expansion which allows us to write (36) for all $\beta < \beta_c = O(1)$. Then we will show how solving the system given by (37) for $a \in [M]$ can be formulated as a minimization problem which is convex in our regime of interest. Next we explain a useful and efficient algorithm for solving said optimization problem. Finally we put all of the bounds together and formulate the full algorithm, in its full complexity-theoretic glory.

3.1. High-temperature cluster expansion

As advertised, we begin by justifying the series expansion in (36), and moreover in particular providing a bound on its radius of convergence β_c.

3.1.1. *Generating functions for Gibbs states*

First we need a useful mathematical object, namely the generating function of correlation functions of our Gibbs state. We write

$$\mathsf{F}(\beta, \lambda_1, ..., \lambda_M) := -\frac{1}{\beta} \log \operatorname{tr} \exp(-\beta H) = -\frac{1}{\beta} \log \operatorname{tr} \exp\left(-\beta \sum_{a \in [M]} \lambda_a E_a \right),$$

which is called the **Helmholtz free energy**, which we will call by its nickname, the 'free energy'. Note that using our notation from before $Z(\beta) = \operatorname{tr}(e^{-\beta H})$, the free energy can be written as $\mathsf{F} = -\frac{1}{\beta} \log Z(\beta)$. The free energy will serve as a generating function due to the following lemma.

Lemma 129 (The free energy is a generating function for the Gibbs state)**.** *Consider a Hamiltonian $H = \sum_{a \in [M]} \lambda_a E_a$ where the λ_a are regarded as formal variables. For non-zero $\beta \in \mathbb{C}$, we have*

$$\operatorname{tr}(E_a \rho_\beta) = \frac{\partial}{\partial \lambda_a} \mathsf{F}(\beta, \lambda_1, ..., \lambda_M)$$

for all $a \in [M]$.

PROOF. Let us write

$$\begin{aligned}
-\frac{1}{\beta}\frac{\partial}{\partial\lambda_a}\operatorname{tr}\exp(-\beta H) &= -\frac{1}{\beta}\sum_{m=0}^{\infty}\frac{1}{m!}\operatorname{tr}\left[\frac{\partial}{\partial\lambda_a}(-\beta H)^m\right] \\
&= -\frac{1}{\beta}\sum_{m=1}^{\infty}\frac{1}{m!}\sum_{k=1}^{m}\operatorname{tr}[(-\beta H)^{k-1}(-\beta E_a)(-\beta H)^{m-k}] \\
&= \sum_{m=1}^{\infty}\frac{1}{m!}\sum_{k=1}^{m}\operatorname{tr}[E_a(-\beta H)^{m-1}] \\
&= \operatorname{tr}[E_a\exp(-\beta H)]\,,
\end{aligned}$$

where we have used the linearity of the trace to move β's outside of it, and the cyclicity of the trace in going from the second line to the third line. We complete the proof by observing that

$$-\frac{1}{\beta}\frac{\partial}{\partial\lambda_a}\log\operatorname{tr}\exp(-\beta H) = -\frac{1}{\beta}\frac{1}{\operatorname{tr}\exp(-\beta H)}\frac{\partial}{\partial\lambda_a}\operatorname{tr}\exp(-\beta H) = \frac{\operatorname{tr}(E_a\exp(-\beta H))}{\operatorname{tr}\exp(-\beta H)}\,.$$

□

Thus to study series expansions of the form (36), it is natural to leverage the free energy F. One minor annoyance of the free energy is that it does not converge as $\beta\to 0$, going as $\sim 1/\beta$. This is not a problem of course, and motivates us to define the ancillary quantity

$$\mathcal{L}(\beta,\lambda_1,...,\lambda_M) := (-\beta)\,\mathsf{F}(\beta,\lambda_1,...,\lambda_M) = \log\operatorname{tr}\exp\left(-\beta\sum_{a\in[M]}\lambda_a\,E_a\right),$$

which goes to a constant as $\beta\to 0$. We will mostly use $\mathcal{L}$ henceforth.

Taking a step back, let us get a sense of what we want to prove. Consider the toy function

$$f(\beta) = \sum_{m=1}^{\infty} c_m\,\beta^m\,.$$

We would like to understand under what conditions on the c_m's is there a non-zero radius of convergence. We recall from complex analysis that the radius of convergence β_c of a function of the form $f(\beta)$ is given by

$$\frac{1}{\beta_c} = \limsup_{m\to\infty}|c_m|^{1/m}\,,$$

and so β_c is non-zero when the c_m's grow *at most* exponentially in m. In our more complicated Gibbs setting, we will show that $\max_{\lambda_1,...,\lambda_M\in[-1,1]}|p_m^{(a)}(\lambda_1,...,\lambda_M)|$ indeed grows at most exponentially in m, which will do the job. Moreover, our β_c will not depend on M and thus on the number of sites n, which is desirable since we would like our temperature bound to be system size independent. In particular, we will show that each $p_m^{(a)}$ satisfies two properties:

(1) Each $p_m^{(a)}$ is a sum of at most $e\mathfrak{d}(1+e(\mathfrak{d}-1))^m$ monomials.
(2) The coefficient in front of any monomial of $p_m^{(a)}$ has magnitude at most $(2e(\mathfrak{d}+1))^{m+1}(m+1)$.

Putting these together, we will have

$$\max_{\lambda_1,...,\lambda_M\in[-1,1]} |p_m^{(a)}(\lambda_1,...,\lambda_M)| \leq e(m+1)\mathfrak{d}(1+e(\mathfrak{d}-1))^m(2e(\mathfrak{d}+1))^{m+1}$$

which manifestly grows at most exponentially in m, and is independent of M. The key to getting M-independence will be to use the spatial locality structure of the Hamiltonian.

To clarify the structure of the series expansion of $\mathcal{L}$ and thus F, it is useful to introduce some notation. We first opt to write

$$\mathcal{L} = \log \operatorname{tr} \exp\left(-\sum_{a\in[M]} z_a E_a\right) \quad \text{where } z_a := \beta\lambda_a\,.$$

We will take $z = (z_1,...,z_M) \in \mathbb{C}^M$, i.e. considering complexified couplings for purposes of assessing convergence. By the chain rule

$$\frac{\partial\mathcal{L}}{\partial\beta} = \sum_{a\in[M]} \frac{\partial z_a}{\partial\beta}\frac{\partial\mathcal{L}}{\partial z_a} = \sum_{a\in[M]} \lambda_a \frac{\partial\mathcal{L}}{\partial z_a}$$

which gives us the multivariable series expansion

$$\begin{aligned}\mathcal{L} &= \sum_{m=0}^{\infty}\frac{\beta^m}{m!}\left(\frac{\partial^m\mathcal{L}}{\partial\beta^m}\bigg|_{\beta=0}\right)\\ &= \sum_{m=0}^{\infty}\frac{\beta^m}{m!}\sum_{a_1,a_2,...,a_m\in[M]}\lambda_{a_1}\cdots\lambda_{a_m}\left(\frac{\partial^m\mathcal{L}}{\partial z_{a_1}\cdots\partial z_{a_m}}\bigg|_{z=(0,...,0)}\right).\end{aligned} \tag{38}$$

In the last equation, for each fixed m, we have an inner sum over m variables $a_1,...,a_m \in [M]$. This, of course, is the standard structure for a multivariable series expansion; it behooves us to write this in a more compact notation so that it is more intelligible. To this end, we have the definition:

Definition 130 (Clusters of multivariate indices)**.** *A* cluster $\boldsymbol{V}$ *is a set of tuples* $\{(a,\mu(a)) : a\in[M]\}$ *where* $\mu : [M]\to\mathbb{Z}_{\geq 0}$ *counts the multiplicity of each* a*. Then the total weight* $|\boldsymbol{V}|$ *of* $\boldsymbol{V}$ *is* $\sum_a \mu(a)$*. We will write* $a\in\boldsymbol{V}$ *if* $\mu(a)\geq 1$*, and define the* support *of* $\boldsymbol{V}$ *as* $\operatorname{Supp}\boldsymbol{V} := \{a\in[M] : \mu(a)\geq 1\}$*. Finally, we define the combinatorial factor* $\boldsymbol{V}! := \prod_{a\in[M]}\mu(a)!$*.*

This is sometimes called **multi-index notation**, where $\mathbf{V}$ is the multi-index. With this notation in mind, let us rewrite (38) in a more compact manner, and define a few more pieces of notation along the way. In particular, (38) can be written as

$$\begin{aligned}\mathcal{L} &= \sum_{m=0}^{\infty}\beta^m\sum_{\mathbf{V}:|\mathbf{V}|=m}\frac{1}{\mathbf{V}!}\prod_{a\in\operatorname{Supp}\mathbf{V}}\lambda_a^{\mu(a)}\left(\prod_{b\in\operatorname{Supp}\mathbf{V}}\frac{\partial^{\mu(b)}}{\partial z_b^{\mu(b)}}\right)\bigg|_{z=(0,...,0)}\mathcal{L}\\ &= \sum_{m=0}^{\infty}\sum_{\mathbf{V}:|\mathbf{V}|=m}\frac{1}{\mathbf{V}!}\underbrace{\prod_{a\in\operatorname{Supp}\mathbf{V}}\lambda_a^{\mu(a)}}_{=:\lambda^{\mathbf{V}}}\underbrace{\left(\prod_{b\in\operatorname{Supp}\mathbf{V}}\frac{\partial^{\mu(b)}}{\partial\lambda_b^{\mu(b)}}\right)\bigg|_{\lambda=(0,...,0)}}_{=:\mathcal{D}_{\mathbf{V}}}\mathcal{L}\\ &= \sum_{m=0}^{\infty}\sum_{\mathbf{V}:|\mathbf{V}|=m}\frac{\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\end{aligned} \tag{39}$$

where we have introduced the notation $\lambda^{\mathbf{V}}$ and $\mathcal{D}_{\mathbf{V}}$. The final form (39) is compactly expressed and readily intelligible, and so was worth our efforts in notational wrangling. The form of (39) makes the source of our difficulty clearer. For $|\mathbf{V}| = m$ and any fixed $\mathbf{V}$, the product rule expansion $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ naïvely has $m!$ terms; then if we supposed that each term has size "1" (in fact, size can be larger), we would have the back-of-the-envelope estimate

$$\sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{1}{\mathbf{V}!}\, m! = M^m\,,$$

which is the number of unique length-m strings of M symbols. This type of M^m growth is exponential in m so would have a finite radius of convergence, but that radius of convergence would go as $\sim \frac{1}{M}$ which gets worse as M (or accordingly, the system size) gets larger, which we do not want. So we need to more cleverly exploit the structure of derivatives of $\mathcal{L}$ and the locality of the Hamiltonian.

To proceed, we first show $\sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!}\, \mathcal{D}_{\mathbf{V}}\mathcal{L}$ contains much fewer than M^m terms. Specifically, we show that $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ is non-zero only when $\mathbf{V}$ is *connected*, in the following sense:

Definition 131 (Connected clusters)**.** *A cluster $\boldsymbol{V} = \{(a, \mu(a)) : a \in [M]\}$ is* connected *if the subgraph of $\mathfrak{G}$ induced by the support of $\boldsymbol{V}$ is connected.*

Then, as advertised, we have the following lemma:

Lemma 132. *Recall that $Z(\beta) = \mathrm{tr}(e^{-\beta H})$. If $\boldsymbol{V}'$ and $\boldsymbol{V}''$ are nonempty and mutually disjoint and if there is no edge in $\mathfrak{G}$ connecting $\boldsymbol{V}'$ and $\boldsymbol{V}''$, then $\mathcal{D}_{\boldsymbol{V}'\cup\boldsymbol{V}''}Z = (\mathcal{D}_{\boldsymbol{V}'}Z)(\mathcal{D}_{\boldsymbol{V}''}Z)$. Thus if a cluster $\boldsymbol{V}$ is not connected, then we have $\mathcal{D}_{\boldsymbol{V}}\mathcal{L} = 0$.*

Proof. Let $H_{\mathbf{V}} := \sum_{a\in \mathrm{Supp}\mathbf{V}} \lambda_a E_a$. Then $H_{\mathbf{V}'}$ and $H_{\mathbf{V}''}$ commute since the supports of their constituent operators do not overlap. Moreover letting $Z_{\mathbf{V}} := \mathrm{tr}\exp(-\beta H_{\mathbf{V}})$, we evidently have $Z_{\mathbf{V}'\cup\mathbf{V}''} = Z_{\mathbf{V}'}Z_{\mathbf{V}''}$, and so we find

$$\mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}Z = \mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}Z_{\mathbf{V}'\cup\mathbf{V}''} = (\mathcal{D}_{\mathbf{V}'}Z_{\mathbf{V}'})(\mathcal{D}_{\mathbf{V}''}Z_{\mathbf{V}''}) = (\mathcal{D}_{\mathbf{V}'}Z)(\mathcal{D}_{\mathbf{V}''}Z)$$

as we claimed.

Letting $\mathcal{L}_{\mathbf{V}} := \log Z_{\mathbf{V}}$, we see that $\mathcal{L}_{\mathbf{V}'\cup\mathbf{V}''} = \mathcal{L}_{\mathbf{V}'} + \mathcal{L}_{\mathbf{V}''}$. Then we have

$$\mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}\mathcal{L} = \mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}\mathcal{L}_{\mathbf{V}'\cup\mathbf{V}''} = \mathcal{D}_{\mathbf{V}'\cup\mathbf{V}''}(\mathcal{L}_{\mathbf{V}'} + \mathcal{L}_{\mathbf{V}''}) = 0\,,$$

which is zero because the $\mathbf{V}''$ part of the derivative annihilates $\mathcal{L}_{\mathbf{V}'}$ and the $\mathbf{V}'$ part of the derivative annihilates $\mathcal{L}_{\mathbf{V}''}$. □

We have thus shown that a term $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ only contributes to (39) if $\mathbf{V}$ is a connected cluster. Next, it will be useful to count the number of connected clusters $\mathbf{V}$ such that $\mathbf{V}$ contains some particular vertex a, and $|\mathbf{V}| = w$, i.e. we want to count the number of clusters with weight w containing a. We do this below.

3.1.2. *Counting the number of connected clusters of fixed weight*

Recall that the dual interaction graph of our Hamiltonian is a graph $\mathfrak{G}$ of maximum degree at most $\mathfrak{d}$. For convenience, let us distinguish a 'root' vertex $a \in V(\mathfrak{G})$. We say that a cluster $\mathbf{V} = \{(a, \mu(a)) : a \in [M]\}$ is *rooted at* a if $a \in \mathrm{Supp}\mathbf{V}$.

For $k \geq 1$ let $N_{\mathfrak{G}}(a, k)$ be the number of connected vertex sets $S \subseteq V(\mathfrak{G})$ with $|S| = k$ and $a \in S$ (we will refer to such an S as a "connected support" of size k at

a). For $w \geq 1$, let $C_{\mathfrak{G}}(a, w)$ be the number of connected clusters $\mathbf{V}$ of total weight w rooted at a. Our goal will be to upper bound

$$\max_{a \in [M]} C_{\mathfrak{G}}(a, w)$$

which will give us a bound on the number of clusters with weight w containing a. For this, two elementary observations will be useful:

(1) If a support S has $|S| = k$ vertices, the number of ways to assign positive multiplicities summing to w is

$$\#\{\mu : \ \textstyle\sum_{b \in S} \mu(b) = w, \ \mu(b) \in \mathbb{Z}_{\geq 1}\} = \binom{w-1}{k-1}.$$

In particular, the multiplicity factor depends only on k (not on the geometry of S).

(2) Consequently,

$$C_{\mathfrak{G}}(a, w) = \sum_{k=1}^{w} N_{\mathfrak{G}}(a, k) \binom{w-1}{k-1}. \tag{40}$$

Hence, if we want to upper bound $C_{\mathfrak{G}}(a, w)$, it suffices to first upper bound each $N_{\mathfrak{G}}(a, k)$.

Our strategy will be to upper bound $N_{\mathfrak{G}}(a, k)$ by

$$N_{\mathfrak{G}}(a, k) \leq \max_{\mathfrak{H} \text{ with degree } \leq \mathfrak{d}} N_{\mathfrak{H}}(a, k)\,,$$

namely to maximize over all graphs $\mathfrak{H}$ with degree at most $\mathfrak{d}$. The next proposition establishes the desired maximization.

Proposition 133 (Tree maximizes rooted supports)**.** *For every $k \geq 1$ and every graph $\mathfrak{G}$ with $\Delta(\mathfrak{G}) \leq \mathfrak{d}$,*

$$N_{\mathfrak{G}}(a, k) \leq N_{T_{\mathfrak{d}}}(r, k), \tag{41}$$

where $T_{\mathfrak{d}}$ is the infinite $\mathfrak{d}$-regular tree and r is its root. Equivalently, among all degree–$\leq \mathfrak{d}$ graphs, the number of connected supports of size k rooted at a is maximized by $T_{\mathfrak{d}}$.

Proof. We construct an injective map from the family of connected supports $S \subseteq V(\mathfrak{G})$ of size k with $a \in S$ to the family of rooted subtrees of $T_{\mathfrak{d}}$ with k vertices containing r. Fix once and for all a total order on $V(\mathfrak{G})$. Given S, run breadth-first search on the induced subgraph $\mathfrak{G}[S]$ starting at a, breaking ties by the fixed order. This yields a *canonical* rooted spanning tree T_S of S. In T_S the root has at most $\mathfrak{d}$ children and each nonroot has at most $\mathfrak{d} - 1$ children.

For each vertex $v \in V(\mathfrak{G})$, fix an injective labeling $\alpha_v : \Gamma_{\mathfrak{G}}(v) \hookrightarrow \{1, 2, \ldots, \mathfrak{d}\}$ of its neighbors. Direct the edges of T_S away from the root and label each parent→child edge $u \to v$ by the *port* number $\ell(u \to v) := \alpha_u(v)$. By construction, siblings of a vertex use distinct port labels.

Now label the $\mathfrak{d}$ edges incident to every vertex of $T_{\mathfrak{d}}$ with the symbols $\{1, \ldots, \mathfrak{d}\}$. Starting at r, read the labeled rooted tree (T_S, ℓ) as instructions: from any vertex in $T_{\mathfrak{d}}$, for each child edge of T_S bearing label j, follow the unique incident edge labeled j. Distinct child labels ensure that the image is a well-defined rooted subtree of size k. Denote the resulting subtree by $\Phi(S)$.

From $\Phi(S)$ one can recover the labeled rooted tree (T_S, ℓ) (reading off port labels along edges), and then recover S itself level-by-level: the children of $u \in S$ are $\alpha_u^{-1}(\{\text{child-labels at } u\})$. Hence Φ is injective, and the claim follows. □

Remark 134. *The proposition formalizes the intuition that "unrolling cycles cannot reduce the number of rooted connected substructures" under a local degree cap; $T_{\mathfrak{d}}$ is the universal cover of any degree–$\leq \mathfrak{d}$ graph. In our cluster expansion, (40) then shows that, for fixed w and $\mathfrak{d}$, the total number of rooted clusters is maximized on $T_{\mathfrak{d}}$.*

We now provide a quantitative bound on $N_{T_{\mathfrak{d}}}(r,k)$, which is the number of rooted subtrees of $T_{\mathfrak{d}}$ with exactly k vertices. This number does not depend on the root since $T_{\mathfrak{d}}$ is self-similar, and so we write $N_{T_{\mathfrak{d}}}(r,k) = N_{T_{\mathfrak{d}}}(k)$. We have the lemma:

Lemma 135. *For $k \in \mathbb{Z}_{\geq 0}$, let $N_{T_{\mathfrak{d}}}(k)$ be the number of all connected rooted subtrees with k nodes in the infinite $\mathfrak{d}$-regular tree. Then*

$$N_{T_{\mathfrak{d}}}(k) = \binom{k(\mathfrak{d}-1)+1}{k-1} \frac{\mathfrak{d}}{k(\mathfrak{d}-1)+1} \leq e\,\mathfrak{d}\,(e(\mathfrak{d}-1))^{k-1}\,.$$

This lemma follows from some standard generating function manipulations in analytic combinatorics, which are carried out in [**HKT22**].

By the above lemma, in conjunction with (40) and (41), we have

$$\begin{aligned} C_{\mathfrak{G}}(a,w) &= \sum_{k=1}^{w} N_{\mathfrak{G}}(a,k) \binom{w-1}{k-1} \\ &\leq \sum_{k=1}^{w} N_{T_{\mathfrak{d}}}(k) \binom{w-1}{k-1} \\ &\leq \sum_{k=1}^{w} e\,\mathfrak{d}(e(\mathfrak{d}-1))^{k-1} \binom{w-1}{k-1} \\ &= e\,\mathfrak{d}(1+e(\mathfrak{d}-1))^{w-1}, \end{aligned}$$

and so we have obtained the following result:

Proposition 136. *Let $\mathfrak{G}$ be any graph with degree $\mathfrak{d} \geq 2$, and fix $a \in V(\mathfrak{G})$. For every $w \in \mathbb{Z}_{>0}$, the number of connected clusters $\boldsymbol{V}$ of total weight w rooted at a satisfies*

$$C_{\mathfrak{G}}(a,w) \leq e\,\mathfrak{d}\,\big(1+e(\mathfrak{d}-1)\big)^{w-1}.$$

In the degenerate case $\mathfrak{d} = 1$, a trivial estimate gives $C_{\mathfrak{G}}(a,w) \leq w$.

3.1.3. *Estimating the size of cluster derivatives*

Having estimated the number of (rooted) clusters with fixed weight, we now turn to bounding the size of $\frac{1}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}$ for fixed $\mathbf{V}$. For $|\mathbf{V}| = m$, we will ultimately find a bound

$$\left|\frac{1}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\right| \leq (2e(\mathfrak{d}+1)\beta)^{m+1}$$

which only depends on the degree of the graph $\mathfrak{d}$, inverse temperature β, and weight m. In order to establish this bound, we will first prove an intermediate lemma which bounds $|\mathcal{D}_{\mathbf{V}}\mathcal{L}|$ in terms of a graph constructed from the data of $\mathbf{V}$ and $\mathfrak{G}$.

To construct such an ancillary graph, consider a fixed $\mathbf{V}$. We define a graph $\mathrm{Gra}(\mathbf{V})$ from $\mathbf{V}$ as follows. The set of vertices of $\mathrm{Gra}(\mathbf{V})$ is taken to be

$$\mathrm{Mar}(\mathbf{V}) := \{(a,i) \in (\mathrm{Supp}\mathbf{V}) \times \mathbb{Z}_{>0} \,:\, 1 \leq i \leq \mu(a)\}\,,$$

where 'mar' stands for 'marked vertices'. Thus, $\mathrm{Gra}(\mathbf{V})$ has $\mu(a)$ vertices corresponding to each $a \in \mathbf{V}$, giving $|\mathbf{V}|$ vertices in total. We impose that in $\mathrm{Gra}(\mathbf{V})$, there is an edge between (a,i) and (a',i') if and only if $a = a'$ or $\mathrm{Supp}(E_a) \cap \mathrm{Supp}(E_{a'}) \neq \emptyset$ in the Hamiltonian. We have the following lemma from [**WA23**]:

Lemma 137 ([**WA23**])**.** *Letting* $\deg(v)$ *denote the number of neighbors of a vertex* $v \in \mathrm{Gra}(\boldsymbol{V})$*, we have the bound*

$$|\mathcal{D}_{\boldsymbol{V}}\mathcal{L}| \leq |\beta|^{|\boldsymbol{V}|} \prod_{v \in \mathrm{Mar}\,\boldsymbol{V}} (2\deg(v))\,.$$

This lemma follows by an elaborate graph coloring argument, which is explicated in a comprehensive manner in [**HKT22**]. For our purposes, this lemma is the main ingredient in our proposition of interest:

Proposition 138. *Let* $\boldsymbol{V}$ *be a cluster with weight* $|\boldsymbol{V}| = m+1 \geq 1$*. Then*

$$\left|\frac{1}{\boldsymbol{V}!}\mathcal{D}_{\boldsymbol{V}}\mathcal{L}\right| \leq (2e(\mathfrak{d}+1)\beta)^{m+1}\,.$$

To prove this, we will need one more elementary algebraic lemma:

Lemma 139. *Let* $\mu_1, ..., \mu_n \in \mathbb{R}_{>0}$ *and* $y_1, ..., y_n \in \mathbb{R}_{\geq 0}$*. Then*

$$\left(\frac{y_1}{\mu_1}\right)^{\mu_1} \cdots \left(\frac{y_n}{\mu_n}\right)^{\mu_n} \leq \left(\frac{y_1 + \cdots + y_n}{\mu_1 + \cdots + \mu_n}\right)^{\mu_1 + \cdots + \mu_n}\,,$$

where equality holds when $\frac{y_j}{\mu_j} = \frac{\sum_i y_i}{\sum_j \mu_j}$ *for all* j*.*

Proof. The inequality holds trivially if any $y_i = 0$, so let us assume $y_i > 0$ for all i. If we take the log of both sides of the inequality and divide by $\sum_i \mu_i$ we find

$$\sum_{i=1}^{n} \frac{\mu_i}{\sum_j \mu_j} \log\left(\frac{y_i}{\mu_i}\right) \leq \log\left(\frac{y_1 + \cdots + y_n}{\mu_1 + \cdots + \mu_n}\right),$$

which is just Jensen's inequality applied to a concave function of the logarithm. □

Now we turn to proving Proposition 138.

Proof of Proposition 138. From the definition of $\mathrm{Gra}(\mathbf{V})$ we have that for any $b \in \mathrm{Supp}\mathbf{V}$,

$$\deg((b,i)) = (\mu(b) - 1) + \sum_{a \in \Gamma(b)} \mu(a)\,, \tag{42}$$

where $\Gamma(b)$ is the set of neighbors of b in $\mathfrak{G}$ that appear in the cluster $\mathbf{V}$. Then we have the simple bound

$$\begin{aligned}\sum_{b\in\mathrm{Supp}\mathbf{V}} \deg((b,1)) &= \sum_{b\in\mathrm{Supp}\mathbf{V}} \left((\mu(b)-1) + \sum_{a\in\Gamma(b)} \mu(a) \right) \\ &\leq m + \sum_{b\in\mathrm{Supp}\mathbf{V}} \sum_{a\in\Gamma(b)} \mu(a) \\ &\leq m + \mathfrak{d}(m+1)\,, \end{aligned} \tag{43}$$

where in going from the first line to the second line we used $\sum_{b\in\mathrm{Supp}\mathbf{V}}(\mu(b)-1) \leq \left(\sum_{b\in\mathrm{Supp}\mathbf{V}} \mu(b)\right) - 1 = (m+1) - 1 = m$, and in going from the second line to the third line we used $\sum_{b\in\mathrm{Supp}\mathbf{V}} \sum_{a\in\Gamma(b)} \mu(a) = \sum_{a\in\mathbf{V}} \mu(a)\,|\{b \in \mathrm{Supp}\mathbf{V} : b \in \Gamma(a)\}| \leq \mathfrak{d} \sum_{a\in\mathbf{V}} \mu(a) = \mathfrak{d}(m+1)$. Using Lemma 137 we have

$$\begin{aligned}\frac{1}{\mathbf{V}!}|\mathcal{D}_{\mathbf{V}}\mathcal{L}| &\leq \frac{(2\beta)^{m+1}}{\mathbf{V}!} \prod_{b\in\mathrm{Supp}\mathbf{V}} \prod_{i=1}^{\mu(b)} \deg((b,i)) \\ &= (2\beta)^{m+1} \prod_{b\in\mathrm{Supp}\mathbf{V}} \frac{1}{\mu(b)!} \left(\mu(b) - 1 + \sum_{a\in\Gamma(b)} \mu(a) \right)^{\mu(b)} \\ &\leq (2e\beta)^{m+1} \prod_{b\in\mathrm{Supp}\mathbf{V}} \left(\frac{\mu(b) - 1 + \sum_{a\in\Gamma(b)} \mu(a)}{\mu(b)} \right)^{\mu(b)} \\ &\leq (2e\beta)^{m+1} \left(\frac{(1+\mathfrak{d})(m+1)}{m+1} \right)^{m+1} = (2e(\mathfrak{d}+1)\beta)^{m+1}, \end{aligned}$$

where in going from the first line to the second line we used (42), in going from the second line to the third line we used $u! \geq u^u e^{-u}$, and in going to the last line we used (43) and Lemma 139. □

3.1.4. *Bounds on the sizes of polynomials*

Proposition 138 gives us a nice bound on the size of $\frac{1}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}$. Looking back to (39), we see that this should allow us to bound the sizes of the polynomials arising in the expansion of $\mathcal{L}$. We put the pieces together below to achieve such a bound, which comes from [**HKT22**]:

Theorem 140 (High-temperature Taylor expansion and size bounds)**.** *Let $H = \sum_{a\in[M]} \lambda_a E_a$ be a Hamiltonian with known traceless Hermitian terms E_a, $\|E_a\| \leq 1$, and unknown coefficients $\lambda_a \in [-1,1]$. Let $\mathfrak{G}$ be its dual interaction graph of maximum degree $\mathfrak{d}$, and write $\rho_\beta = e^{-\beta H}/Z(\beta)$ with $Z(\beta) = \operatorname{tr} e^{-\beta H}$. Then for each $a \in [M]$ we have a (formal) β–series*

$$\langle E_a \rangle_\beta := \operatorname{tr}(E_a \rho_\beta) = \frac{\operatorname{tr}(E_a)}{d} + \sum_{m=1}^{\infty} \beta^m\, p_m^{(a)}(\lambda_1, \ldots, \lambda_M), \tag{44}$$

which holds as an identity whenever the series converges absolutely. Moreover, for every $m \in \mathbb{Z}_{>0}$ the coefficient $p_m^{(a)}$ satisfies:

(1) $p_m^{(a)} \in \mathbb{R}[\lambda_1, \dots, \lambda_M]$ *is a homogeneous polynomial of degree* m *in the Hamiltonian coefficients.*
(2) (Locality of dependence) $p_m^{(a)}$ *can involve* λ_b *only if* $\mathrm{dist}_{\mathfrak{G}}(a,b) \le m$.
(3) (Number of monomials) $p_m^{(a)}$ *contains at most* $e\,\mathfrak{d}\,\big(1+e(\mathfrak{d}-1)\big)^m$ *monomials.*
(4) (Coefficient size) *The magnitude of the coefficient in front of any monomial of* $p_m^{(a)}$ *is at most* $(2e(\mathfrak{d}+1))^{m+1}(m+1)$.

If, in addition, each E_a *is a Pauli string supported on at most* L *qubits, then after an* $O(LM\mathfrak{d}\log\mathfrak{d})$ *pre-processing (basis bookkeeping) the following algorithmic statements hold for every* $m \ge 1$*:*

(A) *The list of monomials appearing in* $p_m^{(a)}$ *can be* enumerated *in time* $O(m\,\mathfrak{d}\,C)$ *where* C *is the number of monomials; in particular in time* $O\big(m\,\mathfrak{d}^2\big(1+e(\mathfrak{d}-1)\big)^m\big)$.
(B) *The coefficient of any specific monomial can be computed* exactly *(as a rational number) in time* $O\big(Lm^3 + 8^m m^5 \log^2 m\big) = (8^m + L)\,\mathrm{poly}(m)$.

A nice consequence of the theorem is as follows. Letting

$$\tau := \big(1+e(\mathfrak{d}-1)\big)\,\big(2e(\mathfrak{d}+1)\big) \le 2e^2(\mathfrak{d}+1)^2,$$

Items 3 and 4 above imply that the series (44) converges absolutely whenever $\beta < 1/\tau$; in particular it suffices that $\beta < \beta_c = \frac{1}{2e^2(\mathfrak{d}+1)^2}$.

We will provide a proof Items (1)-(4) of Theorem 140 using the ingredients we have previously derived, and then discuss (A) and (B) which are proved in [**HKT22**].

Proof of Items (1)-(4) in Theorem 140. Recall that $\mathcal{L}(\lambda) := \log \mathrm{tr}\exp\big(-\beta\sum_b \lambda_b E_b\big)$ so that $-\frac{1}{\beta}\,\partial_{\lambda_a}\mathcal{L}(\lambda) = \mathrm{tr}(E_a\rho_\beta)$. We recall that by analyticity of $\mathcal{L}$ in a neighborhood of the origin and the multivariate Taylor formula, we have the cluster (multi-index) expansion

$$\mathcal{L}(\lambda) = \sum_{m\ge 0}\sum_{\mathbf{V}:|\mathbf{V}|=m} \frac{\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\,. \tag{45}$$

Crucially, only *connected* clusters contribute by virtue of Lemma 132. Using $\langle E_a\rangle_\beta = -\frac{1}{\beta}\,\partial_{\lambda_a}\mathcal{L}(\lambda)$ and differentiating (45) termwise on a domain of absolute convergence, we obtain

$$\langle E_a\rangle_\beta = -\frac{1}{\beta}\sum_{m\ge 0}\sum_{\substack{\mathbf{V}:|\mathbf{V}|=m+1\\ a\in\mathbf{V}}} \frac{\partial_{\lambda_a}\lambda^{\mathbf{V}}}{\mathbf{V}!}\mathcal{D}_{\mathbf{V}}\mathcal{L}\,. \tag{46}$$

Every $\mathcal{D}_{\mathbf{V}}\mathcal{L}$ carries a factor $\beta^{|\mathbf{V}|}$, so after accounting for the overall factor $1/\beta$ in (46) we can regroup terms by β^m with $m = |\mathbf{V}|-1$, arriving at (44) with

$$p_m^{(a)}(\lambda) = (-1)^{m+1}\sum_{\substack{\mathbf{V}:|\mathbf{V}|=m+1\\ a\in\mathbf{V}}} \frac{\partial_{\lambda_a}\lambda^{\mathbf{V}}}{\mathbf{V}!}\,\frac{\mathcal{D}_{\mathbf{V}}\mathcal{L}}{\beta^{m+1}}. \tag{47}$$

Since each $\partial_{\lambda_a}\lambda^{\mathbf{V}}$ is a monomial of total degree m, Item (1) follows. Moreover, because only *connected* clusters $\mathbf{V}$ contribute, any cluster counted in (47) must lie within graph distance $\le m$ of a, giving Item (2).

The number of connected clusters of total weight $w = m + 1$ that contain a is at most $e\,\mathfrak{d}\,(1 + e(\mathfrak{d} - 1))^{w-1}$ by Proposition 136, hence we have Item (3). Since Proposition 138 gives the uniform bound

$$\left| \frac{1}{\mathbf{V}!}\, \mathcal{D}_{\mathbf{V}} \mathcal{L} \right| \leq \big(2e(\mathfrak{d} + 1)\,\beta\big)^{m+1},$$

and differentiating the monomial $\lambda^{\mathbf{V}}$ contributes at most a factor $(m + 1)$ since $|\partial_{\lambda_a} \lambda^{\mathbf{V}}| \leq \mu(a) \leq m + 1$. Dividing by β^{m+1} as in (47) yields Item (4): each monomial coefficient in $p_m^{(a)}$ has size at most $(2e(\mathfrak{d} + 1))^{m+1}(m + 1)$. □

Now let us briefly discuss Items (A) and (B) in Theorem 140. For Item (A), to enumerate all contributing monomials, one enumerates connected clusters of weight m rooted at a by a breadth-first, layer-by-layer procedure (see Algorithm 1, i.e. "`tails`" in [**HKT22**]). Given random-access to neighbors in $\mathfrak{G}$, the total time is $O(m\,\mathfrak{d}\,C)$ where C is the number of clusters (hence monomials), giving Item (A).

For Item (B), to compute an individual coefficient exactly, [**HKT22**] shows how to evaluate the needed cluster derivatives $\mathcal{D}_{\mathbf{V}} \mathcal{L}$ symbolically using faithful Pauli representations, in time $O(Lm^3 + 8^m m^5 \log^2 m)$.

3.2. Finding a solution using convexity

In the previous subsection we established a high-temperature expansion for the observables $\langle E_a \rangle_\beta = \mathrm{tr}(E_a \rho_\beta)$ and proved quantitative bounds on the size and locality of the resulting polynomials in Theorem 140. We now leverage those bounds to show that $\mathcal{L}(\lambda) = \log \mathrm{tr}\Big(e^{-\beta \sum_{a \in [M]} \lambda_a E_a}\Big)$ is *locally strongly convex* in the high-temperature regime. This convexity will be the key ingredient that lets us robustly invert the map from Hamiltonian coefficients to thermal expectations, and thereby learn the coefficients.

Fix a vector $x = (x_1, \ldots, x_M) \in [-1, 1]^M$. By Theorem 140 we may write

$$\langle E_a \rangle_\beta(x) = \sum_{m=1}^{\infty} \beta^m\, p_m^{(a)}(x)\,, \quad p_m^{(a)} \text{ homogeneous of degree } m,$$

where $p_m^{(a)}$ only depends on entries x_b with $\mathrm{dist}_{\mathfrak{G}}(a, b) \leq m$, and its number and size of coefficients obey the bounds from Theorem 140 (Items (3)–(4)). In particular, letting

$$\tau := (1 + e(\mathfrak{d} - 1))\,(2e(\mathfrak{d} + 1)) \leq 2e^2(\mathfrak{d} + 1)^2 \tag{48}$$

as before, the sum of absolute coefficients of $p_m^{(a)}$ is bounded by

$$\begin{aligned} c_m &= e\,\mathfrak{d}\,(1 + e(\mathfrak{d} - 1))^m\,(2e(\mathfrak{d} + 1))^{m+1}(m + 1) \\ &= 2e^2 \mathfrak{d}(\mathfrak{d} + 1)\,\tau^{\,m}(m + 1)\,. \end{aligned} \tag{49}$$

For the learning task we will work with a *shifted, truncated* map $\mathcal{F} : [-1, 1]^M \to \mathbb{R}^M$ whose a-th component is

$$\mathcal{F}_a(x) := \sum_{m=0}^{m_{\max}} \beta^m\, p_m^{(a)}(x) = -\widehat{E}_a - \beta\, x_a + \beta^2 p_2^{(a)}(x) + \cdots + \beta^m p_{m_{\max}}^{(a)}(x)\,, \tag{50}$$

where $\widehat{E}_a$ is an estimate of $\langle E_a \rangle_\beta(\lambda)$ obtained from measurements (so we set $p_0^{(a)} := -\widehat{E}_a$), $p_1^{(a)}(x) = -x_a$ by a short computation, and $m_{\max}$ is a truncation order we

will later choose polylogarithmic in $1/(\beta\varepsilon)$. Our strategy will be to find an x such that $\mathcal{F}_a(x)$ is small for all $a \in [M]$; we will argue that if we can do then, then x is guaranteed to be closed to the true couplings λ by a convexity argument.

Here we will articulate our basic proof strategy. Let $J(x) = d\mathcal{F}(x)$ be the Jacobian of $\mathcal{F}$, namely

$$J_{ab}(x) = \frac{\partial}{\partial x_b}\mathcal{F}_a(x)\,.$$

Recall that the norm $\|\cdot\|_{\infty\to\infty}$ is defined by $\|A\|_{\infty\to\infty} = \max_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|_\infty}$. Then the idea is to use Newton iteration to find a x such that $\|x-\lambda\|_\infty \leq O(\varepsilon)$; doing so will involve bounding the size of the inverse Jacobian J^{-1} which plays an important role in Newton iteration, as well as the size of $\mathcal{F}(\lambda)$ which is the target value of $\mathcal{F}$.

To prepare for our Newton's method procedure, we will want to first establish the following facts:

(1) For suitable conditions on β and $\mathfrak{d}$, we have $\|J(x)^{-1}\|_{\infty\to\infty} \leq 2\beta^{-1}$ for all $m_{\max} \geq 1$.
(2) For any $\varepsilon > 0$, we can choose $m_{\max}$ sufficiently large (with suitable conditions on β and $\mathfrak{d}$) such that $\|\mathcal{F}(\lambda)\|_\infty \leq O(\beta\varepsilon)$.

For the first condition, we really only need the condition to hold for $m_{\max}$ sufficiently large, but in fact we will show that it holds for all $m_{\max} \geq 1$.

We will begin by establishing the first condition, and then treat the second. To this end, we have the following lemma.

Lemma 141. *Suppose that*

$$100e^6(\mathfrak{d}+1)^8\beta \leq 1\,. \tag{51}$$

Then for any $x \in [-1,1]^M$, we have $\|\mathbb{1}+\beta^{-1}J(x)\|_{\infty\to\infty} \leq \frac{1}{2}$ and $\|J(x)^{-1}\|_{\infty\to\infty} \leq 2\beta^{-1}$ for any $m_{\max} \geq 1$.

Proof. We note that if $\|\mathbb{1}+\beta^{-1}J\|_{\infty\to\infty} \leq \frac{1}{2}$, then since

$$J^{-1} = -\frac{1}{\beta}\frac{\mathbb{1}}{\mathbb{1}-(\mathbb{1}-\beta^{-1}J)} = -\frac{1}{\beta}\sum_{k=0}^{\infty}(\mathbb{1}+\beta^{-1}J)^k\,,$$

we would have

$$\|J^{-1}\|_{\infty\to\infty} \leq \beta^{-1}\sum_{k=0}^{\infty}\|\mathbb{1}+\beta^{-1}J\|_{\infty\to\infty}^k \leq 2\beta^{-1}\,.$$

Thus it suffices to show $\|\mathbb{1}+\beta^{-1}J(x)\|_{\infty\to\infty} \leq \frac{1}{2}$, or equivalently $\|\beta\mathbb{1}+J(x)\|_{\infty\to\infty} \leq \frac{\beta}{2}$, for our stated domain of β.

We observe that the Jacobian takes the form

$$J_{ab} = -\beta\,\delta_{ab} + O(\beta^2)\,,$$

and so $\beta\mathbb{1} + J = O(\beta^2)$. As such, we would like to bound the $O(\beta^2)$ remainder. Let $u = (u_1, ..., u_M)$ satisfy $|u_b| \leq 1$ for all b, i.e. $\|u\|_\infty \leq 1$. Then we have

$$\begin{aligned}((J+\beta\mathbb{1})u)_a &= \sum_b (J+\beta\mathbb{1})_{ab} u_b \\ &= \sum_b u_b \left(\beta^2 \frac{\partial}{\partial x_b} p_2^{(a)}(x) + \cdots + \beta^{m_{\max}} \frac{\partial}{\partial x_b} p_{m_{\max}}^{(a)}(x) \right) \\ &= \sum_{k=2}^{m_{\max}} \beta^k \sum_{b\,:\,\mathrm{dist}_{\mathfrak{G}}(a,b)\leq k} u_b \frac{\partial}{\partial x_b} p_k^{(a)}(x)\,,\end{aligned}$$

where in going to the last line we have used Item (2) of Theorem 140. We observe that in the last sum, at each fixed k, the index b ranges over at most $1+\mathfrak{d}+\cdots+\mathfrak{d}^k \leq (\mathfrak{d}+1)^k$ vertices of $\mathfrak{G}$. Now recall that each $p_k^{(a)}$ is a homogeneous polynomial of degree k, and that the sum of the absolute values of the coefficients is bounded by c_k in (49). Therefore $\left|\frac{\partial}{\partial x_b} p_k^{(a)}\right| \leq k c_k$ in the domain of $\mathcal{F}$, and we have

$$\begin{aligned}|((J+\beta\mathbb{1})u)_a| &\leq \sum_{k=2}^{\infty} \beta^k (\mathfrak{d}+1)^k k c_k \\ &\leq 2e^2(\mathfrak{d}+1)^2 (\beta(\mathfrak{d}+1)\tau)^2 \sum_{k=2}^{\infty} (\beta(\mathfrak{d}+1)\tau)^{k-2} k(k+1) \\ &= 2e^2(\mathfrak{d}+1)^4\beta^2\tau^2 \left(\frac{6-6r+2r^2}{(1-r^3)}\bigg|_{r=\beta(\mathfrak{d}+1)\tau} \right) \\ &\leq \frac{25}{2} e^2(\mathfrak{d}+1)^4\beta^2\tau^2\,.\end{aligned}$$

In going from the second line to the third line we used $\beta(\mathfrak{d}+1)\tau < 1$, and in going to the last line we used $\beta(\mathfrak{d}+1)\tau \leq \frac{1}{100}$. Since our u satisfying $\|u\|_\infty \leq 1$ was arbitrary, we have obtained the bound $\|J+\beta\mathbb{1}\|_{\infty\to\infty} \leq \frac{25}{2} e^2(\mathfrak{d}+1)^4\beta^2\tau^2$. Using $\tau \leq 2e^2(\mathfrak{d}+1)^2$ from (48) and $100e^6(\mathfrak{d}+1)^8\beta \leq 1$ from (51), we find our desired bound $\|J+\beta\mathbb{1}\|_{\infty\to\infty} \leq \frac{\beta}{2}$. □

A nice consequence of the above lemma is the following convexity result:

Lemma 142. *If (51) holds, then $\nabla^{\otimes 2}\mathcal{L} \succeq \frac{\beta^2}{2}\mathbb{1}$, namely $\mathcal{L}$ is $(\frac{\beta^2}{2})$-strongly convex.*

Proof. Take $m_{\max} = \infty$ so that $\nabla^{\otimes 2}\mathcal{L} = -\beta J$, where we note that the Jacobian J is Hermitian. For a Hermitian matrix X, we have $\|X\| \leq \|X\|_{\infty\to\infty}$, and so $\|\mathbb{1} + \beta^{-1}J\| \leq \|\mathbb{1} + \beta^{-1}J\|_{\infty\to\infty} \leq \frac{1}{2}$, implying that $\mathbb{1} + \beta^{-1}J \preceq \mathbb{1}/2$ and thus $\beta^{-1}J \preceq -\mathbb{1}/2$, which is equivalent to $\nabla^{\otimes 2}\mathcal{L} \succeq \frac{\beta^2}{2}\mathbb{1}$. □

Next let us show that if $m_{\max}$ is chosen to scale at least logarithmically in $1/(\beta\varepsilon)$, then we can arrange for $\|\mathcal{F}(\lambda)\|_\infty \leq O(\beta\varepsilon)$. First we require the following lemma.

Lemma 143 (Estimating thermal expectations in parallel)**.** *For any $\varepsilon, \delta \in (0,1)$ there is a measurement procedure that (given independent copies of ρ_β) produces estimators $\widehat{E}_a$ such that*

$$\left|\widehat{E}_a - \langle E_a\rangle_\beta\right| \leq \beta\,\varepsilon \qquad \text{for all } a \in [M]$$

simultaneously with probability at least $1-\delta$*, using*

$$O\left(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right)$$

copies of ρ_β *and with time complexity*

$$O\left(\frac{N\,\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right).$$

Proof. We first recall a standard fact: given a quantum state ρ and a Hermitian observable E with $\|E\|\le 1$, one can estimate $\operatorname{tr}(E\rho)$ to additive error ε_0 with success probability at least $1-\delta_0$ using $O\big(\log(1/\delta_0)/\varepsilon_0^2\big)$ independent copies of ρ. Indeed, measuring ρ in the eigenbasis of E yields an i.i.d. random variable in $[-1,1]$ whose expectation is $\operatorname{tr}(E\rho)$; Hoeffding bounds then give the stated sample complexity.

We now apply this in parallel to the family $\{E_a\}_{a\in[M]}$. Color the vertices of the dual interaction graph $\mathfrak{G}$ using at most $\mathfrak{d}+1$ colors (a greedy coloring suffices). By definition of $\mathfrak{G}$, all E_a belonging to a fixed color class act on disjoint sets of qubits. Consequently, on a single copy of ρ_β we can measure *all* E_a in that color class simultaneously: since each E_a is a Pauli string, it suffices to measure each qubit once in the appropriate single-qubit Pauli basis and multiply outcomes to obtain the eigenvalue of each E_a in the class.

Fix a color class and set the target accuracy per observable to $\varepsilon_0 := \beta\varepsilon$. By the single-observable estimate and a union bound over all a in the class, $O\big(\log(1/\delta_0)/\varepsilon_0^2\big)$ copies of ρ_β suffice to ensure that every $\widehat{E}_a$ in that class satisfies $|\widehat{E}_a-\langle E_a\rangle_\beta|\le\varepsilon_0$ with probability at least $1-\delta_0$. Repeating independently for each of the at most $\mathfrak{d}+1$ color classes, the total number of copies is

$$(\mathfrak{d}+1)\,O\Big(\frac{\log(1/\delta_0)}{\varepsilon_0^2}\Big)=O\left(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{1}{\delta_0}\right).$$

Choosing $\delta_0 := \delta/M$ and applying a union bound across all M observables yields simultaneous accuracy $|\widehat{E}_a-\langle E_a\rangle_\beta|\le\beta\varepsilon$ for every $a\in[M]$ with probability at least $1-\delta$, and the stated copy complexity $O\big(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\log\frac{M}{\delta}\big)$ follows.

For the time complexity, note that each copy used in a given color round requires at most N single-qubit Pauli measurements (one per qubit), and there are $(\mathfrak{d}+1)\,O(\log(1/\delta_0)/\varepsilon_0^2)$ such copies overall. This gives time

$$O\left(N\,\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right),$$

as claimed. □

This lemma tells us that we can set $|\widehat{E}_a-\langle E_a\rangle_\beta|\le O(\beta\varepsilon)$ for all a. With this in mind, we have the following.

Lemma 144. *Assume the high-temperature condition* (51)*. Let* τ *be as in* (48) *and set* $r:=\beta\tau$*. Suppose the empirical means obey* $\big|\widehat{E}_a-\langle E_a\rangle_\beta\big|\le\beta\varepsilon$ *for all* $a\in[M]$*. If the truncation order* $m_{\max}$ *in* (50) *satisfies*

$$(2r)^{\,m_{\max}+1}\le\frac{\beta\varepsilon}{4\,e^2\,\mathfrak{d}\,(\mathfrak{d}+1)}\,, \tag{52}$$

then $\|\mathcal{F}(\lambda)\|_\infty \le 2\beta\varepsilon$. *Equivalently, it suffices to take*

$$m_{\max} \ge \left\lceil \frac{\log\left(\dfrac{4\,e^2\,\mathfrak{d}\,(\mathfrak{d}+1)}{\beta\varepsilon}\right)}{\log\left(\dfrac{1}{2\beta\tau}\right)} \right\rceil - 1\,. \tag{53}$$

In particular, for constant $\mathfrak{d}$ *we have* $m_{\max} = O\big(\log(1/(\beta\varepsilon))\big)$.

Proof. By the definition (50) and the triangle inequality,

$$\big|\mathcal{F}_a(\lambda)\big| \le \big|\widehat{E}_a - \langle E_a\rangle_\beta\big| + \sum_{m>m_{\max}} \beta^m \,\big|p_m^{(a)}(\lambda)\big| \le \beta\varepsilon + \sum_{m>m_{\max}} \beta^m\, c_m\,.$$

Thus, with $r = \beta\tau$,

$$\sum_{m>m_{\max}} \beta^m\, c_m = 2e^2\,\mathfrak{d}(\mathfrak{d}+1) \sum_{m>m_{\max}} (m+1)\, r^{\,m}.$$

For $m \ge 1$ we may use $(m+1) \le 2^m$, where (since $2r < 1$)

$$\sum_{m>m_{\max}} (m+1)\, r^m \le \sum_{m>m_{\max}} (2r)^m = \frac{(2r)^{m_{\max}+1}}{1-2r}.$$

The high-temperature hypothesis (51) implies $r \ll 1$ (and hence $2r < 1$); in particular, $1/(1-2r) \le 2$. Therefore

$$\sum_{m>m_{\max}} \beta^m\, c_m \le 4e^2\,\mathfrak{d}(\mathfrak{d}+1)\,(2r)^{\,m_{\max}+1}.$$

Imposing (52) makes the right-hand side at most $\beta\varepsilon$, and hence $\big|\mathcal{F}_a(\lambda)\big| \le 2\beta\varepsilon$ for all a. Taking the maximum over a yields $\|\mathcal{F}(\lambda)\|_\infty \le 2\beta\varepsilon$.

Finally, solving (52) for $m_{\max}$ gives (53); since $2\beta\tau < 1$ under (51), the denominator is a positive constant when $\mathfrak{d}$ is constant, proving the claimed $O(\log(1/(\beta\varepsilon)))$ scaling. □

Finally, we show that we can efficiently find an x such that $\|x-\lambda\|_\infty \le 18\,\varepsilon$.

Theorem 145 (High-temperature learning via projected Newton-Raphson)**.** *Assume the high-temperature condition* (51). *Suppose we are given estimates* $\{\widehat{E}_a\}_{a\in[M]}$ *of the thermal expectations* $\langle E_a\rangle_\beta$ *obeying* $\big|\widehat{E}_a - \langle E_a\rangle_\beta\big| \le \beta\,\varepsilon$ *for all* $a \in [M]$. *Moreover let us take* $\varepsilon \le \frac{1}{12}$. *Then there is a classical algorithm (a projected Newton-Raphson scheme with a truncated Neumann-series inverse) that outputs* $x \in [-1,1]^M$ *such that* $\|x-\lambda\|_\infty \le 18\,\varepsilon$ *in time* $O(\frac{ML}{\varepsilon}\mathrm{poly}(\mathfrak{d}, \log\frac{1}{\beta\varepsilon}))$, *where* L *is the maximum number of qubits on which any Hamiltonian term acts.*

Proof Sketch. Let us choose the judicious bound

$$m_{\max} \ge \left\lceil \frac{e}{e-1}\,\frac{1}{\log\left(\frac{1}{\beta\tau}\right)}\,\log\left(\frac{12e^2(\mathfrak{d}+1)^2}{\beta\varepsilon\log\left(\frac{1}{\beta\tau}\right)}\right)\right\rceil$$

which is compatible with our previous one. The **Newton-Raphson method** ordinarily entails an iteration like $x^{(t+1)} = x^{(t)} - (J^{-1}\mathcal{F})(x^{(t)})$, although to avoid computing the inverse of J we will instead consider an approximation $J(x)^{-1} \approx$

$\beta^{-1}\sum_{k=0}^{K-1}(\mathbb{1}+\beta^{-1}J(x))^k$ for a sufficiently large K that we will specify. Specifically, we consider the iteration

$$x^{(0)}=\vec{0}\,,\quad x^{(t+1)}=\mathrm{Proj}_{[-1,1]^M}\left[x^{(t)}+\beta^{-1}\sum_{k=1}^{K}(\mathbb{1}+\beta^{-1}J(x^{(t)}))^k\mathcal{F}(x^{(t)})\right]$$

where we have used

$$\mathrm{Proj}(u):=\begin{cases}1 & \text{if } u\in(1,\infty)\\ u & \text{if } u\in[-1,1]\\ -1 & \text{if } u\in(-\infty,1)\end{cases},$$

and take $K=\lceil\log_2\left(\frac{3}{2\varepsilon}\right)\rceil$.

Before analyzing the convergence of the iterations, let us examine the error $e^{(t)}$ between $J(x^{(t)})^{-1}\mathcal{F}(x^{(t)})$ and $\beta^{-1}\sum_{k=0}^{K-1}(\mathbb{1}+\beta^{-1}J(x^{(t)}))^k\mathcal{F}(x^{(t)})$. Specifically, we have

$$\begin{aligned}e^{(t)}&:=\left(-J(x^{(t)})^{-1}+\frac{1}{\beta}\sum_{k=0}^{K-1}(\mathbb{1}+\beta^{-1}J(x^{(t)}))^k\right)\mathcal{F}(x^{(t)})\\&=-\frac{1}{\beta}\sum_{k=K}^{\infty}(\mathbb{1}+\beta^{-1}J(x^{(t)})^k\mathcal{F}(x^{(t)})\\&=J^{-1}(x^{(t)})(\mathbb{1}+\beta^{-1}J(x^{(t)}))^K\mathcal{F}(x^{(t)})\,,\end{aligned}$$

which by Lemma 141 decays exponentially in K in the $\|\cdot\|_\infty$ norm. This will be useful for us shortly.

With the error $e^{(t)}$ under control, let us examine the convergence of $x^{(t)}$ under our Newton-Raphson iteration. Let $\mathcal{F}_a(s):[0,1]\to\mathbb{R}$ by defined by $\mathcal{F}_a(s):=\mathcal{F}_a(x+s(\lambda-x))$. By Taylor's remainder theorem, there exists an $s'\in[0,1]$ such that

$$\underbrace{\mathcal{F}_a(1)}_{=\mathcal{F}_a(\lambda)}=\underbrace{\mathcal{F}_a(0)}_{=\mathcal{F}_a(x)}+(\partial_s\mathcal{F}_a)(0)+\frac{1}{2}(\partial_s^2\mathcal{F}_a)(s')\,.$$

Using $\partial_s=\sum_b(\lambda_b-x_b)\partial_b$ and setting $y^{(a)}:=s'\lambda+(1-s')x$, we find

$$\mathcal{F}_a(\lambda)=\mathcal{F}_a(x)+\sum_b(\lambda_b-x_b)\underbrace{(\partial_b\mathcal{F}_a)(x)}_{=J_{ab}(x)}+\frac{1}{2}\sum_{b,c}(\lambda_b-x_b)(\lambda_c-x_c)(\partial_b\partial_c\mathcal{F}_a)(y^{(a)})\,.$$

Letting $\Delta^{(t)} := x^{(t)} - \lambda$ (and similarly $\Delta^{(t+1)} := x^{(t+1)} - \lambda$) where $\Delta_d^{(t)}$ denotes the dth coordinate, we have the following:

$$\begin{aligned}
|\Delta_d^{(t+1)}| &= \left|\mathrm{Proj}_{[-1,1]}[(x - (J^{-1}\mathcal{F})(x) + e)_d] - \lambda_d\right| \\
&\leq \left|(x - (J^{-1}\mathcal{F})(x) + e)_d - \lambda_d\right| \\
&= \left|e_d^{(t)} + \Delta_d^{(t)} - \sum_a (J(x^{(t)})^{-1})_{da}\mathcal{F}_a(x^{(t)})\right| \\
&= \left|e_d^{(t)} + \Delta_d^{(t)} - \sum_a J(x^{(t)})^{-1}_{da}\left(\mathcal{F}_a(\lambda) - \sum_b (\lambda_b - x_b^{(t)})J_{ab}(x^{(t)})\right.\right. \\
&\qquad\qquad \left.\left. - \frac{1}{2}\sum_{b,c}(\lambda_b - x_b^{(t)})(\lambda_c - x_c^{(t)})[\partial_b\partial_c\mathcal{F}_a](y^{(a)})\right)\right| \\
&= \left|\left[e^{(t)} + \Delta^{(t)} - J(x^{(t)})^{-1}\mathcal{F}(\lambda) - J(x^{(t)})^{-1}J(x^{(t)})\Delta^{(t)}\right]_d\right. \\
&\qquad\qquad \left. + \frac{1}{2}\sum_{a,b,c} J(x^{(t)})^{-1}_{da}\Delta_b^{(t)}\Delta_c^{(t)}[\partial_b\partial_c\mathcal{F}](y^{(a)})\right| \\
&= \left|\left[J(x^{(t)})^{-1}\left((\mathbb{1} + \beta^{-1}J(x^{(t)}))^K\mathcal{F}(x^{(t)}) - \mathcal{F}(\lambda)\right)\right]_d\right. \\
&\qquad\qquad \left. + \frac{1}{2}\sum_{a,b,c} J(x^{(t)})^{-1}_{da}\Delta_b^{(t)}\Delta_c^{(t)}[\partial_b\partial_c\mathcal{F}_a](y^{(a)})\right| . \qquad (54)
\end{aligned}$$

We will bound each term in the last equation in turn. For the first part, we have

$$\begin{aligned}
&\left|\left[J(x^{(t)})^{-1}\left((\mathbb{1} + \beta^{-1}J(x^{(t)}))^K\mathcal{F}(x^{(t)}) - \mathcal{F}(\lambda)\right)\right]_d\right| \\
&\leq \|J(x^{(t)})^{-1}\|_{\infty\to\infty}\left(\|\mathbb{1} + \beta^{-1}J(x^{(t)})\|^K_{\infty\to\infty}\|\mathcal{F}(x^{(t)})\|_\infty + \|\mathcal{F}(\lambda)\|_\infty\right) \\
&\leq 2\beta^{-1}\left(2^{-K}(2 + \beta\varepsilon) + 2\beta\varepsilon\right) \leq 6\varepsilon\,. \qquad (55)
\end{aligned}$$

In going to the last line we have used Lemma 141 and Lemma 144, in conjunction with

$$\begin{aligned}
|\mathcal{F}_a(x)| &\leq \left|\widehat{E}_a + \sum_{k=1}^{m_{\max}}\beta^k|p_k^{(a)}(x)|\right| \\
&\leq |\widehat{E}_a - \langle E_a\rangle_\beta| + \left|-\langle E_a\rangle_\beta + \sum_{k=1}^{m_{\max}}\beta^k|p_k^{(a)}(x)|\right| \\
&\leq \beta\varepsilon + 2\,.
\end{aligned}$$

For the last term in (54), we have for all indices d the inequalities

$$\begin{aligned}
&\left|\frac{1}{2}\sum_{a,b,c} J(x^{(t)})^{-1}_{da}\Delta_b^{(t)}\Delta_c^{(t)}[\partial_b\partial_c\mathcal{F}](y^{(a)})\right| \\
&\le \frac{1}{2}\|J(x^{(t)})^{-1}\|_{\infty\to\infty}\max_a\left|\sum_{b,c}\Delta_b^{(t)}\Delta_c^{(t)}[\partial_b\partial_c\mathcal{F}_a](y^{(a)})\right| \\
&\le \frac{1}{\beta}\max_a\sum_{k=0}^{\infty}\sum_{b,c}|\Delta_b^{(t)}\Delta_c^{(t)}|\,\beta^k\,|\partial_b\partial_c p_k^{(a)}(y)| \\
&\le \frac{1}{\beta}\sum_{k=0}^{\infty}\sum_{\substack{b,c:\\ \mathrm{dist}_{\mathfrak{G}}(b,a)\le k\\ \mathrm{dist}_{\mathfrak{G}}(c,a)\le k}}\|\Delta^{(t)}\|_\infty^2\,\beta^k k(k-1)c_k \\
&\le \frac{1}{\beta}\sum_{k=0}^{\infty}(\mathfrak{d}+1)^{2k}\|\Delta^{(t)}\|_\infty^2\,\beta^k k(k-1)c_k \\
&= \frac{12e^2}{\beta}\|\Delta^{(t)}\|_\infty^2(\mathfrak{d}+1)^2\frac{(\beta(\mathfrak{d}+1)^2\tau)^2}{(1-\beta(\mathfrak{d}+1)^2\tau)^4} \\
&\le \frac{25}{2}\,e^2\beta(\mathfrak{d}+1)^6\tau^2\|\Delta^{(t)}\|_\infty^2\,, \qquad (56)
\end{aligned}$$

where in going to the second-to-last line we have used that $\beta(\mathfrak{d}+1)^2\tau<1$ and in going to the last line we have used that $\beta\mathfrak{d}^2\tau\le 1-\left(\frac{24}{25}\right)^{1/4}$. Putting together (55) and (56) we find

$$\|\Delta^{(t+1)}\|_\infty \le 6\varepsilon + \frac{25}{2}\,e^2\beta(\mathfrak{d}+1)^6\tau^2\|\Delta^{(t)}\|_\infty^2\,.$$

By solving the recursion, one can show that so long as $\|\Delta^{(0)}\|_\infty \le \frac{1}{25e^2\beta(\mathfrak{d}+1)^6\tau^2}\le 1$, we achieve $\|x^{(T)}-\lambda\|_\infty\le 18\varepsilon$ for

$$T=\lceil -\log_2(300e^6(\mathfrak{d}+1)^{10}\beta\varepsilon)\rceil\,.$$

Finally, let us sketch the runtime bound. For each $a\in[M]$, the truncated series $\mathcal{F}_a(x)=\sum_{k=0}^m\beta^k p_k^{(a)}(x)-\widehat{E}_a$ is a degree-m polynomial whose support is contained in the radius-k neighborhoods of a in $\mathfrak{G}$; the number of contributing terms at order k is at most $\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^k$ and each coefficient can be evaluated in time $O(L\,\mathrm{poly}(k))$. Hence, evaluating all M coordinates of $F(x)$ and forming (or applying) the nonzeros of the sparse Jacobian $J(x)$ at a given point x costs

$$O\Big(M\,L\,\mathrm{poly}(\mathfrak{d})\sum_{k=0}^{m}(\mathfrak{d}+1)^k\Big)=O\big(M\,L\,\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^{O(m)}\big).$$

One Newton step uses the truncated Neumann-series inverse $\beta^{-1}\sum_{k=0}^{K-1}\big(\mathbb{1}+\beta^{-1}J(x)\big)^k$, which requires K sparse matrix–vector multiplies with $J(x)$, and thus has cost $O\big(K\,M\,L\,\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^{O(m)}\big)$ at iteration $x=x^{(t)}$. The projection $\mathrm{Proj}_{[-1,1]^M}$ adds only $O(M)$ time. With T Newton iterations in total, the overall runtime is

$$O\Big((K+1)\,T M\,L\,\mathrm{poly}(\mathfrak{d})\,(\mathfrak{d}+1)^{O(m)}\Big).$$

Substituting in our parameter choices yields the stated time complexity:

$$O\left(\frac{M\,L}{\varepsilon}\mathrm{poly}\left(\mathfrak{d},\log\frac{1}{\beta\varepsilon}\right)\right).$$

□

To summarize, we have succeeded in establishing that, for suitable β, $\mathfrak{d}$, and $m_{\max}$, we have $\|x-\lambda\|_\infty \le O(\varepsilon)$. Below we will put together all of our results thus far to get the final, overarching algorithm and associated bounds.

3.3. Putting the bounds together

We can combine all of the results above to get the main result of [**HKT22**]. We recapitulate some of the notation we have collected along the way.

Theorem 146 (Learning from high-temperature Gibbs states)**.** *Let $H=\sum_{a\in[M]}\lambda_a E_a$ be a* low-intersection *Hamiltonian on N qubits: each non-identity Pauli term E_a acts on at most $L=O(1)$ qubits and the dual interaction graph has maximum degree $\mathfrak{d}=O(1)$. Fix inverse temperature $\beta>0$ obeying the high-temperature condition (51) (equivalently $\beta<\beta_c(\mathfrak{d})$ for a universal constant $\beta_c>0$ depending only on $\mathfrak{d}$), and let $\rho_\beta=e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$.*

For any $\varepsilon\in(0,\frac{1}{12})$ and failure probability $\delta\in(0,1)$, there is a classical algorithm which, given independent copies of ρ_β, outputs $\widehat{\lambda}\in[-1,1]^M$ satisfying

$$\|\widehat{\lambda}-\lambda\|_\infty \le 18\varepsilon$$

with probability at least $1-\delta$, using

$$S_\infty = O\left(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right)$$

copies of ρ_β. In particular, when $\mathfrak{d}=O(1)$ and $M=\Theta(N)$, this is

$$S_\infty = O\left(\frac{\log N}{\beta^2\varepsilon^2}\right).$$

Consequently, to achieve ℓ_2-error $\|\widehat{\lambda}-\lambda\|_2\le\varepsilon$ it suffices to use

$$S_2 = O\left(\frac{M}{\beta^2\varepsilon^2}\,\log\frac{M}{\delta}\right) = O\left(\frac{N}{\beta^2\varepsilon^2}\,\log\frac{N}{\delta}\right).$$

The total running time is linear in the sample size (i.e. $O(SN)$ where S is the number of copies used), up to polylogarithmic factors in $1/(\beta\varepsilon)$.

Proof. Assume (51) and let τ be as in (48). We can estimate all thermal expectations in parallel (via Lemma 143) to obtain $\{\widehat{E}_a\}_{a\in[M]}$ with $|\widehat{E}_a-\langle E_a\rangle_\beta|\le\beta\varepsilon$ for every a using $S_\infty=O\big(\frac{\mathfrak{d}}{\beta^2\varepsilon^2}\log\frac{M}{\delta}\big)$ copies of ρ_β, with success probability $\ge 1-\delta$.

Define $\mathcal{F}$ as in (50) and choose the truncation order $m_{\max}$ as in (53). Then Lemma 144 gives $\|\mathcal{F}(\lambda)\|_\infty \le 2\beta\varepsilon$. By the high-temperature conditioning in Lemma 141, we have

$$\|\mathbb{1}+\beta^{-1}J(x)\|_{\infty\to\infty}\le\tfrac{1}{2}\quad\text{and}\quad\|J(x)^{-1}\|_{\infty\to\infty}\le 2\beta^{-1},\quad\text{for all } x\in[-1,1]^M.$$

We run the projected Newton–Raphson update with truncated Neumann inverse in Theorem 145 from $x^{(0)}=\vec{0}$ and with $K=\lceil\log_2\big(\frac{3}{2\varepsilon}\big)\rceil$. The one-step analysis yields the recursion $\|\Delta^{(t+1)}\|_\infty\le 6\varepsilon+C\,\beta\,\|\Delta^{(t)}\|_\infty^2$ with $C=\frac{25}{2}e^2(\mathfrak{d}+1)^6\tau^2$.

Solving this recursion with $T = \lceil -\log_2 \left(300 e^6 (\mathfrak{d}+1)^{10} \beta\varepsilon\right) \rceil$ gives $\|x^{(T)} - \lambda\|_\infty \le 18\varepsilon$. We set $\widehat{\lambda} := x^{(T)}$ to obtain the claimed accuracy with probability $\ge 1-\delta$.

The sample bound is exactly that of Lemma 143, and for $\mathfrak{d} = O(1)$ and $M = \Theta(N)$ it simplifies to $S_\infty = O\big(\frac{\log N}{\beta^2\varepsilon^2}\big)$. The ℓ_2 statement follows by targeting $\|\widehat{\lambda} - \lambda\|_\infty \le \varepsilon/\sqrt{M}$, which replaces ε by $\varepsilon/\sqrt{M}$ in Lemma 143, yielding $S_2 = O\big(\frac{M}{\beta^2\varepsilon^2} \log \frac{M}{\delta}\big)$. The runtime is $O(S_\infty N)$ for data collection plus $O\big(\frac{ML}{\varepsilon}\,\mathrm{poly}(\mathfrak{d}, \log \frac{1}{\beta\varepsilon})\big)$ for classical postprocessing, which is linear in the sample size up to polylogarithmic factors in $1/(\beta\varepsilon)$. $\square$

CHAPTER 9

Learning Gibbs States: Low Temperature

In the previous chapter, we saw an algorithm for learning Gibbs states at temperatures above some absolute constant depending on the geometry of the Hamiltonian. Although the algorithm we considered breaks down at lower temperatures, there is no *a priori* reason why there shouldn't exist any algorithm that succeeds in that regime. Indeed, one can show that at least *information-theoretically*, one can learn at arbitrary temperatures with sample complexity scaling exponentially in β and inversely in $\mathrm{poly}(\beta)$ [**AAKS21**]. For a while, it was an open question whether one could achieve this rate with a *computationally* efficient algorithm. This was resolved in a recent breakthrough of [**BLMT24**] which gave an algorithm with runtime and sample complexity $\mathrm{poly}(m, (1/\epsilon)^{2^\beta})$, where m is the number of terms in the Hamiltonian; this doubly exponential dependence on β was subsequently improved to singly exponential dependence by [**Nar24**]. These papers rely on a powerful algorithmic framework called *sum-of-squares programming*; unfortunately, a complete exposition of this approach would take us too far afield, and instead we will consider a different algorithm due to the very recent work of [**CAN25**]. This last paper gave an alternative algorithm with better dependence on the system size, and using an arguably more intuitive approach.

1. Technical Preliminaries

Throughout, fix a Hamiltonian

$$H = \sum_a \lambda_a P_a = \sum_\eta \eta \Pi_\eta \,,$$

where η ranges over the distinct eigenvalues ("energies") of H, and $\Pi_\eta = |\eta\rangle \langle\eta| \, \eta$ denotes projection to the eigenspace corresponding to eigenvalue η. Throughout, $\rho \propto e^{-\beta H}$ will denote its Gibbs state.

Any such Gibbs state naturally induces the following inner product which generalizes the classical L_2 inner product with respect to a probability measure:

Definition 147 (KMS inner product)**.** *Given operators A, B and a density matrix ρ, their **KMS inner product** is given by*

$$\langle A, B \rangle_\rho = \mathrm{tr}(A \rho^{1/2} B^\dagger \rho^{1/2}) \,.$$

*This induces the **KMS norm** $\|A\|_\rho^2 \triangleq \langle A, B \rangle_\rho$.*

We will often be interested in *differences* between energies $\eta - \eta'$:

Definition 148 (Bohr frequencies and Bohr decomposition)**.** *The set of **Bohr frequencies**, denoted $B(H)$, of a Hamiltonian H consist of all differences $\eta - \eta'$ between eigenvalues of H. We will always use the letter ν, possibly with superscripts,*

to denote Bohr frequencies, and $\sum_\nu$ *to denote* $\sum_{\nu \in B(H)}$ *when* H *is clear from context.*

Any operator A *can naturally be decomposed into blocks* $\Pi_{\eta'} A \Pi_\eta$ *corresponding to different pairs of eigenspaces, and the Bohr frequencies give a natural set of "bands" for grouping together these blocks. Given operator* A *and Bohr frequency* ν*, define*

$$A_\nu = \sum_{\eta' - \eta = \nu} \Pi_{\eta'} A \Pi_\eta \,.$$

Counterintuitively, the algorithm we will describe for learning Gibbs states, which are inherently *static* objects, will arise from reasoning about *dynamics* associated to the Hamiltonian. These were introduced briefly in Section 3. We will discuss their relevance to the learning algorithm and analysis in Section 2, but for now we simply recall their definition:

Definition 149 (Time evolution)**.** *Given a time* $t \in \mathbb{R}$ *and a Hermitian* H*, the associated* ***time evolution*** *operator is* e^{-iHt}*.*[1] *Under the* ***Schrödinger picture****, a state* ρ_0 *undergoing time evolution becomes* $\rho_t = e^{-iHt} \rho_0 e^{iHt}$ *at time* t*. Dually, one can consider the time evolution of* observables*. Under the* ***Heisenberg picture****, an observable* A_0 *undergoing time evolution becomes* $A_t = e^{iHt} A_0 e^{-iHt}$ *at time* t*. We will adopt some shorthand for the latter: given an operator* A*, define*

$$A_H(t) \triangleq e^{iHt} A e^{-iHt} \,.$$

While time evolution is defined with $t \in \mathbb{R}$ *(indeed, this is essential for* e^{iHt} *to be unitary), we can also consider conjugating operators by the Gibbs state instead of by* e^{iHt}*, which would correspond to* imaginary t*, to get*

$$e^{-\beta H} A e^{\beta H} \,.$$

We will refer to this as ***imaginary time evolution*** *and occasionally abuse notation by writing this as* $A_H(i\beta)$*.*

The Bohr decomposition behaves nicely under time evolution and imaginary time evolution:

Lemma 150. *For any* $t \in \mathbb{C}$*,*

$$A_H(t) = \sum_\nu e^{i\nu t} A_\nu \,.$$

Proof. By definition $e^{iHt} = \sum_\eta e^{i\eta t} \Pi_\eta$ has the same eigenvectors as H, so $\Pi_\eta e^{-iHt} = e^{-i\eta t} \Pi_\eta$ and $\Pi_{\eta'} e^{iHt} = e^{i\eta' t}$. Therefore, for any η, η' for which $\eta' - \eta = \nu$, we have

$$e^{iHt} \Pi_{\eta'} A \Pi_\eta e^{-iHt} = e^{i\nu t} \Pi_{\eta'} A \Pi_\eta \,,$$

from which the claim follows by linearity. □

[1] Here we have flipped the sign from what was defined in Section 3 as it is more convenient for some of the subsequent calculations.

2. Learning by Exploiting Detailed Balance

The starting point for the proof is the **Kubo-Martin-Schwinger (KMS) condition**, which can be thought of as a quantum analogue of detailed balance.

Theorem 151 (KMS condition)**.** *Let $\rho' = e^{-\beta H'}/tr(e^{-\beta H'})$ for some Hamiltonian H'. The equation*

$$\mathrm{tr}(\rho' A_H(t) O) = \mathrm{tr}(\rho' O A_H(t + i\beta)) \tag{57}$$

holds for all operators O and A and all times $t \in \mathbb{C}$[2] if and only if $H' = H + c\mathrm{Id}$ for some absolute constant $c \in \mathbb{R}$.

While this statement is incredibly powerful, as we will see, its proof is rather trivial.

Proof. One can readily verify that for $\rho = e^{-\beta H}/\mathrm{tr}(e^{-\beta H})$,

$$\begin{aligned}
\mathrm{tr}(\rho A_H(t) O) &= \frac{1}{Z}\mathrm{tr}(e^{iH(t+i\beta)} A e^{-iHt} O) \\
&= \frac{1}{Z}\mathrm{tr}(e^{iH(t+i\beta)} A e^{-iH(t+i\beta)} e^{-\beta H} O) \\
&= \mathrm{tr}(A_H(t + i\beta)\rho O) \\
&= \mathrm{tr}(\rho O A_H(t + i\beta)) .
\end{aligned}$$

To show the converse, note that Eq. (57) holding for all O, A, t is equivalent to the condition that

$$\rho' A_H(t) = A_H(t + i\beta)\rho' = \rho A_H(t)\rho^{-1}\rho'$$

for all A, t. This in particular implies that $A = (\rho^{-1}\rho')^{-1} A (\rho^{-1}\rho')$ for all A, which implies that the mixed states ρ and ρ' are equal, as desired. □

In other words, the only state that satisfies the KMS condition with respect to H is the Gibbs state. The rest of this section is about making this insight quantitative in order to extract out a learning algorithm. This requires answering two questions. First, if $\rho' \propto e^{-\beta H'}$ is *close*, in an appropriate sense, to satisfying the KMS condition, does that imply H' is close to H? Second, how does one computationally efficiently find an H' for which this is the case?

2.1. A KMS alternative that sees locality

The above argument that the only state that satisfies the KMS condition with respect to H is the Gibbs state is unfortunately rather *global* in nature as it involves multiplying by inverses of matrix exponentials. This is a horribly ill-conditioned operation as many of the eigenvalues of ρ are exponentially small. In this section, we will outline an approach to making this argument more "local" by carefully designing O and A.

First, let us slightly shift perspectives by switching the roles of H, ρ and H', ρ' in Theorem 151. For convenience, let us also replace t by $t - i\beta/2$ just so that instead of $(t, t + i\beta)$, we get $(t - i\beta/2, t + i\beta/2)$. This gives rise to the following equivalent formulation of the KMS condition:

[2]The KMS condition was originally devised as an alternative characterization of Gibbs states that can extend to *infinite dimensions*; in those contexts one has to be a bit careful about issues of analyticity and t is taken to be real, but in finite dimensions this is not an issue.

Theorem 152 (Dual KMS)**.** *Given a Hamiltonian H', operators O and A, and time $t \in \mathbb{C}$, define the observable*

$$\Delta[\![H'; O, A, t]\!] \coloneqq OA_{H'}(t + i\beta/2) - A_{H'}(t - i\beta/2)O \tag{58}$$

measuring the extent to which some equation in the KMS condition is violated. Then

$$\operatorname{tr}(\rho\Delta[\![H'; O, A, t]\!]) = 0$$

for all O, A, t if and only if $H = H' + c\mathrm{Id}$ for some absolute constant c.

If we had some way of enumerating over all H', O, A, t, then we might imagine using our copies of the Gibbs state ρ to estimate all of these observable values and hope that if we have found some H' for which the corresponding observable values $\operatorname{tr}(\rho\Delta^*_{H';O,A,t})$ were all very small, then $H' \approx H$.

Thus far, we haven't done anything new on top of Theorem 151, but we are now in a position to try making things more local. The first key insight towards doing so is to realize that there is a certain weighted combination of the observables $\Delta[\![H'; O, A, t]\!]$'s for *local* operators O and A which is small if and only if H and H' are close (up to additive shift).

Theorem 153 (Identifiability equation)**.** *For any operators O and A, we have*

$$\frac{\beta}{2}\langle O, [A, H - H']\rangle_\rho = \operatorname{tr}(\rho\overline{\Delta}[\![H'; O, A]\!])\,,$$

for

$$\overline{\Delta}[\![H'; O, A]\!] \coloneqq \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \Delta[\![H'; O_H^\dagger(t), A, t]\!]\, g_\beta(t)\,\mathrm{d}t$$

where $g_\beta(t) \coloneqq \frac{2}{\beta}g(2t/\beta)$ for

$$g(t) \triangleq -\frac{\pi^{3/2}}{2\sqrt{2}(1 + \cosh(\pi t))}$$

(see Figure 1 for a plot – the particular functional form is not important, but the fact that it is rapidly decaying is).

We will prove this in Section 2.2. Although this result is stated in terms of general operators O and A, the following result morally tells us that it suffices to consider 1-local Paulis A and $O = [A, H - H']$ and motivates why we consider $\langle O, [A, H - H']\rangle_\rho$ in the first place:

Lemma 154. *Suppose $H = \sum_a \lambda_a P_a$ and $H' = \sum_a \lambda_a E'_a$ are Hamiltonians with the same set of couplings. If $\frac{1}{2^n}\|[A, H - H']\|_F^2 \le \epsilon^2$ for all three Pauli operators $A \in \{X_i, Y_i, Z_i\}$ acting solely on the i-th qubit, then $|\lambda_a - \lambda'_a| \le \epsilon$ for every term P_a acting on qubit i.*

Proof. For any A, we have

$$\frac{1}{2^n}\|[A, H - H']\|_F^2 = \frac{1}{2^n}\operatorname{tr}([A, [A, H - H']](H - H'))\,.$$

If $A \in \{X_i, Y_i, Z_i\}$ and P is some Pauli operator, then $[A, P] = 0$ if P acts as A or Id on the i-th qubit, and otherwise $[A, P]$ is the operator which is identical to P off of the i-th qubit and equal to some signed Pauli on the i-th qubit. Moreover, in this case $[A, [A, P]] = 4P$.

We conclude that $\sum_{A\in\{X_i,Y_i,Z_i\}}[A,[A,H-H']] = 8\sum_{a\sim i}(\lambda_a-\lambda'_a)P_a$, where the sum is over terms a which act on i, and thus

$$3\epsilon^2 \geq \frac{1}{2^n}\sum_{A\in\{X_i,Y_i,Z_i\}}\mathrm{tr}([A,[A,H-H']](H-H')) = 8\sum_{a\sim i}(\lambda_a-\lambda'_a)^2\,,$$

from which the claim follows. □

Remark 155. *There is a slight but nontrivial catch that Lemma 154 pertains to the* Frobenius *norm of* $[A,H-H']$*, whereas Theorem 153 involves the* KMS *norm. For local operators like* $[A,H-H']$*, these can be related up to a* $e^{\mathrm{poly}(\beta)}$ *factor using ideas from Section 3 below. Proving this would take us too far afield, and we defer the interested reader to* [**CAN25**, Lemma III.6].

Modulo this remark, combining Theorem 153 and Lemma 154, we conclude that if H' was such that $\mathrm{tr}(\rho\overline{\Delta}[\![H';O,A]\!])$ was small for all 1-local Paulis A and $O=[A,H-H']$, then this would ensure that H and H' are equivalent. As we saw in the proof of Lemma 154, $[A,H-H']$ only consists of terms a which act on the i-th qubit, and for local Hamiltonians this is a constant number of terms. Furthermore, the operator $O_H^\dagger(t)$ that appears in the definition of $\overline{\Delta}[\![H';O,A]\!]$ is also approximately local, because intuitively the locality of H ensures that the time-evolved operator $O_H^\dagger(t)$ doesn't "spread out" too much in a short amount of time - this is the content of **Lieb-Robinson bounds**, which we discuss in Section 4.

There are however two important challenges remaining to "localizing" the KMS condition into something that can be algorithmically useful. First, the observables $\overline{\Delta}[\![H';O,A]\!]$ require knowledge of H, at the very least in order to write down $O_H^\dagger(t)$. Second, recall from the definition of Δ in Eq. (58) that they still involve the scary-looking imaginary-time-evolved operators $A_{H'}(t\pm i\beta/2)$. Imaginary time evolution involves conjugating by a fractional power of $e^{-\beta H'}$, which again might have exponentially small eigenvalues. So it would appear that we still haven't sidestepped the need to invert by ill-conditioned matrices.

The former issue is not so bad: even without knowing H, we can simply enumerate over guesses of the Hamiltonian in a way that we make precise in Section 4. The latter is the more fundamental issue, and we deal with this in Section 3 using a subtle **regularization trick** from the literature on quantum Gibbs sampling.

2.2. Proof of identifiability equation

In this section we prove Theorem 153. The key technical tools will be a *nested* Bohr decomposition with respect to the Bohr frequencies of H and H'.

Given Hamiltonians H_1 and H_2 with Bohr frequencies $B(H_1)=\{\nu_1\}$ and $B(H_2)=\{\nu_2\}$ and an operator A, we will use the following "double" decomposition:

$$(A_{\nu_1})_{\nu_2} = \sum_{\eta'_2-\eta_2=\nu_2}\sum_{\eta'_1-\eta_1=\nu_1}\Pi_{\eta'_2}\Pi_{\eta'_1}A\Pi_{\eta_1}\Pi_{\eta_2}\,.$$

Here η_1,η'_1 (resp. η_2,η'_2) denote eigenvalues of H_1 (resp. η_2), and the Π's are projectors to the corresponding eigenspaces.

This double decomposition gives us a way to analyze objects like the commutator on the left-hand side of the identifiability equation.

Lemma 156 (Calculations with double decomposition)**.** *The following identities hold:*

$$[A, H_2 - H_1] = -\sum_{\nu_1,\nu_2} (A_{\nu_1})_{\nu_2}(\nu_2 - \nu_1)\,.$$

$$e^{H_2}e^{-H_1}Ae^{H_1}e^{-H_2} - e^{-H_2}e^{H_1}Ae^{-H_1}e^{H_2} = \sum_{\nu_1,\nu_2} (A_{\nu_1})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1)\,.$$

Proof. We have

$$\begin{aligned}
[A, H_1] &= AH_1 - H_1A \\
&= \sum_{\nu_1} \sum_{\eta' - \eta = \nu_1} \Pi_{\eta'} A \Pi_\eta H_1 - H_1 \Pi_{\eta'} A \Pi_\eta \\
&= \sum_{\nu_1} \sum_{\eta' - \eta = \nu_1} (\eta - \eta') \Pi_{\eta'} A \Pi_\eta \\
&= -\sum_{\nu_1} \nu_1 A_{\nu_1}\,,
\end{aligned}$$

and similarly for $[A, H_2]$. So the first part follows.

For the second part,

$$e^{-H_1}Ae^{H_1} = \sum_{\nu_1} \sum_{\eta_1' - \eta_1 = \nu_1} e^{-(\eta_1' - \eta_1)} \Pi_{\eta_1'} A \Pi_{\eta_1} = \sum_{\nu_1} e^{-\nu_1} A_{\nu_1}\,,$$

and similarly

$$e^{H_2} A_{\nu_1} e^{-H_2} = \sum_{\nu_2} e^{\nu_2} (A_{\nu_1})_{\nu_2}\,,$$

so

$$e^{H_2}e^{-H_1}Ae^{H_1}e^{-H_2} = \sum_{\nu_1,\nu_2} e^{\nu_2 - \nu_1} (A_{\nu_1})_{\nu_2}$$

and similarly

$$e^{-H_2}e^{H_1}Ae^{-H_1}e^{H_2} = \sum_{\nu_1,\nu_2} e^{\nu_1 - \nu_2} (A_{\nu_1})_{\nu_2}\,.$$

Note that $e^{\nu_2 - \nu_1} - e^{\nu_1 - \nu_2} = 2\sinh(\nu_2 - \nu_1)$, so the second part follows. □

As sinh is a bijective function, the above lemma gives a crucial link between commutator differences and interleaved imaginary time evolution differences, formalized as follows:

Lemma 157 (Commutator difference in time domain)**.**

$$[A, H_2 - H_1] =$$
$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Big[e^{H_2}e^{-H_1} A_{H_1}(t) e^{H_1}e^{-H_2} - e^{-H_2}e^{H_1} A_{H_1}(t) e^{-H_1}e^{H_2} \Big]_{H_2} (-t) \cdot g(t)\,\mathrm{d}t$$

for

$$\hat{g}[\omega] = -\frac{\omega}{2\sinh[\omega]}$$

and

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{g}[\omega] e^{-i\omega t}\,\mathrm{d}t = -\frac{\pi^{3/2}}{2\sqrt{2}(1 + \cosh(\pi t))}\,.$$

PROOF. We have

$$\begin{aligned}
[A, H_2 - H_1] &= -\sum_{\nu_1,\nu_2} (A_{\nu_1})_{\nu_2}(\nu_2 - \nu_1) \\
&= \sum_{\nu_1,\nu_2} \hat{g}(\nu_2 - \nu_1) \cdot (A_{\nu_1})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1) \\
&= \frac{1}{\sqrt{2\pi}} \sum_{\nu_1,\nu_2} \int_{-\infty}^{\infty} g(t) e^{-i(\nu_2-\nu_1)t} (A_{\nu_1})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1)\, \mathrm{d}t \\
&= \frac{1}{\sqrt{2\pi}} \sum_{\nu_1,\nu_2} \int_{-\infty}^{\infty} e^{-i\nu_2 t} g(t) (A_{\nu_1} e^{i\nu_1 t})_{\nu_2} \cdot 2\sinh(\nu_2 - \nu_1)\, \mathrm{d}t\,.
\end{aligned}$$

where in the third step we used Fourier inversion. Using Lemma 150 and Lemma 156, the claimed identity follows. □

Note that the function $g(t)$ is rapidly decaying, see Figure 1 below.

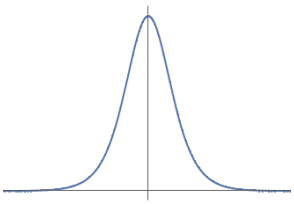


FIGURE 1. Plot of $g(t)$ from Theorem 153

We can now complete the proof of the identifiability equation.

PROOF OF THEOREM 153. Let ρ and ρ' denote the Gibbs states for H and H'. Taking $H_1 = \beta H'/2$ and $H_2 = \beta H/2$ in Lemma 157, consider the first term in the integral. Using that time evolution via H commutes with left- and right-multiplication by $\sqrt{\rho}$, we find that the first term equals

$$\begin{aligned}
&\mathrm{tr}(\sqrt{\rho} O^\dagger \sqrt{\rho} [\sqrt{\rho^{-1}} \sqrt{\rho'} A_{\beta H'/2}(t) \sqrt{\rho'^{-1}} \sqrt{\rho}]_{\beta H/2}(-t)) \\
&\quad = \mathrm{tr}(O^\dagger [\sqrt{\rho'} A_{\beta H'/2}(t) \sqrt{\rho'^{-1}} \rho]_{\beta H/2}(-t)) \\
&\quad = \mathrm{tr}(O^\dagger_{\beta H/2}(t) \cdot \sqrt{\rho'} A_{\beta H'/2}(t) \sqrt{\rho'^{-1}} \rho) \\
&\quad = \mathrm{tr}(O^\dagger_H(t\beta/2) \cdot \sqrt{\rho'} A_{H'}(t\beta/2) \sqrt{\rho'^{-1}} \rho)\,,
\end{aligned}$$

where in the penultimate step we pushed the reverse time evolution onto $O^\dagger$. Similarly

$$\begin{aligned}
&\mathrm{tr}(\sqrt{\rho} O^\dagger \sqrt{\rho} [\sqrt{\rho} \sqrt{\rho'^{-1}} A_{\beta H'/2}(t) \sqrt{\rho'} \sqrt{\rho^{-1}}]_{\beta H/2}(-t)) \\
&\quad = \mathrm{tr}(O^\dagger [\rho \sqrt{\rho'^{-1}} A_{\beta H'/2}(t) \sqrt{\rho'}]_{\beta H/2}(-t)) \\
&\quad = \mathrm{tr}(O^\dagger_{\beta H/2}(t) \cdot \rho \sqrt{\rho'^{-1}} A_{\beta H'/2}(t) \sqrt{\rho'}) \\
&\quad = \mathrm{tr}(O^\dagger_H(t\beta/2) \cdot \rho \sqrt{\rho'^{-1}} A_{H'}(t\beta/2) \sqrt{\rho'})\,.
\end{aligned}$$

We conclude that

$$\frac{\beta}{2} \langle O, [A, H - H'] \rangle_\rho = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{tr}(\rho \Delta[\![H'; O^\dagger_H(t\beta/2), A, t\beta/2]\!])\, \mathrm{d}t\,.$$

The identifiability equation follows by a change of variable $t \mapsto 2t/\beta$. □

3. Regularization

3.1. Preliminaries

Definition 158 (Operator Fourier transform)**.** *Given a Hamiltonian H and an operator A, define the* ***operator Fourier transform (FT)*** *$\hat{A}_H$ by*

$$\hat{A}_H[\omega] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A_H(t) e^{-i\omega t} f(t)\, \mathrm{d}t\,,$$

where $f(t) = e^{-\sigma^2 t^2} \sqrt{\sigma\sqrt{2/\pi}}$ is a Gaussian filter. The "regularizing" role of $f(t)$ will be become clearer in the sequel. Its Fourier transform $\hat{f}[\omega] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega t} f(t)\, \mathrm{d}t$ satisfies $\hat{f}[\omega] = \frac{1}{\sqrt{\sigma\sqrt{2\pi}}} \exp(-\omega^2/4\sigma^2)$.

Note that the operator FT commutes with imaginary time evolution:

$$e^{\beta H} \hat{A}_H[\omega] e^{-\beta H} = (\widehat{e^{\beta H} A e^{-\beta H}})_H[\omega]$$

Taking the operator FT of both sides of Lemma 150 results in the following useful identity:

$$\hat{A}_H[\omega] = \sum_{\nu} A_\nu \hat{f}[\omega - \nu]\,.$$

In other words, the operator FT gives "soft" access to the components in the Bohr decomposition of A. We have a corresponding "soft" Bohr decomposition, by Fourier duality.

Lemma 159. *For any operator A and Hermitian H,*

$$A = C_\sigma \int_{-\infty}^{\infty} \hat{A}_H[\omega]\, \mathrm{d}\omega$$

for $C_\sigma := \frac{1}{\sqrt{2\sigma\sqrt{2\pi}}}$.

Importantly, a straightforward calculation shows that the Gaussian filter ensures the operator FT decays exponentially in the frequency ω:

Lemma 160. *For any frequency ω and operator A satisfying $\|A\|_{\mathsf{op}} \leq 1$,*

$$\hat{A}_H[\omega] = e^{-\beta\omega + \sigma^2\beta^2} e^{\beta H} \hat{A}_H[\omega - 2\sigma^2\beta] e^{-\beta H}$$

To see why this is useful, note that because $\|\hat{A}_H[\omega]\|_{\mathsf{op}} \leq \hat{f}(0) = O(\sigma^{-1/2})$, this ensures that $\|e^{\beta H} \hat{A}_H[\omega'] e^{-\beta H}\|_{\mathsf{op}} \lesssim e^{\sigma^2\beta^2 + \beta\omega'} \sigma^{-1/2}$. Crucially, the right-hand scales *exponentially in the frequency ω'*, rather than exponentially in the system size! In contrast, norm of the imaginary time-evolved observable $\|e^{\beta H} A e^{-\beta H}\|_{\mathsf{op}}$ can scale exponentially in the system size.

3.2. Truncating the identifiability observable

Using Lemma 159, we can decompose A in $\langle O, [A, H - H']\rangle_\rho$ into low-frequency and high-frequency terms under operator FT with respect to H' and apply the

identifiability equation in Theorem 153 to obtain

$$\frac{\beta}{2C_\sigma}\langle O,[A,H-H']\rangle_\rho = \int_{|\omega'|\le\Omega'}\operatorname{tr}(\rho\overline{\Delta}[\![H';O,\hat{A}_{H'}[\omega']]\!])\,\mathrm{d}\omega' + \frac{\beta}{2}\int_{|\omega'|\ge\Omega'}\langle O,[\hat{A}_{H'}[\omega'],H-H']\rangle_\rho\,\mathrm{d}\omega'\,.$$

Let us try to write down a slightly more palatable expression for the first integral that doesn't involve the operator FT. Note that

$$\hat{A}_{H'}[\omega']_{H'}(t-i\beta/2) = (\sqrt{\rho'}\hat{A}_{H'}[\omega']\sqrt{\rho'^{-1}})_{H'}(t)\,,$$

and

$$\begin{aligned}&\int_{|\omega'|\le\Omega'}\sqrt{\rho'}\hat{A}_{H'}[\omega']\sqrt{\rho'^{-1}}\,\mathrm{d}\omega'\\ &\quad=\int_{|\omega'|\le\Omega'}\hat{A}_{H'}[\omega'-\sigma^2\beta]e^{-\beta\omega'/2+\sigma^2\beta^2/4}\,\mathrm{d}\omega'\\ &\quad=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}A_{H'}(t')\underbrace{\int_{|\omega'|\le\Omega'}e^{-i(\omega'-\sigma^2\beta)t'}e^{-\beta\omega'/2+\sigma^2\beta^2/4}\,\mathrm{d}\omega' f(t')}_{h_+(t')}\,\mathrm{d}t'\,,\end{aligned}$$

where in the first step we used Lemma 160, and similarly

$$\begin{aligned}&\int_{|\omega'|\le\Omega'}\sqrt{\rho'^{-1}}\hat{A}_{H'}[\omega']\sqrt{\rho'}\,\mathrm{d}\omega'\\ &\quad=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}A_{H'}(t')\underbrace{\int_{|\omega'|\le\Omega'}e^{-i(\omega'+\sigma^2\beta)t'}e^{\beta\omega'/2+\sigma^2\beta^2/4}\,D\omega' f(t')}_{h_-(t')}\,\mathrm{d}t'\,.\end{aligned}$$

Observe that

$$|h_+(t)|,|h_-(t)|\le O\Big(\frac{\sqrt{\sigma}}{\beta}e^{-\sigma^2t^2+\beta\Omega'/2+\sigma^2\beta^2/4}\Big)\,,$$

i.e. these functions are rapidly decaying in t.

Summarizing, we have the following:

Lemma 161. *Let $\Omega'>0$ and define the* truncated observable

$$\begin{aligned}&\overline{\Delta}^{\le\Omega'}[\![H';O,A]\!]\\ &:=\frac{1}{\sqrt{2\pi}}\iint_{-\infty}^{\infty}\Big(h_+(t')O_H^\dagger(t)A_{H'}(t'+t)-h_-(t')A_{H'}(t'+t)O_H^\dagger(t)\Big)\,g_\beta(t)\,\mathrm{d}t'\mathrm{d}t\,.\end{aligned}\tag{59}$$

Then

$$\frac{\beta}{2C_\sigma}\langle O,[A,H-H']\rangle_\rho=\operatorname{tr}(\rho\overline{\Delta}^{\le\Omega'}[\![H';O,A]\!])+\frac{\beta}{2}\int_{|\omega'|\ge\Omega'}\langle O,[\hat{A}_{H'}[\omega'],H-H']\rangle_\rho\,\mathrm{d}\omega'\,.$$

Let's take stock of what this buys us. First, because g_β, h_+, and h_- are rapidly decaying, the bulk of the double integral in the truncated observable $\overline{\Delta}^{\le\Omega'}[\![H';O,A]\!]$ is coming from short-time evolutions of O and A, which are local by the aforementioned Lieb-Robinson bounds. In short, if we only look at the "low-degree" term

in Lemma 161, we now have an observable which is entirely local which captures the discrepancy between H and H'.

It still remains to control the truncation error term $\int_{|\omega'|\geq\Omega'}$. For this, we can use Lemma 160 in conjunction with locality of $H - H'$ and A to show that for $\Omega' \geq \Omega(\sigma^2/\mathfrak{d})$, where $\mathfrak{d}$ is the degree of the dual interaction graph, the truncation error is negligible. We defer the details to [**CAN25**, Lemma III.5].

As discussed above, there is still one important missing piece before we can turn the above into a learning algorithm. The issue is that the truncated observable in Eq. (59) ultimately still depends on H through $O_H^\dagger(t)$. We explain the workaround for this next.

4. Learning Algorithm

In this section we describe how to exploit the ingredients from the preceding sections, deferring a complete proof of correctness to [**CAN25**].

To sidestep the issue that the truncated observable defined in Eq. (59) depends on H, we first define a broader class of observables that contains this observable.

Definition 162 (General truncated observables)**.** *Fix $\Omega' > 0$. Given operators K, O, A, G, with K and G Hermitian, define*

$$\Delta^*[\![G, K; O, A]\!] := \frac{1}{\sqrt{2\pi}} \iint_{-\infty}^{\infty} \Big(h_+(t')O_G^\dagger(t)A_K(t'+t) - h_-(t')A_K(t'+t)O_G^\dagger(t)\Big) g_\beta(t)\,\mathrm{d}t'\mathrm{d}t\,. \quad (60)$$

Note that $\Delta^*[\![H, H'; O, A]\!] = \overline{\Delta}^{\leq\Omega'}[\![H'; O, A]\!]$.

By design, we have the following:

Proposition 163. *When $K = H$, then $\mathrm{tr}(\rho\Delta^*[\![G, K; O, A]\!]) = 0$ for all O, G, A.*

Proof. In the proof of Lemma 161, instead of passing to h_+, h_-, we can directly express the "low-degree" term $\mathrm{tr}(\rho\overline{\Delta}^{\mathrm{trunc}}(H'; O, A))$ as

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{tr}\Big[\rho \int_{|\omega'|\leq\Omega'} \Big(O_H^\dagger(t)\hat{A}_{H'}[\omega']_{H'}(t+i\beta/2) - \hat{A}_{H'}[\omega']_{H'}(t-i\beta/2)O_H^\dagger(t)\Big)\,\mathrm{d}\omega'\Big]\,\mathrm{d}t\,.$$

In the definition of $\Delta^*[\![G, K; O, A]\!]$, H' and H above are replaced by K and G respectively, yielding

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{tr}\Big[\rho \int_{|\omega'|\leq\Omega'} \Big(O_G^\dagger(t)\hat{A}_K[\omega']_K(t+i\beta/2) - \hat{A}_K[\omega']_K(t-i\beta/2)O_H^\dagger(t)\Big)\,\mathrm{d}\omega'\Big]\,\mathrm{d}t\,.$$

If $K = H$ however, then mirroring the proof of the KMS condition, we have that

$$\mathrm{tr}(\rho O_G^\dagger(t)\hat{A}_H[\omega']_H(t+i\beta/2)) = \mathrm{tr}(\sqrt{\rho}\hat{A}_K[\omega']_H(t)\sqrt{\rho}) = \mathrm{tr}(\rho\hat{A}_H[\omega']_H(t-i\beta/2)O_H^\dagger(t))\,,$$

so $\Delta^*[\![G, H; O, A]\!] = 0$ as claimed. □

This suggests that we can simply brute-force enumerate over a net of different K's, and for each one we check whether $\mathrm{tr}(\rho\Delta^*[\![G, K; O, A]\!]) \approx 0$ for all 1-local Paulis A, and O, G in a suitable net. Previously we considered taking $O = [A, H - H']$, but given that this depends on H, we can instead use the fact that

$$\|[A, H - H']\|_\rho^2 \leq 2\mathfrak{d} \max_a |\langle [A, P_a], [A, H - H']\rangle_\rho| \quad (61)$$

to restrict to $O = [A, P_a]$ for all 1-local Paulis A and terms a in the support of the Hamiltonian.

The (regularized) identifiability equation in Lemma 161, combined with Lemma 154 and the inequality in Eq. (61), ensures that if $\mathrm{tr}(\rho\Delta^*[\![G, K; O, A]\!]) \approx 0$ for all 1-local A, $O = [A, P_a]$, and G in a suitable net, then $K \approx H$.

Only one step remains: how do we enumerate over G, K? Naively, if the Hamiltonian has m terms, this would require enumerating over a net over $O(m)$-dimensional parameter space and incurring a runtime scaling exponentially in $O(m)$. Fortunately, there is a workaround that again exploits locality. The intuition is that in the definition of $\Delta^*[\![G, K; O, A]\!]$ in Eq. (60), if A is a 1-local Pauli acting on site i, then $A_K(t'+t)$ and $O^\dagger_G(t) = [A, P_a]^\dagger_G(t)$ are roughly supported on a small neighborhood around i (because t', t are not too large because of the exponential damping of g_β, h_+, h_-). Moreover, Lieb-Robinson bounds ensure that these operators do not change much when G and K are replaced by their truncations to a suitable neighborhood around the i-th site. Formally, we have the following estimate:

Lemma 164 (Lieb-Robinson bound)**.** *If Hamiltonian $H = \sum_a \lambda_a P_a$ with coefficients satisfying $|\lambda_a| \le 1$ has interaction degree $\mathfrak{d}$, then for any operator A acting on subsystem $S \subseteq [n]$ and satisfying $\|A\|_{\mathsf{op}} \le 1$, if H_ℓ is given by removing all terms from H at distance at least ℓ from S, then*

$$\|A_{H_\ell}(t) - A_H(t)\|_{\mathsf{op}} \le O\Big(|S| \cdot \frac{(2\mathfrak{d}|t|)^\ell}{\ell!}\Big).$$

The proof of this will be the subject of one of the homework exercises.
With this in hand, we essentially have a complete, albeit informal, description of the algorithm:

- For each qubit $i \in [n]$:
  (1) Enumerate over a net of local Hamiltonians K_ℓ acting on the neighborhood $V(\ell, i)$ of radius ℓ around the i-th site
  (2) For each such K_ℓ, use $O(\log n)$ copies of ρ to estimate the observable values $\mathrm{tr}(\rho\Delta^*[\![G_\ell, K_\ell; [A, P_a], A]\!])$ for all local Hamiltonians G_ℓ acting on $V(\ell, i)$ and all terms P_a and 1-local Paulis A.
  (3) If for any such K_ℓ all of these observable values are small, then we will take our estimate of H over the local patch $V(\ell, i)$ to be K_ℓ.

The quantitative details are somewhat dense and do not provide much additional insight beyond the intuition outlined above, so we defer these to [**CAN25**].

CHAPTER 10

Learning Stabilizer States

We now introduce an important family of states that lies at the boundary between classical and quantum computation: **stabilizer states**. A salient property of these states is that they can be highly entangled, yet they are *classically simulable* in the sense that one can sample from the probability distribution encoded by their amplitudes using classical computation – this is the content of the *Gottesman-Knill theorem*, which we will not prove in this course [GOT98]. Indeed as we will see in the next lecture, they offer a useful yardstick by which to quantify the extent to which a given quantum computation is truly quantum. Stabilizer states also come with rich algebraic structure and are characterized by their symmetries. This structure makes them particularly useful for robustly encoding quantum information to be resilient to noise, and for this reason they play a central role in the field of *quantum error correction*.

In this lecture, we continue the theme of learning structured classes of states by giving an efficient algorithm for learning stabilizer states, due to Montanaro [Mon07]. Unlike the algorithms we saw for shallow circuit states and Gibbs states, this algorithm will heavily exploit the algebraic structure native to stabilizer states. This will also give us our first glimpse at a powerful and ubiquitous primitive: **Bell sampling**.

1. Stabilizer State Basics

Stabilizer states can be defined either in terms of the quantum circuits that prepare them, or in terms of the symmetries that they possess.

Definition 165 (Clifford group)**.** *The **Clifford group** $\mathcal{C}_n$ on n qubits is the group of unitaries U for which $UPU^{-1} \in \mathcal{P}_n$ for all $P \in \mathcal{P}_n$. One choice of gates generating $\mathcal{C}_n$ are Hadamard, phase, and CNOT:*

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \qquad \text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

*We will refer to elements of $\mathcal{C}_n$ as **Clifford circuits** as they can be built out of these gates.*

Definition 166 (Stabilizer states – Clifford circuit formulation)**.** *A state $|\psi\rangle$ is a* stabilizer state *if it can be written as $|\psi\rangle = U|0^n\rangle$ for $U \in \mathcal{C}_n$.*

Lemma 167 (Symmetry formulation)**.** *A state $|\psi\rangle$ is a stabilizer state if and only if there are exactly 2^n commuting Pauli operators $P \in \{\pm I, \pm X, \pm Y, \pm Z\}^{\otimes n}$ that* stabilize *$|\psi\rangle$, that is, for which $P|\psi\rangle = |\psi\rangle$.*

PROOF. Suppose $|\psi\rangle = U|0^n\rangle$ for $U \in \mathcal{C}_n$. There are exactly 2^n Paulis P for which $P|0^n\rangle = |0^n\rangle$, namely $P \in \{I, Z\}^{\otimes n}$. Because U is Clifford, for every $P \in \mathcal{P}_n$ we have $UP = P_U U$ for a unique $P_U \in \mathcal{P}_n$. Therefore, for each $P \in \{I, Z\}^{\otimes n}$,

$$P_U|\psi\rangle = P_U U|0^n\rangle = UP|0^n\rangle = U|0^n\rangle = |\psi\rangle ,$$

so there are exactly 2^n Paulis that stabilize $|\psi\rangle$. Furthermore, they commute because they stabilize $|\psi\rangle$ and all Paulis either commute or anti-commute with each other.

For the converse, we provide a sketch of the proof. Let $P_1, \ldots, P_n \in G$ be a set of generators for the abelian group consisting of stabilizers of $|\psi\rangle$, and let $Z_1, \ldots, Z_n$ denote the Z operators on qubits $1, \ldots, n$. Each P_j can be expressed as $\prod_i Z_i^{a_{ij}}$, so we can effectively perform Gaussian elimination to obtain $U \in \mathcal{C}_n$ for which $UP_iU^{-1} = Z_i$ for all i (the details for this are provided in Lemma 177). As $P_i|\psi\rangle = |\psi\rangle$, we have that

$$U|\psi\rangle = UP_i|\psi\rangle = Z_iU|\psi\rangle ,$$

so $U|\psi\rangle$ is stabilized by all Paulis in $\{I, Z\}^{\otimes n}$. Therefore, $|\psi\rangle = U^\dagger|\phi\rangle$ for $|\phi\rangle \in \{|0\rangle, |1\rangle\}^{\otimes n}$, and thus $|\psi\rangle = U'|0^n\rangle$ for some $U' \in \mathcal{C}_n$. □

The group of Paulis stabilizing a stabilizer state is important enough to merit a name:

Definition 168. *Given a stabilizer state $|\psi\rangle$, the group of 2^n Pauli operators P for which $P|\psi\rangle = |\psi\rangle$ is called the **stabilizer group** of $|\psi\rangle$, denoted* $\mathrm{Stab}(|\psi\rangle)$.

Example 169. *As we saw in the proof above, the simplest stabilizer state is $|0^n\rangle$, whose stabilizer group is $\{I, Z\}^{\otimes n}$. More generally, any state which is a product of single-qubit states from $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i\rangle, |-i\rangle\}$ is a stabilizer state.*

Another example to keep in mind is the n-qubit cat state $\frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$, which can be prepared starting from $|0^n\rangle$ by applying H to one qubit and then taking CNOTs with all of the remaining qubits.

There is a third formulation of stabilizer states in terms of their amplitudes in the computational basis:

Lemma 170. *If $|\psi\rangle$ is stabilizer, then it is equal, up to phase, to*

$$\frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{\ell(x)} (-1)^{q(x)} |x\rangle \tag{62}$$

for some affine subspace $A \subseteq \mathbb{F}_2^n$, linear function $\ell : \mathbb{F}_2^n \to \mathbb{F}_2$, and quadratic function $q : \mathbb{F}_2^n \to \mathbb{F}_2$.

PROOF SKETCH. This follows by direct calculation by inducting on the number of gates in the Clifford circuit preparing $|\psi\rangle$. It is not hard to see that applying phase or CNOT gates will respectively modify ℓ and linearly transform A. The trickiest part to verify is that applying a Hadamard gate to a state of the form Eq. (62) results in another state of the same form but with A, ℓ, and q all modified. The complete calculation is provided in Appendix A of [**VDN10**]. □

In fact the converse holds, though we will not prove or use this.

2. Symplectic Vector Spaces – A First Glimpse

When ρ is pure, the distribution over measurement outcomes under Bell sampling has a useful characterization. It will be convenient to adopt the following mapping between Pauli operators and bitstrings.

Definition 171 (Bijection between Paulis and strings)**.** *Given a string* $\vec{s} = (s_{1,1}, \ldots, s_{1,n}, s_{2,1}, \ldots, s_{2,n}) \in \mathbb{F}_2^{2n}$*, define*

$$P_{\vec{s}} \triangleq \bigotimes_{j=1}^{n} i^{s_{1,j} \cdot s_{2,j}} X_j^{s_{1,j}} Z_j^{s_{2,j}} .$$

Every $P \in \mathcal{P}_n$ *can then be written as* $\phi \cdot P_{\vec{s}}$ *for some phase* $\phi \in \{\pm 1, \pm i\}$*.*

With this identification, we can naturally associate to every element of $\mathrm{Stab}(|\psi\rangle)$ a string as follows:

Definition 172 (Unsigned stabilizer group)**.** *The* ***unsigned stabilizer group*** *of a stabilizer state* $|\psi\rangle$*, denoted* $\mathrm{Weyl}(|\psi\rangle)$*, is the set of* $\vec{s} \in \mathbb{F}_2^{2n}$ *for which either* $P_{\vec{s}}|\psi\rangle = |\psi\rangle$ *or* $P_{\vec{s}}|\psi\rangle = -|\psi\rangle$*. Note that the elements of* $\mathrm{Weyl}(|\psi\rangle)$ *and* $\mathrm{Stab}(|\psi\rangle)$ *are in one-to-one correspondence up to sign.*[1]

In fact the mapping in Definition 171 has even richer structure: commutation relations between Pauli operators correspond to linear algebraic relations between their associated vectors in $\mathbb{F}_2^{2n}$ equipped with the *symplectic* inner product.

Definition 173 (Symplectic inner product)**.** *Given* $\vec{s}, \vec{t} \in \mathbb{F}_2^{2n}$ *with entries*

$$\vec{s} = (s_{1,1}, \ldots, s_{1,n}, s_{2,1}, \ldots, s_{2,n})$$
$$\vec{t} = (t_{1,1}, \ldots, t_{1,n}, t_{2,1}, \ldots, t_{2,n}) ,$$

their ***symplectic inner product****, which we will denote by* $[\vec{s}, \vec{t}]$*, is given by*

$$[\vec{s}, \vec{t}] = \sum_{j=1}^{n} (s_{1,j} t_{2,j} + s_{2,j} t_{1,j}) .$$

Our motivation for using this notation is the following elegant fact:

Lemma 174 (Commutation as symplectic orthogonality)**.** *For any* $\vec{s}, \vec{t} \in \mathbb{F}_2^{2n}$*,* $[\vec{s}, \vec{t}] = 0$ *(resp. 1) if and only if* $P_{\vec{s}}$ *and* $P_{\vec{t}}$ *commute (resp. anti-commute).*

Proof. It suffices to show this at the level of a single qubit. Let $s = (s_1, s_2)$ and $t = (t_1, t_2)$, so that $P_s = i^{s_1 s_2} X^{s_1} Z^{s_2}$ and $P_t = i^{t_1 t_2} X^{t_1} Z^{t_2}$. Then

$$P_s P_t - P_t P_s = i^{s_1 s_2 + t_1 t_2} (X^{s_1} Z^{s_2} X^{t_1} Z^{t_2} - X^{t_1} Z^{t_2} X^{s_1} Z^{s_2}) .$$

It can be verified that the first term in the parentheses is $(-1)^{s_2 t_1} \cdot X^{s_1 + t_1} Z^{s_2 + t_2}$, and likewise the second term is $(-1)^{s_1 t_2} \cdot X^{s_1 + t_1} Z^{s_2 + t_2}$. So the above expression is zero if and only $s_2 t_1 = s_1 t_2$. □

Definition 175. *A subspace* $T \subseteq \mathbb{F}_2^{2n}$ *is* ***isotropic*** *if* $[\vec{s}, \vec{t}] = 0$ *for all distinct* $\vec{s}, \vec{t} \in T$*.*

[1] However, be wary that despite the terminology, the unsigned stabilizer group is not necessarily a group – for example, consider the stabilizer group $\{II, XX, -YY, ZZ\}$, for which the corresponding unsigned stabilizer group $\{II, XX, YY, ZZ\}$ is not closed under multiplication.

Lemma 176. *For any (abelian) subgroup $G \subseteq \mathcal{P}_n$, the corresponding strings form an (isotropic) subspace of $\mathbb{F}_2^{2n}$. Conversely, any (isotropic) subspace of $\mathbb{F}_2^{2n}$ corresponds to an (abelian) subgroup $G \subseteq \mathcal{P}_n$.*

Proof. This equivalence follows immediately from Lemma 174 and the fact that $P_{\vec{s}}P_{\vec{t}}$ and $P_{\vec{s}\oplus\vec{t}}$ are equal to each other up to phase for any strings $\vec{s}, \vec{t}$. □

This symplectic structure will be especially useful in the next lecture. For now, the most important takeaway from the association between $\{I, X, Y, Z\}^{\otimes n}$ and $\mathbb{F}_2^{2n}$ is that it allows us to efficiently and classically perform manipulations of Pauli operators. This will allow us to flesh out the details in the proof sketch of the second part of Lemma 167.

Lemma 177. *Given a collection of vectors $\vec{s}^1, \ldots, \vec{s}^m \in \mathbb{F}_2^{2n}$ which are mutually orthogonal with respect to the symplectic inner product, there is an $O(n^3)$-time classical algorithm that outputs a minimal subset $S \subseteq [m]$ such that $\{P_{\vec{s}^j}\}_{j\in S}$ generates all of $\{P_{\vec{s}^j}\}_{j\in[m]}$. If $|S| = n$, the algorithm additionally outputs a classical description of a Clifford circuit $U \in \mathcal{C}_n$ for which the corresponding stabilizer state is $U|0^n\rangle$, and for which $UP_{\vec{s}^j}U^{-1} = Z_j$ for all j.*

The proof of this can be skipped upon first reading, as it essentially amounts to Gaussian elimination.

Proof. This amounts to finding a *basis* for the rows of the matrix

$$M = \left(\begin{array}{ccc|ccc} s^1_{1,1} & \cdots & s^1_{1,n} & s^1_{2,1} & \cdots & s^1_{2,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s^m_{1,1} & \cdots & s^m_{1,n} & s^m_{2,1} & \cdots & s^m_{2,n} \end{array}\right).$$

We will refer to the submatrix on the left (resp. right) of the divider as the "X block" (resp. "Z block"). We can find a row basis for M via Gaussian elimination. More specifically, we can apply column operations to this matrix to place it in reduced column echelon form, with nonzero columns within the Z block. This is done by a combination of (1) swapping columns within the X block, (2) swapping a column in the X block with a corresponding column in the Z block, (3) adding columns in the X block to the corresponding columns in the Z block, and (4) adding the i-th column in both the X and Z blocks to the j-th column in both blocks respectively for various i, j. (1) corresponds to SWAPs which can be implemented with Clifford gates, (2) can be implemented with H gates, (3) can be implemented with phase gates, and (4) can be implemented with CNOT gates. Note that these column operations do not change the commutation relations among the Paulis associated to the rows.

The first part of the lemma then follows by selecting the rows of the resulting matrix corresponding to the identity block in reduced column echelon form. The second part of the lemma follows from the fact that if this block is $n \times n$, it occupies the entire Z block, and the X block is zero. Because the stabilizer state associated to this matrix is simply $|0^n\rangle$, the Clifford circuit U in the lemma statement can be read off from the sequence of gates that were used to implement the above column operations. □

The upshot is that in order to learn a classical description of a stabilizer state, it suffices to learn a classical description of its stabilizer group. This is what we will do in the rest of the lecture.

3. Bell Sampling and Bell Difference Sampling

We now introduce the key primitive behind the algorithm that we will present for learning stabilizer states. We begin by defining an important measurement basis:

Definition 178 (Bell basis)**.** *Define the **Bell states***

$$\sigma_{00} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \qquad \sigma_{01} = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$
$$\sigma_{10} = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \qquad \sigma_{11} = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)\,.$$

*The **Bell basis** consists of all 2^n states on $2n$ qubits which take the form*

$$\sigma_{\vec{s}} = \sigma_{s_1} \otimes \cdots \otimes \sigma_{s_n}\,,$$

where $s_1, \ldots, s_n \in \{00, 01, 10, 11\}$.

*Given an n-qubit state ρ, **Bell sampling** is the operation of performing a measurement of $\rho^{\otimes 2}$ in the Bell basis.*

Definition 179 (Characteristic distribution)**.** *Given a pure state $\rho = |\psi\rangle\langle\psi|$, define the **characteristic distribution** q_ρ to be the distribution over $\vec{s} \in \{0,1\}^{2n}$ with probability mass function*

$$q_\rho(\vec{s}) \triangleq 2^{-n} \langle\psi| P_{\vec{s}} |\psi\rangle^2\,.$$

(Exercise: verify that this is indeed a probability distribution)

Proposition 180. *If $|\psi\rangle$ is stabilizer, then the characteristic distribution is uniform over* $\mathrm{Weyl}(|\psi\rangle)$.

Proof. For $P \in \mathrm{Weyl}(|\psi\rangle)$, we have $2^{-n}\langle\psi| P |\psi\rangle^2 = 2^{-n}$. There are exactly 2^n Paulis in the unsigned stabilizer group, so the claim follows. □

Lemma 181. *When one performs Bell sampling on pure state $\rho = |\psi\rangle\langle\psi|$, one observes measurement outcome $s \in \{0,1\}^{2n}$ with probability $2^{-n}|\langle\psi| P_{\vec{s}} |\psi^*\rangle|^2$, where $|\psi^*\rangle$ denotes the entrywise conjugation of $|\psi\rangle$ in the computational basis.*

Proof. Note that the Bell states satisfy $\sigma_s = \frac{1}{\sqrt{2}}\mathrm{vec}(P_s)$ for all $s \in \{00, 01, 10\}$ and $\sigma_{11} = i \cdot \frac{1}{\sqrt{2}}\mathrm{vec}(P_{11})$, so

$$\sigma_{\vec{s}} = C_{\vec{s}} \cdot \mathrm{vec}(P_{\vec{s}}) \qquad \text{for } C_{\vec{s}} \triangleq 2^{-n/2} i^{\#\{j:(s_{2j}, s_{2j+1}) = (1,1)\}}\,.$$

As $|\psi\rangle|\psi\rangle = \mathrm{vec}(|\psi\rangle\langle\psi^*|)$ – where vec denotes the operation that flattens a matrix into a vector, and $|\psi^*\rangle$ denotes the entrywise conjugation of $|\psi\rangle$ in the computational basis – we have

$$|\langle\sigma_{\vec{s}}|\psi\rangle|\psi\rangle|^2 = 2^{-n}|\mathrm{tr}(P_{\vec{s}}|\psi\rangle\langle\psi^*|)|^2 = 2^{-n}|\langle\psi| P_{\vec{s}} |\psi^*\rangle|^2\,.$$

as claimed. □

Corollary 182. *When one performs Bell sampling on stabilizer state $\rho = |\psi\rangle\langle\psi|$, there is some $\vec{t} \in \{0,1\}^{2n}$ such that the distribution over measurement outcomes places mass $q_\rho(\vec{s} \oplus \vec{t})$ on any string $s \in \{0,1\}^{2n}$.*

PROOF. By Lemma 170, $|\psi\rangle$ takes the form of Eq. (62). As $i^{\ell(x)} = \prod_{j\in S} i^{x_j}$ for some $S \subseteq [n]$, and $\overline{i^a} \cdot |a\rangle = i^a \cdot Z|a\rangle$ for $a \in \{0,1\}$, we find that up to phase, $|\psi^*\rangle$ is given by $Z^{\otimes S}|\psi\rangle$. By Lemma 181, we conclude that under Bell sampling, we observe outcome s with probability

$$2^{-n}|\langle\psi| P_{\vec{s}} Z^{\otimes S}|\psi\rangle|^2 = q_\rho(\vec{s}\oplus\vec{t})$$

for some fixed string t satisfying $P_{\vec{s}}Z^{\otimes S} = P_{\vec{s}\oplus\vec{t}}$, as claimed. □

The upshot is that if one performs Bell sampling on a stabilizer state, one obtains a sample from the characteristic distribution, but with an unknown *shift* $\vec{t}$. By Proposition 180, this is the uniform distribution over strings of the form $\vec{s}\oplus\vec{t}$ where $\vec{s}$ ranges over strings corresponding to elements of $\mathrm{Weyl}(|\psi\rangle)$.

If we could get rid of this shift, we would be able to access uniformly random elements of $\mathrm{Weyl}(|\psi\rangle)$. The key idea is to run Bell sampling *twice*, and then XOR the two samples, which exactly cancels out the shift! This trick, due to [Mon07], has a name:

Definition 183 (Bell difference sampling)**.** ***Bell difference sampling*** *is the procedure of measuring a given state ρ in the Bell basis twice to get measurement outcomes $\vec{s}\oplus\vec{t}$ and $\vec{s'}\oplus\vec{t}$, and then outputting their XOR, namely $\vec{s}\oplus\vec{s'}$. We will denote the distribution over strings (or equivalently, their associated Pauli operators) obtained in this fashion by $\mathcal{B}_\rho$.*

Lemma 184. *If $|\psi\rangle$ is a stabilizer state, then Bell difference sampling results in a random sample from* $\mathrm{Weyl}(|\psi\rangle)$.

PROOF. By design, Bell difference sampling results in a sample of the form $\vec{s}\oplus\vec{s'}$, where both $\vec{s},\vec{s'}$ are uniform over $\mathrm{Weyl}(|\psi\rangle)$. As the uniform measure over a symplectic linear subspace is invariant under translation by any fixed element of that subspace, $\vec{s}\oplus\vec{s'}$ is itself distributed as a uniform sample from $\mathrm{Weyl}(|\psi\rangle)$. □

4. Learning Algorithm

This yields a simple algorithm, due to [Mon07], for learning the stabilizer group of an unknown stabilizer state:

Algorithm 7: LEARNSTABILIZERGROUP($|\psi\rangle$)

Input: Copies of stabilizer state $|\psi\rangle$
Output: Classical description of Clifford circuit preparing $|\psi\rangle$

Perform Bell difference sampling $2n$ times to get strings $\vec{s}^1,\ldots,\vec{s}^{2n}$
Compute a basis $\{\vec{s}^{j_1},\ldots,\vec{s}^{j_n}\}$ for $\vec{s}^1,\ldots,\vec{s}^{2n}$
Let $U \in \mathcal{C}_n$ be the Clifford circuit for which $UP_{\vec{s}^{j_i}}U^{-1} = Z_i$ for all $i \in [n]$
for $i \in [n]$ **do**
    Measure $|\psi\rangle$ in the eigenbasis of $P_{\vec{s}^{j_i}}$ to determine whether $P_{\vec{s}^{j_i}}|\psi\rangle = -|\psi\rangle$ or $P_{\vec{s}^{j_i}}|\psi\rangle = |\psi\rangle$
end
Let Q be the operator which acts as X in all qubits $i \in [n]$ for which $P_{\vec{s}^{j_i}}|\psi\rangle = -|\psi\rangle$
Output $U^\dagger Q$

Theorem 185. *Given access to copies of an unknown stabilizer state $|\psi\rangle$, the algorithm* LearnStabilizerGroup$(|\psi\rangle)$ *performs $O(n)$ two-copy measurements and with probability at least $1-2^{-n}$ outputs the classical description of a Clifford circuit preparing $|\psi\rangle$.*

Proof. The algorithm fails if and only if the strings $\vec{s}^1, \ldots, \vec{s}^{2n}$ are all contained in a subspace of dimension at most $n-1$. There are 2^n such subspaces, so by a union bound, this happens with probability at most $2^{-2n} \cdot 2^n = 2^{-n}$.

The operator Q is meant to correct for the fact that the group generated by $\{P_{\vec{s}^{j_i}}\}_{i\in[n]}$ is only the *unsigned* stabilizer group for $|\psi\rangle$. In particular, while $U^\dagger |0^n\rangle$ is the stabilizer state associated to those Paulis, the true stabilizer state is given by $U^\dagger |x\rangle$ for a string $x \in \{0,1\}^n$ whose 1 entries are precisely those i for which $P_{\vec{s}^{j_i}} |\psi\rangle = -|\psi\rangle$. □

While the algebraic structure in this learning gives rise to a very elegant protocol, it is clear that the guarantee we have proven is incredibly brittle. In particular, even if $|\psi\rangle$ were corrupted by a small amount of noise, it is no longer clear which parts of this lecture can be salvaged, if any. In the next lecture, we will show that remarkably, there is a way to redeem Bell difference sampling even in the presence of such "model misspecification."

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*

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

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 $F = \{P_1, \ldots, P_m\} \subset \{I, X, Y, Z\}^{\otimes n}$ 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 $P_1, \ldots, P_m$ 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 ρ, error parameters $0 < \epsilon, \delta < 1$
Output: ϵ-high-correlation basis of n commuting Paulis

1 Run Bell difference sampling $m = \Theta((n + \log 1/\delta)/\epsilon)$ times to obtain strings $\vec{s}_1, \ldots, \vec{s}_m$
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 ρ
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,

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

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 $p \leq \epsilon$, we are already done. Otherwise if $p > \epsilon$, by Chernoff bound $|S_{\mathsf{high}}| \geq pm/2$ 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, $\Pr_{\vec{s}\sim D_h}[y \notin \mathrm{span}(S_{\mathsf{high}})] \leq \epsilon/p$. 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 $\mathrm{tr}(O^2\rho) \geq \mathrm{tr}(O\rho)^2$ implies that $\mathrm{tr}(P\rho)^2 + \mathrm{tr}(Q\rho)^2 \leq 1$. □

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,*

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

PROOF. Let $V_i \triangleq \operatorname{span}(x_1, \ldots, x_i)$, 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 $\operatorname{Stab}(|\psi\rangle)$ 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 $\operatorname{Stab}(|\psi\rangle)$.

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 $\operatorname{Stab}(|\psi\rangle)$, the high-correlation Paulis in $\operatorname{Stab}(|\psi\rangle)$ are mostly concentrated around a *proper subspace* of $\operatorname{Stab}(|\psi\rangle)$. Anti-concentration then ensures that the next Bell difference sample has a non-negligible chance of yielding an element of $\operatorname{Stab}(|\psi\rangle)$ 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 $\operatorname{span}(F) \neq \operatorname{Weyl}(|\psi\rangle)$, then*

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

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

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

Then

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

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

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

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 $\mathrm{Stab}(|\psi\rangle)$ 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

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

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 $n-1$, $\sum_{\vec{s}\in V^*\setminus W}\langle\zeta|P_{\vec{s}}|\zeta\rangle \gtrsim 2^n\tau$.*

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 $|\pm i 0^{n-1}\rangle$ 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 $|\pm i 0^{n-1}\rangle$, which are given by

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

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 $\max_b \mathrm{Re}(a_0\cdot\overline{b a_1}) = a_0|a_1|C_\theta$ 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 $\vec{s}\in W$, $\{\vec{s}\oplus\vec{t}\colon \vec{s}\in V^*\backslash W\} = V^*\backslash W$, 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. □

CHAPTER 12

Learning short-range entangled states

Short-range entangled states form a fundamental class in condensed matter physics, characterized by area-law entanglement and the ability to be transformed into product states through constant-depth local unitary circuits. These states include ground states of gapped local Hamiltonians without topological order. The central question we investigate is whether copies of an unknown SRE state contain sufficient information to learn a quantum circuit that prepares the state, and whether this learning can be accomplished in polynomial time.

This lecture presents a polynomial-time algorithm for learning quantum states prepared by shallow circuits on finite-dimensional lattices. The algorithm uses novel geometric constructions that avoid solving constraint satisfaction problems. We show how careful application of local inversion operations, combined with covering schemes tailored to lattice geometry, enables efficient global state reconstruction from local reduced density matrices.

1. The Learning Problem

We formalize the computational problem of learning quantum states with low circuit complexity.

1.1. Problem Statement

Definition 204 (Learning short-range entangled states)**.** *Given access to N copies of an unknown n-qubit state $|\psi\rangle$, with the promise that $|\psi\rangle = U|0^n\rangle$ where U is an unknown depth-d quantum circuit acting on a k-dimensional lattice, the goal is to output a quantum circuit W such that*

$$\left\| W|0^{n+m}\rangle - |\psi\rangle \otimes |junk\rangle \right\|_1 \leq \epsilon$$

with high probability, where $m = \mathcal{O}(n)$ is the number of ancilla qubits and $\epsilon > 0$ is the target error.

The challenge is to achieve polynomial sample complexity $N = \text{poly}(n)$ and polynomial time complexity when $d = \mathcal{O}(1)$. The naive approach of performing full quantum state tomography requires exponentially many measurements and is computationally intractable.

2. Local Inversions and the Replacement Process

The key algorithmic primitive is the concept of local inversion, which enables us to systematically disentangle local regions without disturbing the global state.

2.1. Local Inversions

Definition 205 (Local Inversion)**.** *Given a state $|\psi\rangle$ on n qubits and a subset $A \subseteq [n]$, a unitary operator V is a **local inversion** of A if:*

(i) V acts nontrivially only on $B(A,d)$, the radius-d ball around A;
(ii) $V|\psi\rangle = |0\rangle_A \otimes |\phi\rangle$ for some state $|\phi\rangle$ on the remaining qubits.

When $|\psi\rangle = U|0^n\rangle$ is prepared by a depth-d circuit, local inversions always exist and can be constructed by inverting the gates in the lightcone of region A.

Lemma 206 (Existence of Local Inversions)**.** *Let $|\psi\rangle = U|0^n\rangle$ where U is a depth-d circuit on a k-dimensional lattice. For any region $A \subseteq [n]$, there exists a local inversion V of A such that:*

(i) V is supported nontrivially only on $B(A,d)$;
(ii) V can be implemented as a depth-d circuit whose gates lie entirely within the lightcone $\mathcal{L}(A,d)$.

Proof. Consider the lightcone $\mathcal{L}(A,d)$ of region A in the circuit U, defined as the set of all gates in U that can causally influence any qubit in A. This lightcone has depth at most d and acts on qubits within $B(A,d)$.

Let $U = U_d U_{d-1} \cdots U_1$ where U_j denotes the j-th layer. Define V by inverting all gates in $\mathcal{L}(A,d)$ in reverse order. More precisely, for each layer $j = 1, \ldots, d$, let G_j be the set of gates in layer U_j that belong to $\mathcal{L}(A,d)$. Then

$$V = \left(\prod_{G \in G_1} G^{-1}\right)\left(\prod_{G \in G_2} G^{-1}\right)\cdots\left(\prod_{G \in G_d} G^{-1}\right).$$

Note that within each product, the gates commute since they come from the same layer of non-overlapping gates.

Applying V to $|\psi\rangle = U|0^n\rangle$ undoes precisely the gates in the lightcone. Since the lightcone contains all gates affecting A, the qubits in A return to state $|0\rangle_A$. The qubits outside $B(A,d)$ are never affected by gates in the lightcone, so they remain in some state $|\phi\rangle$. Therefore $V|\psi\rangle = |0\rangle_A \otimes |\phi\rangle$ for some state $|\phi\rangle$ on the remaining $n - |A|$ qubits. □

2.2. Learning Approximate Local Inversions

While exact local inversions require knowledge of the unknown circuit U, we can learn approximate local inversions from copies of $|\psi\rangle$.

Definition 207 (ϵ-Approximate Local Inversion)**.** *A unitary V acting on $B(A,d)$ is an **ϵ-approximate local inversion** of region A for state $|\psi\rangle$ if*

$$\langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}\left[V|\psi\rangle\langle\psi|V^\dagger\right]|0\rangle_A \geq 1 - \epsilon.$$

Lemma 208 (Learning Local Inversions)**.** *Algorithm 9 outputs a ϵ_1-approximate local inversion using $M = 2^{\mathcal{O}(c)}/\epsilon_1^2$ copies of $|\psi\rangle$ and runs in time $T = 2^{\mathcal{O}(c)}/\epsilon_1^2 + (kdc/\epsilon_1)^{\mathcal{O}(dc)}$, where $c = |B(A,d)|$.*

Proof. By standard quantum state tomography results, learning the reduced density matrix $\rho_{B(A,d)}$ to error $\epsilon_1/3$ in trace distance requires $N = 2^{\mathcal{O}(c)}/\epsilon_1^2$ copies and time proportional to N. The size of an $(\epsilon_1/3)$-net over depth-d circuits on c qubits is $(kdc/\epsilon_1)^{\mathcal{O}(dc)}$ by standard covering number arguments for unitary groups.

Algorithm 9: LearnLocalInv($|\psi\rangle$, region A, depth d, error ϵ_1)

Input: Copies of state $|\psi\rangle$, region A, circuit depth d, error tolerance ϵ_1
Output: ϵ_1-approximate local inversion V of region A

Use quantum state tomography to learn $\rho_{B(A,d)}$ to error $\epsilon_1/3$;
Let $c = |B(A,d)|$ be the number of qubits in the ball;
Construct an $(\epsilon_1/3)$-net $\mathcal{N}$ over depth-d circuits on c qubits;
for *each circuit* $V \in \mathcal{N}$ **do**
    Compute $p = \langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V \rho_{B(A,d)} V^\dagger] |0\rangle_A$;
    if $p \geq 1 - 2\epsilon_1/3$ **then**
        return V;
    end
end
return *Failure*;

By Lemma 206, there exists an exact local inversion V_{exact} of region A satisfying $V_{\text{exact}} |\psi\rangle = |0\rangle_A \otimes |\phi\rangle$ for some state $|\phi\rangle$. Since the net is an $(\epsilon_1/3)$-net, it contains some circuit V_0 such that $\|V_0 - V_{\text{exact}}\|_{\text{op}} \leq \epsilon_1/3$.

We now show that the algorithm outputs some ϵ_1-approximate local inversion. Let us first analyze the circuit V_0 in the net. The probability computed by the algorithm for V_0 is

$$p_0 = \langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V_0 \rho_{B(A,d)} V_0^\dagger] |0\rangle_A .$$

First, we bound the difference between using the learned density matrix $\rho_{B(A,d)}$ versus the true reduced density matrix $|\psi\rangle\langle\psi|_{B(A,d)}$. By the tomography error and the fact that partial trace and unitary conjugation are non-increasing under trace distance (data-processing inequality),

$$\left| p_0 - \langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V_0 |\psi\rangle\langle\psi|_{B(A,d)} V_0^\dagger] |0\rangle_A \right| \leq \epsilon_1/3.$$

Second, we bound the difference between V_0 and V_{exact}. Since $\|V_0 - V_{\text{exact}}\|_{\text{op}} \leq \epsilon_1/3$, for any unit vector $|\alpha\rangle$, we have $\|V_0 |\alpha\rangle - V_{\text{exact}} |\alpha\rangle\|_2 \leq \epsilon_1/3$. By the relationship between fidelity and 2-norm distance, for any state $|\alpha\rangle$,

$$|\langle 0|_A (\text{anything}) |0\rangle_A|^2 \text{ changes by at most } 2\epsilon_1/3 + (\epsilon_1/3)^2 \leq \epsilon_1$$

when we replace V_{exact} with V_0. More precisely, since $V_{\text{exact}} |\psi\rangle = |0\rangle_A \otimes |\phi\rangle$,

$$\langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V_{\text{exact}} |\psi\rangle\langle\psi|_{B(A,d)} V_{\text{exact}}^\dagger] |0\rangle_A = 1.$$

Using the Fuchs-van de Graaf inequality, which relates trace distance to fidelity, and the operator norm bound, we obtain

$$\langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V_0 |\psi\rangle\langle\psi|_{B(A,d)} V_0^\dagger] |0\rangle_A \geq 1 - \epsilon_1/3.$$

Combining both sources of error, we have

$$p_0 \geq 1 - \epsilon_1/3 - \epsilon_1/3 = 1 - 2\epsilon_1/3.$$

Therefore, when the algorithm tests $V_0 \in \mathcal{N}$, the condition $p_0 \geq 1 - 2\epsilon_1/3$ is satisfied, and the algorithm outputs some circuit V (not necessarily V_0, but some circuit that passes the test).

To show that the output V is an ϵ_1-approximate local inversion, note that if V is output by the algorithm, then

$$p_V = \langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V \rho_{B(A,d)} V^\dagger] |0\rangle_A \geq 1 - 2\epsilon_1/3.$$

By the tomography error,

$$\begin{aligned}\langle 0|_A \operatorname{Tr}_{B(A,d)\setminus A}[V |\psi\rangle \langle\psi|_{B(A,d)} V^\dagger] |0\rangle_A &\geq p_V - \epsilon_1/3 \\ &\geq 1 - 2\epsilon_1/3 - \epsilon_1/3 = 1 - \epsilon_1.\end{aligned}$$

Therefore V is an ϵ_1-approximate local inversion of region A for $|\psi\rangle$, and the algorithm succeeds. □

2.3. The Replacement Process

The replacement process provides a way to apply local inversions without disturbing the global state, enabling systematic reconstruction.

Definition 209 (Replacement Process)**.** *For a region A and local inversion V, the A-**replacement process** applied to $|\psi\rangle$ consists of three steps:*

(1) Apply the local inversion: $|\psi\rangle \mapsto V|\psi\rangle$;
(2) Reset region A to $|0\rangle_A$: trace out qubits in A and replace with $|0\rangle_A$;
(3) Undo the local inversion: apply $V^\dagger$.

As a quantum channel, the replacement process is $\mathcal{R}_A^V = V^\dagger \circ \mathcal{T}_A \circ V$, where $\mathcal{T}_A$ is the reset channel on region A.

Lemma 210 (Invariance Under Replacement)**.** *For any state $|\psi\rangle$ and exact local inversion V of region A, the A-replacement process leaves $|\psi\rangle$ invariant:*

$$\mathcal{R}_A^V(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|.$$

Proof. Since V is a local inversion, $V|\psi\rangle = |0\rangle_A \otimes |\phi\rangle$ for some state $|\phi\rangle$ on the remaining $n - |A|$ qubits. After step 1, the state becomes $V|\psi\rangle = |0\rangle_A \otimes |\phi\rangle$. Step 2 traces out the qubits in region A and replaces them with $|0\rangle_A$. Since the qubits in A are already in state $|0\rangle_A$, this operation acts as the identity: tracing out $|0\rangle_A$ and replacing with $|0\rangle_A$ yields $|0\rangle_A \otimes |\phi\rangle$ again. Step 3 applies $V^\dagger$ to obtain $V^\dagger(|0\rangle_A \otimes |\phi\rangle) = V^\dagger V|\psi\rangle = |\psi\rangle$. Therefore $\mathcal{R}_A^V(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|$. □

The power of the replacement process is subtle: while it leaves $|\psi\rangle$ unchanged, it provides a circuit representation where part of the state has been replaced by known operations. This becomes clear when we examine the backward lightcone.

3. Covering Schemes and Reconstruction

To reconstruct the global state, we apply replacement processes for a carefully chosen collection of regions that satisfy geometric constraints.

3.1. Covering Schemes

Definition 211 (Covering Scheme)**.** *An (ℓ, c, d) **covering scheme** for an n-qubit system on a k-dimensional lattice is a collection of subsets*

$$\mathcal{S} = \{S_j^i : 1 \leq i \leq \ell, 1 \leq j \leq m_i\}$$

satisfying:

(i) ***Bounded size:*** *For all i, j, we have $|B(S_j^i, d)| \leq c$;*
(ii) ***Layer disjointness:*** *For each fixed layer i, the sets $\{B(S_j^i, d)\}_{j=1}^{m_i}$ are pairwise disjoint;*
(iii) ***Complete coverage:*** *For every qubit $v \in [n]$, there exists some S_j^i such that $B(\{v\}, (2\ell-1)d) \subseteq S_j^i$.*

The covering scheme organizes regions into ℓ layers. Within each layer, the regions are sufficiently separated (Condition ii) to allow parallel application of replacement processes. Across layers, the regions overlap significantly to ensure complete coverage (Condition iii).

Theorem 212 (Lattice Covering Scheme)**.** *For any k-dimensional lattice and depth d, there exists a $(k+1, c, d)$ covering scheme where*

$$c \leq \left((8k^2 + 14k + 2)d\right)^k.$$

We defer the full proof to later in this section, but first show how covering schemes enable state reconstruction.

3.2. The Reconstruction Process

Definition 213 (Reconstruction Process)**.** *Given a covering scheme $\mathcal{S} = \{S_j^i\}$ and local inversions V_j^i for each S_j^i, the **reconstruction process** Φ applies replacement processes layer by layer:*

$$\Phi = \Phi_\ell \circ \Phi_{\ell-1} \circ \cdots \circ \Phi_1,$$

where each layer operation Φ_i simultaneously applies all replacement processes for layer i:

$$\Phi_i = \prod_{j=1}^{m_i} \mathcal{R}_{S_j^i}^{V_j^i}.$$

The product notation indicates parallel composition, which is well-defined by the disjointness property [Condition (ii) of Definition 211].

By Lemma 210, if all V_j^i are exact local inversions, then $\Phi(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|$. The key insight is that despite this invariance, we can extract a circuit for preparing $|\psi\rangle$ by examining the backward lightcone.

3.3. Backward Lightcone and Reconstruction

To formalize the reconstruction, we need a precise notion of backward lightcone applicable to circuits with reset operations.

Definition 214 (Backward Lightcone)**.** *Let W be a quantum circuit consisting of unitary gates and reset operations, and let $|\phi\rangle = W|0^n\rangle$. For a subset of output qubits $A \subseteq [n]$, the **backward lightcone** $\mathcal{B}(A, W)$ is constructed as follows:*

(1) Initialize all output wires corresponding to A as active*;*
(2) Process the circuit backward from output to input:
- *If a unitary gate G has at least one active wire at its output, mark G as part of the lightcone and mark both input wires of G as active;*
- *If an active wire encounters a reset operation, stop propagating activity backward through that wire;*

(3) The backward lightcone consists of all marked gates and active input wires.

The backward lightcone captures the minimal subcircuit needed to compute the reduced density matrix on A. Reset operations act as barriers that block backward lightcone propagation. The reset operations are central to the reconstruction.

Theorem 215 (Reconstruction via Backward Lightcone)**.** *Let $|\psi\rangle = U|0^n\rangle$ where U is a depth-d circuit on a k-dimensional lattice, and let (ℓ, c, d) be a covering scheme. Let Φ be the reconstruction process using exact local inversions. Then the backward lightcone $\mathcal{B}([n], \Phi)$ of all output qubits is a depth-$(2\ell-1)d$ circuit C s.t.*

$$C|0^{n+m}\rangle = |\psi\rangle \otimes |junk\rangle,$$

where $m = \mathcal{O}(n)$ is the number of ancilla qubits.

Proof. Consider an arbitrary output qubit $v \in [n]$. We must show that the backward propagation from v terminates at reset operations within Φ and does not reach the input state $|\psi\rangle$ at the bottom. By Condition (iii) of Definition 211, there exists S_j^i in the covering scheme such that

$$B(\{v\}, (2\ell-1)d) \subseteq S_j^i.$$

Consider the backward propagation starting from qubit v at the output of Φ. As the propagation moves backward through the circuit, it can spread to neighboring qubits. In the worst case, the propagation starts at the top layer ℓ and must pass through all ℓ layers before potentially reaching the input.

Each layer Φ_k for $k > i$ consists of replacement processes that add depth at most $2d$: depth d for V_j^k and depth d for $(V_j^k)^\dagger$. Thus, after passing through layers $\ell, \ell-1, \ldots, i+1$, the backward propagation has depth at most $2(\ell-i)d \leq 2(\ell-1)d$. When the propagation reaches layer i where the replacement process for S_j^i occurs, it enters the reset operation for region S_j^i after passing through $(V_j^i)^\dagger$. At this point, the propagation has spread at most distance $(2\ell-2)d + d = (2\ell-1)d$ from the original qubit v. By the choice of S_j^i, this entire spread is contained within S_j^i, meaning all active wires encounter reset operations in $\mathcal{R}_{S_j^i}^{V_j^i}$. Therefore, the backward propagation terminates and does not reach the input state $|\psi\rangle$. Since this argument holds for every output qubit v, the complete backward lightcone $\mathcal{B}([n], \Phi)$ is contained within Φ and does not depend on the input state. The depth of this circuit is at most $(2\ell-1)d$ from the analysis above.

Finally, note that C is a unitary circuit acting on n system qubits plus m ancilla qubits (the qubits reset during the replacement processes). The output is a pure state, and the reduced density matrix on the system qubits equals $|\psi\rangle\langle\psi|$ by construction, which implies the system and ancilla are in a product state. □

Part 4

Learning Quantum Channels

CHAPTER 13

Learning Pauli Channels

- Definition of channel
- Pauli transfer matrix, Choi state, twirling
- Eigenvalues vs. error rates
- Randomized benchmarking
- Bell sampling
- Population recovery - probably only have time to sketch

Part 5

Lower Bounds

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. □

CHAPTER 15

Lower Bounds for Pauli Shadow Tomography

In this chapter, we will apply the learning tree formalism to obtain lower bounds for shadow tomography of *Pauli observables*. Concretely, the task here is to estimate $\mathrm{tr}(P\rho)$ for all n-qubit Pauli operators, given copies of unknown state ρ. Like purity testing, this turns out to be a problem for which two-copy measurements are exponentially more statistically efficient than single-copy measurements. In Section 1, we give a protocol that solves this task with two-copy measurements on only $O(n/\epsilon^4)$ copies of ρ. In Section 2.1, we give an exponential lower bound for this problem when one can only perform single-copy measurements. Intriguingly, the two-copy measurement protocol requires *adaptive* choice of measurements; in Section 3 we show that without adaptivity, there is an exponential lower bound even for two-copy measurements. Finally, in Section 4, we prove a qualitative strengthening of the lower bound from Section 2.1 demonstrating that even if one has a small amount of additional quantum memory, but not enough to perform two-copy measurements, then there is still an exponential lower bound for Pauli shadow tomography.

1. Upper bound using two-copy measurements

The protocol here involves two stages: first learning the *absolute values* $|\mathrm{tr}(P\rho)|$ and then learning the *signs* of the observable values. The first stage is based on a protocol due to [**HKP21**, **CCHL22**], and the second stage is based on a protocol due to [**CGY24**, **KGKB25**].

1.1. Learning the absolute values

The starting point for the first stage of the protocol is an idea from the chapter on learning stabilizer states, namely that even though the Pauli operators P do not commute, their tensor squares $P^{\otimes 2}$ do commute. Indeed, recall from Chapter 10 that these $\{P^{\otimes 2}\}$ can be simultaneously diagonalized, and their joint eigenbasis is the *Bell basis*.

By measuring in the Bell basis, we can thus obtain estimates for the *two-copy* observables $\mathrm{tr}(P^{\otimes 2}\rho^{\otimes 2}) = \mathrm{tr}(P\rho)^2$. This gives rise to the following guarantee:

Theorem 229 (Learning absolute values)**.** *There is a protocol which takes as input $O(n/\epsilon^4)$ copies of ρ, performs two-copy measurements in the Bell basis, and outputs estimates for all quantities $\{|\mathrm{tr}(P\rho)|\}$ up to additive error ϵ with high probability.*

Proof. Using Bell basis measurements on $O(n/\epsilon^4)$ copies, we can simultaneously obtain ϵ^2-accurate estimates for all quantities $\{\mathrm{tr}(P\rho)^2\}$ with high probability. It remains to convert these to estimates for $\{|\mathrm{tr}(P\rho)|\}$. For this, observe that for any

scalars $0 \leq x, y \leq 1$,

$$(\sqrt{x} - \sqrt{y})^2 = \frac{|x-y|^2}{(\sqrt{x}+\sqrt{y})^2} = |x-y| \cdot \frac{|x-y|}{x+y+\sqrt{2xy}} \leq |x-y|\,,$$

so if we simply output the square roots of our estimates for $\{\mathrm{tr}(P\rho)^2\}$, these will be ϵ-accurate. □

1.2. Resolving the signs

Theorem 230. *Given ϵ-accurate estimates of $\{|\mathrm{tr}(P\rho)|\}$, and given the ability to perform* single-copy *measurements on $O(n/\epsilon^4)$ additional copies of ρ, there is a protocol which estimates $\{\mathrm{tr}(P\rho)\}$ to additive error $O(\epsilon)$ with high probability.*

We will consider Algorithm 10 below.

Algorithm 10: LearnPauliSigns$(\epsilon, \{f_P\}, \rho)$

Input: Accuracy $\epsilon > 0$, estimates $\{f_P\}$ of absolute values $\{|\mathrm{tr}(P\rho)|\}$, copies of ρ

Output: Estimates $\{\hat{E}_P\}$ of true values $\{\mathrm{tr}(P\rho)\}$

1 Find a state σ such that $|f_P - |\mathrm{tr}(P\sigma)|| \leq \epsilon$ for every P.

2 Perform Bell measurements on $O(n/\epsilon^4)$ copies of $\sigma \otimes \rho$ to estimate all $\mathrm{tr}(P^{\otimes 2}(\sigma \otimes \rho)) = \mathrm{tr}(P\sigma)\mathrm{tr}(P\rho)$ to error ϵ^2 with high probability.

3 Denote each estimate of $\mathrm{tr}(P\sigma)\mathrm{tr}(P\rho)$ by g_P.

4 If $f_P < 2\epsilon$, set $\hat{E}_P = 0$. Otherwise if $f_P > 2\epsilon$, set $\hat{E}_P = g_P/\mathrm{tr}(\sigma P)$.

return $\{\hat{E}_P\}_{P \in A}$

Proof. In the first step, we can always find a σ such that $|f_P - |\mathrm{tr}(P\rho)|| \leq \epsilon$ for all P, as ρ itself satisfies the condition. Once we have an explicit description of this σ, we get exact access to $\mathrm{tr}(P\sigma)$.

In the second step, we obtain ϵ^2-accurate estimates g_P of $\mathrm{tr}(P \otimes P\ \sigma \otimes \rho)$ for every P. Observe that performing a Bell measurement on $\sigma \otimes \rho$ given explicit access to σ can be done using a single-copy measurement on ρ. If $f_P < 2\epsilon$, we set $\hat{E}_P = 0$ and this is trivially an 3ϵ-accurate estimate of $\mathrm{tr}(P\rho)$:

$$|\hat{E}_P - \mathrm{tr}(P\rho)| = |\mathrm{tr}(P\rho) - f_P + f_P| \leq |\mathrm{tr}(P\rho) - f_P| + |f_P| \leq \epsilon + 2\epsilon = 3\epsilon.$$

If $f_P > 2\epsilon$, we set $\hat{E}_P = g_P/\mathrm{tr}(P\sigma)$, in which case

$$|\hat{E}_P - \mathrm{tr}(P\rho)| = \frac{|g_P - \mathrm{tr}(P\sigma)\mathrm{tr}(P\rho)|}{|\mathrm{tr}(P\sigma)|} \leq \frac{\epsilon^2}{f_P - \epsilon} \leq \epsilon$$ □

2. Lower bound using single-copy measurements

The protocol in the previous section uses two-copy measurements. It turns out that with only single-copy measurements, it is not possible to solve Pauli shadow tomography with such a small number of copies of ρ. The proof of this will use the learning tree framework from the previous chapter.

2.1. General lower bound based on second moment

Here we give a general recipe for proving lower bounds for shadow tomography using single-copy measurements; the lower bound on the number of copies needed will depend on a certain quantity that characterizes the extent to which the observables in question approximately commute.

Consider m traceless observables $O_1, \ldots, O_m$ with $\|O_i\|_\infty = 1$ which additionally satisfy

$$O_i = -O_{i+m/2}, \text{ and all eigenvalues of } O_i \text{ are } \pm 1, \quad \forall i = 1, \ldots, m/2\,. \tag{66}$$

We will characterize the hardness of estimating these observable values using single-copy measurements via the following quantity,

$$\delta(O_1, \ldots, O_m) = \sup_{|\phi\rangle} \frac{2}{m} \sum_{i=1}^{m/2} \langle\phi| O_i |\phi\rangle^2\,, \tag{67}$$

where $\sup_{|\phi\rangle}$ is taken over all n-qubit pure states. We have the following general lower bound, which leverages a convexity argument first used in [**BCL20**]:

Theorem 231 (General shadow tomography lower bound)**.** *Let $O_1, \ldots, O_m$ be traceless observables satisfying Eq. (66). Then any learning protocol which only performs single-copy measurements on copies of unknown state ρ and outputs ϵ-accurate estimates for $\{\mathrm{tr}(O_i\rho)\}$ with high probability requires $\Omega\Big(\frac{1}{\epsilon^2 \delta(O_1,\ldots,O_m)}\Big)$ copies.*

Proof. Following the framework in Chapter 14, we will show a lower bound for the easier task of distinguishing between whether the unknown state is maximally mixed or whether it is given by

$$\rho_i := \frac{\mathrm{Id} + 3\epsilon O_i}{2^n}\,.$$

Note that if one had an algorithm for estimating $\mathrm{tr}(O_i\rho)$ for all i to error ϵ, then we could distinguish between these two ensembles.

Now consider any learning protocol using T copies of ρ, and consider its tree representation $\mathcal{T}$. For any leaf ℓ of $\mathcal{T}$, denote the sequence of edges on the path from root r to leaf ℓ by $\{e_{u_t,s_t}\}_{t=1}^T$, where $u_1 = r$ and $s_T = \ell$. If the POVM element corresponding to edge e_{u_t,s_t} is $\{2^n w_{s_t}^{u_t} |\psi_{s_t}^{u_t}\rangle \langle\psi_{s_t}^{u_t}|\}$, then the probability of reaching leaf ℓ for a given state ρ is given by

$$p^\rho(\ell) = \prod_{t=1}^T w_{s_t}^{u_t} 2^n \langle\psi_{s_t}^{u_t}|\rho|\psi_{s_t}^{u_t}\rangle\,.$$

We then have the following calculation:

$$\begin{aligned}
\frac{\left(\mathbb{E}_{i\sim[m]} p^{\rho_i}(\ell)\right)}{p^{\mathrm{Id}/2^n}(\ell)} &= \mathbb{E}_i \prod_{t=1}^{T} \left(\frac{w_s^u 2^n + 3\epsilon w_s^u 2^n \left\langle \psi_{s_t}^{u_t} \right| O_i \left| \psi_{s_t}^{u_t} \right\rangle}{w_s^u 2^n} \right) \\
&= \mathbb{E}_i \exp\left(\sum_{t=1}^{T} \log\left(1 + 3\epsilon \left\langle \psi_{s_t}^{u_t} \right| O_i \left| \psi_{s_t}^{u_t} \right\rangle\right) \right) \\
&\geq \exp\left(\sum_{t=1}^{T} \mathbb{E}_i \log\left(1 + 3\epsilon \left\langle \psi_{s_t}^{u_t} \right| O_i \left| \psi_{s_t}^{u_t} \right\rangle\right) \right) \qquad (68) \\
&\geq \exp\left(\sum_{t=1}^{T} \frac{1}{m} \sum_{i=1}^{m/2} \log\left(1 - 9\epsilon^2 \left\langle \psi_{s_t}^{u_t} \right| O_i \left| \psi_{s_t}^{u_t} \right\rangle^2\right) \right) \qquad (69) \\
&\geq \exp\left(-\sum_{t=1}^{T} \frac{18}{m} \sum_{i=1}^{m/2} \epsilon^2 \left\langle \psi_{s_t}^{u_t} \right| O_i \left| \psi_{s_t}^{u_t} \right\rangle^2 \right) \qquad (70) \\
&\geq \exp\left(-9T\epsilon^2 \delta(O_1, \ldots, O_m)\right) \qquad (71) \\
&\geq 1 - 9T\epsilon^2 \delta(O_1, \ldots, O_m)\,.
\end{aligned}$$

Eq. (68) uses Jensen's inequality, Eq. (69) uses the fact that the set of observables is closed under negation from Eq. (66), Eq. (70) uses the elementary inequality $\log(1-x) \geq -2x, \forall x \in [0, 3/4]$ which is satisfied given $\epsilon < 1/4$, and Eq. (71) uses the definition of δ in Eq. (67).

As long as $T \leq c/(\epsilon^2 \delta(O_1, \ldots, O_m))$ for sufficiently small constant $c > 0$, we obtain a sufficiently strong one-sided bound on the likelihood ratio for all leaves ℓ to deduce that one cannot distinguish between maximally mixed and $\{\rho_i\}$ with sufficiently good advantage. □

2.2. Instantiating the bound for Pauli observables

It remains to compute the $\delta(O_1, \ldots, O_m)$ quantity in the case of Pauli observables. We will take $m = 2(4^n - 1)$ and take $O_1, \ldots, O_m$ to be the set of nontrivial n-qubit Pauli observables together with their negations. This collection certainly satisfies Eq. (66). We now compute δ:

Lemma 232 (δ for Pauli observables)**.**

$$\delta\left(P_1, \ldots, P_{2(4^n-1)}\right) = \sup_{|\phi\rangle} \frac{1}{4^n - 1} \sum_{i=1}^{4^n-1} \left\langle \phi \right| P_i \left| \phi \right\rangle^2 = \frac{1}{2^n + 1}.$$

Proof. For any n-qubit pure state $|\phi\rangle$, we have

$$\begin{aligned}
\frac{1}{4^n - 1} \sum_{i=1}^{4^n-1} \left\langle \phi \right| P_i \left| \phi \right\rangle^2 &= \frac{1}{4^n - 1} \mathrm{tr}\left(\left(\sum_{i=1}^{4^n-1} P_i \otimes P_i \right) \left| \phi \right\rangle \left\langle \phi \right|^{\otimes 2} \right) \\
&= \frac{1}{4^n - 1} \mathrm{tr}\left(\left(2^n \mathrm{SWAP}_n - \mathrm{Id}^{\otimes 2n}\right) \left| \phi \right\rangle \left\langle \phi \right|^{\otimes 2} \right) \\
&= \frac{2^n - 1}{4^n - 1} = \frac{1}{2^n + 1}\,.
\end{aligned}$$

□

The second step uses the fact that the uniform distribution over the Clifford group has the same first (in fact three) moments as the Haar measure.

Combining this with Theorem 231 yields the following lower bound for Pauli shadow tomography.

Corollary 233 (Shadow tomography lower bound for Pauli observables)**.** *Any learning protocol which only performs single-copy measurements on copies of unknown state ρ and outputs ϵ-accurate estimates for $\{\mathrm{tr}(P\rho)\}$ for all Pauli operators P with high probability requires $\Omega(2^n/\epsilon^2)$ copies of ρ.*

3. Lower bound for non-adaptive two-copy measurements

Note that the learning protocol in Section 1 crucially relies on adaptivity in order to determine the appropriate measurement to perform in the second stage of the protocol where one resolves the signs. In this section, we show that this adaptivity is necessary: any protocol that performs *non-adaptive* two-copy measurements requires an exponential amount of copies of ρ:

Theorem 234. *Any protocol that estimates $\{\mathrm{tr}(P\rho)\}$ for all Pauli operators P to additive error less than 1 using* non-adaptive *two-copy measurements requires $\Omega(2^{n/2})$ copies of ρ.*

As in the previous lower bounds, we assume that the learning protocol uses rank-1 POVMs of the form $\{w_s |\psi_s\rangle\langle\psi_s|\}_s$ for for $2n$-qubit pure states $\{|\psi_s\rangle\}$ and nonnegative weights w_s with $\sum_s w_s = 2^{2n}$.

To prove the nonadaptive lower bound in Theorem 234, we consider a distinguishing task in which we are given access to copies of an unknown n-qubit quantum state ρ and want to distinguish between two cases:

- ρ is sampled from $\{\rho_a^+ = \frac{I+P_a}{2^n}\}_{a\in[4^n-1]}$ uniformly over non-identity Paulis P_a
- ρ is sampled from $\{\rho_a^- = \frac{I-P_a}{2^n}\}_{a\in[4^n-1]}$ uniformly over non-identity Paulis P_a.

If we have a protocol that can estimate all Pauli observables to additive error less than 1 with high probability, then we can solve this distinguishing task using the same protocol. It thus suffices to prove a lower bound on the number of copies of ρ needed to solve the distinguishing problem with high probability. Note that this distinguishing problem is different from the distinguishing problem we considered for single-copy measurements; there we wished to distinguish between maximally mixed versus states of the form $(I \pm P_a)/2^n$, whereas here we wish to distinguish between *two different* ensembles over nontrivial projectors.

Given an $a \in \{1, ..., 4^n - 1\}$, and given the t-th POVM $\mathcal{M}_t = \{w_s^t |\psi_s^t\rangle\langle\psi_s^t|\}$, we denote by $p_{a,t}^+$ and $p_{a,t}^-$ the probability distributions over outcomes s_t from measuring ρ_a^+ and ρ_a^- respectively. We also denote by p_a^+ and p_a^- the probability distribution over the T measurement outcomes $s_1, ..., s_T$ for measuring ρ_a^+ and ρ_a^-.

Our goal is to show a bound on the total variation distance between $\mathbb{E}_a[p_a^+]$ and $\mathbb{E}_a[p_a^-]$ for any nonadaptive sequence of two-copy POVM measurements $\{\mathcal{M}_t\}_{t=1}^T$, i.e.,

$$\mathrm{TV}(\mathbb{E}_a[p_a^+], \mathbb{E}_a[p_a^-]) \leq o(1)\,,$$

as Le Cam's lemma would then imply that any nonadaptive protocols using at most T two-copy measurements cannot distinguish between these two cases with high probability.

Note that for a sequence of nonadaptive measurements, the probability distribution p_a^+ and p_a^- can be written as a tensor product of the probability distributions for each individual measurement $p_{a,t}^+$ and $p_{a,t}^-$. We thus have

$$\begin{aligned}\mathrm{TV}(\mathbb{E}_a[p_a^+], \mathbb{E}_a[p_a^-]) &\leq \mathbb{E}_a[\mathrm{TV}(p_a^+, p_a^-)] \\ &= \mathbb{E}_a\Big[\mathrm{TV}\Big(\bigotimes_{t=1}^T p_{a,t}^+, \bigotimes_{t=1}^T p_{a,t}^-\Big)\Big] \\ &\leq \sum_{t=1}^T \mathbb{E}_a[\mathrm{TV}(p_{a,t}^+, p_{a,t}^-)] \\ &\leq T \max_{\text{two copy } \mathcal{M}_t} \mathbb{E}_a[\mathrm{TV}(p_{a,t}^+, p_{a,t}^-)]\,.\end{aligned}$$

Note that this sequence of inequalities crucially uses the fact that the measurements are non-adaptive.

It thus suffices to bound the average total variation distance for a *single* two-copy measurement. Define

$$\mathbf{N} \triangleq \left(2^n(\mathrm{SWAP}_{1,3} + \mathrm{SWAP}_{1,4} + \mathrm{SWAP}_{2,3} + \mathrm{SWAP}_{2,4}) - 4I^{\otimes 4}\right)\,.$$

For any such $\mathcal{M}_t$, we have

$$\begin{aligned}&\mathbb{E}_a[\mathrm{TV}(p_{a,t}^+, p_{a,t}^-)] \\ &\quad= \mathbb{E}_a\Big[\frac{1}{2}\sum_s \Big|\mathrm{tr}\Big[\Big(\frac{I+P_a}{2^n}\Big)^{\otimes 2} w_s^t \left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|\Big] - \mathrm{tr}\Big[\Big(\frac{I-P_a}{2^n}\Big)^{\otimes 2} w_s^t \left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|\Big]\Big|\Big] \\ &\quad= \mathbb{E}_a\Big[\frac{1}{2^{2n}}\sum_s w_s^t \left|\mathrm{tr}[(I\otimes P_a + P_a\otimes I)\left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|]\right|\Big] \\ &\quad\leq \frac{1}{2^{2n}}\mathbb{E}_a\Big[\sqrt{\sum_s w_s^t \,\mathrm{tr}^2[(I\otimes P_a + P_a\otimes I)\left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|]\cdot\sum_s w_s^t}\Big] \\ &\quad\leq \frac{1}{2^n}\sqrt{\mathbb{E}_a\Big[\sum_s w_s^t \,\mathrm{tr}^2\left[(I\otimes P_a + P_a\otimes I)\left|\psi_s^t\right\rangle\left\langle\psi_s^t\right|\right]\Big]} \\ &\quad= \frac{1}{2^n}\sqrt{\mathbb{E}_a\Big[\sum_s w_s^t \left\langle\psi_s^t\right|\left\langle\psi_s^t\right|(I\otimes P_a + P_a\otimes I)^{\otimes 2}\left|\psi_s^t\right\rangle\left|\psi_s^t\right\rangle\Big]} \\ &\quad= \frac{1}{2^n}\sqrt{\frac{1}{2^{2n}-1}\sum_s w_s^t \left\langle\psi_s^t\right|\left\langle\psi_s^t\right|\mathbf{N}\left|\psi_s^t\right\rangle\left|\psi_s^t\right\rangle} \\ &\quad\leq \frac{1}{2^n}\sqrt{\frac{1}{2^{2n}-1}\sum_s w_s^t(4\cdot 2^n - 4)} \\ &\quad= \frac{1}{2^n}\sqrt{\frac{4\cdot 2^{2n}}{2^n+1}} = O\Big(\frac{1}{2^{n/2}}\Big)\end{aligned}$$

where the third step follows by Cauchy-Schwarz inequality, the fourth step follows from the fact that $\sum_s w_s^t = 2^{2n}$ and Jensen's inequality, and the sixth step follows from the two-design property of the Clifford group (here $\mathrm{SWAP}_{i,j}$ denotes the SWAP operator on the i-copy and the j-th copy). Therefore, we have

$$\mathrm{TV}(\mathbb{E}_a[p_a^+], \mathbb{E}_a[p_a^-]) \leq O(T\cdot 2^{-n/2}).$$

This indicates that any nonadaptive protocols with $T \leq o(2^{n/2})$ two-copy measurement can not solve this distinguishing task with high probability, which yields the $\Omega(2^{n/2})$ lower bound in Theorem 234 for two-copy nonadaptive protocols for solving Pauli shadow tomography.

4. Lower bound for protocols with limited quantum memory

In this section, we consider an extension of the setting of Section 2.1 to one in which the learner has access to a nonzero but small amount of additional quantum memory - more than is needed to perform single-copy measurements, but not enough to perform two-copy measurements. We prove that the exponential lower bound of Section 2.1 persists in this setting.

First, we define the model that we work with.

Definition 235 (Learning with bounded quantum memory)**.**

CHAPTER 16

Tools from Probability Theory

[sitan: if we move this to the lower bounds unit, this lecture could cover:

- Hypothesis testing: Le Cam and Fano
- Example: classical identity testing (Le Cam)
- Example: distribution learning (Fano)

]

CHAPTER 17

State Tomography with Unentangled Measurements

- Full proof of unentangled lower bound

Bibliography

[AAA+13] Junaid Aasi, Joan Abadie, BP Abbott, Richard Abbott, TD Abbott, MR Abernathy, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light. *Nature Photonics*, 7(8):613–619, 2013. 12

[AAKS21] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, 2021. 151

[Aar04] Scott Aaronson. Is quantum mechanics an island in theoryspace? *quant-ph/0401062*, 2004. 29

[Aar18] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 325–338, 2018. 97

[ACH+18] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. *Advances in neural information processing systems*, 31, 2018. 111

[ACQ22] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(1):887, 2022. 193

[BC17] Jacob C Bridgeman and Christopher T Chubb. Hand-waving and interpretive dance: an introductory course on tensor networks. *Journal of physics A: Mathematical and theoretical*, 50(22):223001, 2017. 55

[BCL20] Sebastien Bubeck, Sitan Chen, and Jerry Li. Entanglement is necessary for optimal quantum property testing. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 692–703. IEEE, 2020. 193, 209

[BCS+25] Jacob Beckey, Luke Coffman, Ariel Shlosberg, Louis Schatzki, and Felix Leditzky. Product testing with single-copy measurements. *arXiv:2510.07820*, 2025. 204

[BLMT24] Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Learning quantum hamiltonians at any temperature in polynomial time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1470–1477, 2024. 151

[CAN25] Chi-Fang Chen, Anurag Anshu, and Quynh T Nguyen. Learning quantum gibbs states locally and efficiently. *arXiv preprint arXiv:2504.02706*, 2025. 151, 155, 160, 161

[CCHL22] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585. IEEE, 2022. 55, 193, 204, 207

[CGY24] Sitan Chen, Weiyuan Gong, and Qi Ye. Optimal tradeoffs for estimating pauli observables. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1086–1105. IEEE, 2024. 207

[CGYZ25] Sitan Chen, Weiyuan Gong, Qi Ye, and Zhihan Zhang. Stabilizer bootstrapping: A recipe for efficient agnostic tomography and magic estimation. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 429–438, 2025. 173, 176

[CHL+23] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning? In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 391–404. IEEE, 2023. 75

[CS+04] Imre Csiszár, Paul C Shields, et al. Information theory and statistics: A tutorial. *Foundations and Trends® in Communications and Information Theory*, 1(4):417–528, 2004. 89

[CW20] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical review letters*, 124(10):100401, 2020. 97

[Dir81] Paul Adrien Maurice Dirac. *The principles of quantum mechanics*. Number 27. Oxford university press, 1981. 33

[EGH+09] Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina. Introduction to representation theory. *arXiv preprint arXiv:0901.0827*, 2009. 91

[EHF19] Tim J Evans, Robin Harper, and Steven T Flammia. Scalable bayesian hamiltonian learning. *arXiv:1912.07636*, 2019. 97

[FH13] William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013. 85, 89

[GIKL24] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Improved stabilizer estimation via bell difference sampling. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1352–1363, 2024. 176

[GKKT20] Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020. 75

[GNW21] David Gross, Sepehr Nezami, and Michael Walter. Schur–weyl duality for the clifford group with applications: Property testing, a robust hudson theorem, and de finetti representations. *Communications in Mathematical Physics*, 385(3):1325–1393, 2021. 177

[GOT98] D GOTTESMAN. The heisenberg representation of quantum computers. In *Proc. XXII International Colloquium on Group Theoretical Methods in Physics*, pages 32–43, 1998. 163

[GP88] Robert Grone and Stephen Pierce. Permanental inequalities for correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 9(2):194–201, 1988. 204

[Gri25] Darij Grinberg. An introduction to the symmetric group algebra. *arXiv preprint arXiv:2507.20706*, 2025. 89

[HHR+05] Hartmut Häffner, Wolfgang Hänsel, CF Roos, Jan Benhelm, D Chek-al Kar, M Chwalla, T Körber, UD Rapol, M Riebe, PO Schmidt, et al. Scalable multiparticle entanglement of trapped ions. *Nature*, 438(7068):643–646, 2005. 74

[HK19] Hsin-Yuan Huang and Richard Kueng. Predicting features of quantum systems from very few measurements. *arXiv preprint arXiv:1908.08909*, 2019. 97

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 363–390. IEEE, 2023. 13

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020. 97

[HKP21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021. 193, 207

[HKT22] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum hamiltonians from high-temperature gibbs states. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 135–146. IEEE, 2022. 131, 137, 138, 139, 140, 141, 149

[KGKB25] Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. Triply efficient shadow tomography. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 914–946. SIAM, 2025. 207

[KRT14] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *ArXiv*, abs/1410.6913, 2014. 97

[Lan11] Joseph M Landsberg. *Tensors: geometry and applications: geometry and applications*, volume 128. American Mathematical Soc., 2011. 55

[Mah18] Urmila Mahadev. Classical verification of quantum computations. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 259–267. IEEE, 2018. 40

[Mon07] Ashley Montanaro. Learning stabilizer states by bell sampling. *Proceedings of the Royal Society A*, 463(2088), 2007. 163, 168

[Nar24] Shyam Narayanan. Improved algorithms for learning quantum hamiltonians, via flat polynomials. *arXiv preprint arXiv:2407.04540*, 2024. 151

[VDN10] Maarten Van Den Nes. Classical simulation of quantum computation, the gottesman-knill theorem, and slightly beyond. *Quantum Information & Computation*, 10(3):258–271, 2010. 164

[VN18] John Von Neumann. *Mathematical foundations of quantum mechanics: New edition*. Princeton university press, 2018. 33

[WA23] Dominik S Wild and Álvaro M Alhambra. Classical simulation of short-time quantum dynamics. *PRX Quantum*, 4(2):020340, 2023. 138

[WB24] Adam Bene Watts and John Bostanci. Quantum event learning and gentle random measurements. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pages 97–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024. 116

[Web15] Zak Webb. The clifford group forms a unitary 3-design. *arXiv preprint arXiv:1510.02769*, 2015. 105, 107

[Wri16] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016. 97

[Yu97] Bin Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435. Springer, 1997. 199

[Zhu17] Huangjun Zhu. Multiqubit clifford groups are unitary 3-designs. *Physical Review A*, 96(6):062336, 2017. 107

Index