Improved bounds for state certification, separability testing, and shadow tomography

Costin Badescu

CMU-CS-25-148

December 2025

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Ryan O'Donnell (Chair)
Aayush Jain
David Woodruff
John Wright (University of California, Berkeley)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science.

This research was sponsored by: the National Science Foundation under award numbers CCF1618679, CCF1909310 and CCF1717606; and the University of Sydney under award number G207324.

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: quantum computation, sample complexity, property testing

Abstract. We present improved sample complexity bounds for three fundamental quantum information tasks: state certification, separability testing, and shadow tomography. Given measurement access to n identical copies of an unknown quantum state ρ, we consider:

i. *State certification*: The task of verifying ρ is equal to a reference state σ or at least ϵ-far in trace distance. We present a testing algorithm for state certification that uses $O(d/\epsilon^2)$ copies of ρ.

ii. *Separability testing*: For a bipartite state ρ on a d^2-dimensional system, we prove a lower bound of $\Omega(d^2/\epsilon^2)$ copies are necessary to distinguish separability from being ϵ-far in trace distance from the set of all separable states.

iii. *Shadow tomography*: The problem of estimating the expectation values $\mathrm{tr}(\rho A_i)$ for m observables $A_1, \ldots, A_m$ to $\pm\epsilon$ accuracy. We present an algorithm that accomplishes this with $O(\log^2(m)\log(d)/\epsilon^4)$ copies, which simultaneously achieves the best known dependence on each parameter m, d, and ϵ.

Acknowledgments

I would like to thank my advisor, Ryan O'Donnell, whose guidance was invaluable throughout my degree. I am sincerely grateful for the time and expertise he so kindly and generously shared with me, and for his great patience and understanding.

I would also like to thank John Wright for sharing his insights with me early on.

My stay in Pittsburgh would not have been as enjoyable without the company of the friends I made there — thank you to all of you.

Finally, none of this would have been possible without the unconditional support of my parents, Felicia and Ovidiu; my wife, Elena; and my brother Victor.

Contents

CHAPTER 1

Introduction

The axioms of quantum mechanics postulate that the state of a physical system can be exactly represented by a mathematical object called a *density operator* and that there are two types of physically realizable state transitions: *unitary transformations* and *quantum measurements*.

For a d-dimensional quantum system, a density operator ρ is a $d \times d$ complex matrix satisfying specific properties. A unitary transformation is a reversible state transition induced by a $d \times d$ unitary matrix U mapping the density operator ρ to the new state $U\rho U^\dagger$. Since $U^{-1} = U^\dagger$, the original state can always be recovered from ρ' by applying the reverse unitary transformation: $\rho' \mapsto U^\dagger \rho' U = \rho$. Thus, no information about ρ is lost under unitary evolution. However, quantum systems are inherently *opaque* — a classical observer cannot extract any information about an unknown state ρ using unitary transformations alone.

Quantum measurements, the other type of transition, provide the only means for a classical (i.e. nonquantum) observer to interact with and extract information from a quantum system. Unlike unitary transformations, measurements are irreversible. The act of measurement causes the system to transition to a new state, a phenomenon known as *wave function collapse*.

A quantum measurement $\mathcal{M}$ defines a set of possible measurement outcomes, say $[k] = \{1, \dots, k\}$. For a given state ρ, $\mathcal{M}$ induces a probability distribution over $[k]$, where different states may yield different distributions. When ρ is measured and a random outcome $\boldsymbol{i} \in [k]$ is observed, the state collapses to a new state ρ', which depends on ρ, $\mathcal{M}$, and $\boldsymbol{i}$.

The pair $(\rho, \mathcal{M})$ defines a probability distribution that an experimenter can sample from. Crucially, quantum mechanics dictates that this is the *only* way an observer can gain information about the system.

This framework raises a fundamental challenge in estimation tasks. After measuring ρ and observing outcome $\boldsymbol{i}$, the state collapses to ρ', modifying the original probability distribution to $(\rho', \mathcal{M})$ for future measurements and possibly incurring information loss. To obtain another sample from the initial distribution $(\rho, \mathcal{M})$, one must prepare a fresh copy of ρ and measure it again.

One might wonder whether quantum states can be cloned to circumvent this issue. However, the well-known quantum *no-cloning theorem* [59] prohibits any physical process from perfectly copying an unknown quantum state. Thus, copies of ρ are finite resources that are consumed upon measurement.

This limitation motivates the central challenge: solving estimation problems while minimizing the number of state copies used. Statistical inference typically requires multiple samples, but preparing identical quantum states is resource-intensive.

This thesis investigates the efficient extraction of statistical information from measurements of unknown quantum states, where efficiency is quantified by the number of state copies required.

We focus on three fundamental problems: *state certification*, *separability testing*, and *shadow tomography*.

1.1. State certification

Let σ denote a fixed known quantum state represented as a $d \times d$ density matrix. Given measurement access to copies of an unknown state ρ, the *quantum state certification* task is to with high probability distinguish between the case $\rho \approx \sigma$ and $d_*(\rho, \sigma) > \epsilon$, where d_* is a chosen distance measure between quantum states and ϵ is a proximity parameter.

State certification addresses the problem of verifying that a quantum device works as intended. In practice, σ may be the predicted theoretical output of a quantum gate on a quantum computer and ρ may be the actual output, which could differ from σ due to operational noise and imperfections in the construction. Alternatively, ρ could be the state of a quantum system obtained in a lab experiment and σ could be the state predicted by theory. The problem is to minimize the number of copies of ρ needed to certify that ρ is close to σ or ϵ-far from σ. This constitutes the quantum analog of classical identity testing for probability distributions, extended to the noncommutative setting of density matrices.

A straightforward approach to the state certification problem is to perform *quantum state tomography* on the n copies of ρ to learn an estimate $\widehat{\rho}$ and then check if the estimate is close to σ. Since learning to ϵ-accuracy in trace distance requires $\Theta(d^2/\epsilon^2)$ copies of ρ [**28**, **46**], this approach yields an algorithm for quantum state certification with respect to trace distance using $O(d^2/\epsilon^2)$ copies of ρ. Yet, learning a matrix involving $\Omega(d^2)$ parameters just to answer a yes-no question seems inefficient, raising the question of whether more efficient methods exist.

In Chapter 3, based on joint work with Ryan O'Donnell and John Wright [**12**], we confirm that such methods exist. Specifically, we show that $O(d/\epsilon^2)$ copies of ρ are sufficient for state certification with respect to trace distance, denoted $d_{\mathrm{tr}}(\_,\_)$:

THEOREM. *There is an algorithm that, given $n = O(d/\epsilon^2)$ copies each of unknown d-dimensional mixed quantum states ρ and σ, with high probability distinguishes between $\rho = \sigma$ or $d_{\mathrm{tr}}(\rho, \sigma) > \epsilon$.*

Our algorithm also works in the case where both ρ and σ are unknown. Furthermore, if either of the two states is of low rank, we obtain the following improvement:

THEOREM. *If either ρ or σ is close to having rank k, in the sense that the sum of its largest k eigenvalues is at least $1 - \delta$, then there is an algorithm that, given $n = O(k/\epsilon^2)$ copies of each ρ and σ, with high probability distinguishes between $d_{\mathrm{HS}}(\rho, \sigma) \leq 0.58\epsilon/\sqrt{k}$ or $d_{\mathrm{tr}}(\rho, \sigma) > \epsilon + \delta$.*

We also study the state certification problem with respect to Bures χ^2-divergence, denoted $d_{\chi^2}(\_,\_)$, a quantum counterpart to classical χ^2-divergence, and prove that $O(d/\epsilon^2)$ copies are sufficient in this case as well:

THEOREM. *Let σ be a d-dimensional known quantum state with smallest eigenvalue at least $c\epsilon^2/d$ for some $c > 0$. There is an algorithm that, given $n = O(d/\epsilon^2)$ copies of ρ, with high probability distinguishes between $d_{\chi^2}(\rho, \sigma) \leq 0.99\epsilon^2$ and $d_{\chi^2}(\rho, \sigma) > \epsilon^2$.*

The result above leads to a state certification algorithm with respect to fidelity, denoted $\mathrm{F}(\_,\_)$, perhaps the second most important dissimilarity measure in quantum information science after trace distance:

COROLLARY. *Let σ be a known d-dimensional quantum state. There is an algorithm that, given $n = O(d/\epsilon)$ copies of ρ, with high probability distinguishes between $d_{\chi^2}(\rho, \sigma) \leq 0.49\epsilon$ and $\mathrm{F}(\rho, \sigma) < 1 - \epsilon$.*

Since fidelity measures the "overlap" between two states, $\mathrm{F}(\rho, \sigma) \approx 1$ if ρ is close to σ. For a worst-case σ, our results have optimal copy complexity, up to constant factors, as a consequence of a lower bound due to O'Donnell–Wright [45] which implies that state certification when $\sigma = \frac{\mathbf{1}}{d}$ is the *maximally mixed state* requires $\Omega(d/\epsilon^2)$ copies of ρ. On an instance-by-instance basis, a recent result of O'Donnell–Wadhwa [48] gives nearly instance-optimal bounds for the quantum state certification problem using a novel quantum analog of the Ingster–Suslina method [35].

1.1.1. Related work. When σ is a known pure state, it is a folk result that $\Theta(1/\epsilon)$ copies are necessary and sufficient for state certification with respect to fidelity (see, e.g. [41, Section 4.1.1]). This result also implies that $\Theta(1/\epsilon^2)$ copies are necessary and sufficient for state certification in trace distance. Furthermore, if σ is assumed to be a pure state *and* the class of measurements is restricted, e.g. to include only Pauli measurements, then it has been shown [5, 15, 19] that $O(d/\epsilon^2)$ copies are sufficient to solve the certification problem with respect to fidelity. At the other end of the spectrum, the state certification problem when $\sigma = \frac{\mathbf{1}}{d}$ is the maximally mixed state was studied in [45], where it is shown that $\Theta(d/\epsilon^2)$ copies of ρ are necessary and sufficient for state certification in trace distance. To the best of our knowledge, our work in [12] is the first study of state certification for general mixed states σ.

1.2. Separability testing

A bipartite quantum state ϱ on $\mathbb{C}^d \otimes \mathbb{C}^d$ is said to be *separable* if it can be written as a convex combination of product states, meaning states of the form $\rho_1 \otimes \rho_2$ where ρ_1 and ρ_2 are quantum states on $\mathbb{C}^d$. Separable quantum states are precisely those states which do not exhibit any form of quantum entanglement. These are the only states that can be prepared by separated parties who can only share classical information. Understanding the general structure and properties of the set of separable states in higher dimensions is a difficult problem and is the subject of much ongoing research. For instance, deciding whether a given $d^2 \times d^2$ matrix represents a separable state on $\mathbb{C}^d \otimes \mathbb{C}^d$ – also known as the *separability problem* in the quantum literature – is NP-hard [26]. In this work, we study the following property testing version of the separability problem:

> Given unrestricted measurement access to n copies of an unknown quantum state ϱ on $\mathbb{C}^d \otimes \mathbb{C}^d$, decide with high probability if ϱ is separable or ϵ-far from all separable states in trace distance.

The ultimate goal is to determine the number of copies of ϱ that is necessary and sufficient to solve the problem, up to constant factors, as a function of d and ϵ.

By estimating (i.e. fully learning) ϱ using recent algorithms for quantum state tomography [28, 47] and checking if the estimate is sufficiently close to a separable state, this problem can be solved using $O(d^4/\epsilon^2)$ copies of ϱ.

In Chapter 4, based on joint work with Ryan O'Donnell [7], we prove the following lower bound:

Theorem. *Let $\mathcal{P}$ denote the set of separable states on $(\mathbb{C}^d)^{\otimes 2}$ and suppose $\epsilon = \Omega(1/\sqrt{d})$. Any algorithm that, given measurement access to n copies of an unknown state ρ, with high probability distinguishes between $\rho \in \mathcal{P}$ and ρ being ϵ-far from $\mathcal{P}$ in trace distance, requires $n = \Omega(d^2/\epsilon^2)$ copies of ρ.*

Closing the gap between the known bounds seems like a difficult problem and while we have no particularly strong feeling about whether the tight bound is the upper bound, the lower bound, or something in between, at least one paper [6] contains some evidence that $\widetilde{\Theta}(d^3)$ might be the true complexity for constant ϵ.

Given the difficulty of closing the gap, we have sought a classical analogue of the separability testing problem to try as a first step. Analogies between quantum states and classical probability distributions have proven to be a helpful source of inspiration throughout quantum theory. Unfortunately, entanglement is understood to be a purely quantum phenomenon; every finitely-supported discrete distribution can be expressed as a convex combination of product point distributions, so there are no "entangled" distributions. But motivated by the characterization of separable quantum states using symmetric extensions and the quantum de Finetti theorem [18], we propose as a kind of analogue the study of mixtures of i.i.d. bivariate distributions, which arise in the classical de Finetti theorem. Doherty et al. [18] used the quantum de Finetti theorem to show that a quantum state ϱ on $\mathbb{C}^d \otimes \mathbb{C}^d$ is separable (i.e. a mixture of product states) if and only if ϱ has a symmetric extension to $\mathbb{C}^d \otimes (\mathbb{C}^d)^{\otimes k}$ for any positive integer k. Somewhat analogously, the classical de Finetti theorem states that a sequence of real random variables is a mixture of i.i.d. sequences of random variables if and only if it is exchangeable [16].

We call distributions which are mixtures of i.i.d. bivariate distributions *completely positive*, due to their connection with completely positive matrices. We show that:

Theorem. *Given sample access to an unknown distribution p over $[d] \times [d]$, at least $\Omega(d/\epsilon^2)$ samples are necessary to decide with high probability if p is completely positive or ϵ-far from all completely positive distributions in total variation distance.*

Our proof is a generalization of Paninski's lower bound for testing if a distribution is uniform [49].

1.2.1. Related work. The property testing version of the separability problem, as defined above, appears in [42], where a lower bound of $\Omega(d^2)$ is proven for constant ϵ. As in [42], our proof also reduces the problem of testing if a state is separable to the problem of testing if a state is the maximally mixed state. However, we do not pass through the notion of entanglement of formation, as [42] does, and instead rely on results about the convex structure of the set of separable states. This approach yields a more direct proof that certain random states are with high probability far from separable, which allows us to take advantage of a lower bound from [45].

We believe that the separability testing problem has seen further study, but that there has been a lack of results due to its difficulty. There *is* a very extensive literature on the subject of entanglement detection (see e.g. [25, 33]), which is concerned with establishing different criteria for detecting or verifying entanglement. However, it is not obvious how these results can be applied in the property testing setting. In particular, few of these criteria are specifically concerned with states that are far from separable in trace distance and many only apply to certain restricted classes of quantum states.

As regards our classical analogue — testing if a bipartite distribution is completely positive (mixture of i.i.d.) — we are not aware of previous work in the literature. The proof of our $\Omega(d/\epsilon^2)$ lower bound is inspired by, and generalizes, Paninski's lower bound for testing if a distribution is uniform [49].

1.3. Shadow tomography

A *quantum event* E is a matrix representing an outcome of a quantum measurement which can be assigned a probability $\mathrm{tr}(\rho E)$ of occurring with respect to a quantum state ρ. Given m quantum events $E_1, \dots, E_m$ on $\mathbb{C}^d$ and measurement access to copies of an unknown d-dimensional quantum state ρ, the *shadow tomography* problem is to w.h.p. estimate the probabilities of each of the m events, $\mathrm{tr}(\rho E_1), \dots, \mathrm{tr}(\rho E_m)$, to $\pm\epsilon$ accuracy using as few copies of ρ as possible.

There are two obvious approaches to shadow tomography. By performing quantum state tomography on $O(d^2/\epsilon^2)$ copies of ρ, one can learn an ϵ-close estimate $\widehat{\rho}$ and output the probabilities $\mathrm{tr}(\widehat{\rho}E_1), \dots, \mathrm{tr}(\widehat{\rho}E_m)$. Alternatively, for each event E_i, with $i = 1, \dots, m$, one can repeatedly measure $O(1/\epsilon^2)$ copies of ρ with the binary measurement $(E_i, \overline{E}_i)$, thereby sampling from a Bernoulli distribution with $p = \mathrm{tr}(\rho E_i)$, and output an ϵ-accurate estimate of $\mathrm{tr}(\rho E_i)$, using $O(m/\epsilon^2)$ copies overall. Both of these approaches require a number of copies of ρ that is polynomial in m or d.

Aaronson introduced the shadow tomography problem in [1] and proved that there exists an algorithm for shadow tomography that requires only $\widetilde{O}(\log^4(m)\log(d)/\epsilon^4)$ copies, an exponential improvement over the naive approaches.

In Chapter 5, based on joint work with Ryan O'Donnell [11], we improve quadratically the dependence on $\log(m)$:

Theorem. *Given m quantum events $A_1, \dots, A_m$ and measurement access to n copies of an unknown d-dimensional quantum state ρ, there exists an algorithm that, with high probability, produces estimates $\widehat{\mu}_1, \dots \widehat{\mu}_m$ such that $|\widehat{\mu}_i - \mathrm{tr}(\rho A_i)| \leq \epsilon$ for all $i = 1, \dots, m$, using $n = \widetilde{O}(\log^2(m)\log(d)/\epsilon^4)$ copies of ρ.*

Furthermore, this algorithm is online*: the events A_i can be presented one-by-one to the algorithm and the corresponding estimate $\widehat{\mu}_i$ is produced before the next event A_{i+1} is determined, so the sequence of events can be chosen adversarially.*

In [1], Aaronson uses a combination of a *gentle search lemma* and a quantum state learning algorithm to prove their result. Roughly speaking, the learning algorithm maintains an estimate $\widehat{\rho}$ of the state ρ, starting with $\widehat{\rho}$ being the maximally mixed state, and iteratively improves the estimate $\widehat{\rho}$ by finding a "bad" event A_i whose probability estimate $\mathrm{tr}(\widehat{\rho}A_i)$ is off by $O(\epsilon)$ and postselecting on A_i or its complement, i.e. conditioning the state $\widehat{\rho}$ on the occurrence of the event A_i or $\mathbf{1} - A_i$. To find such a bad event, the gentle search lemma uses binary search combined with a result of Harrow–Lin–Montanaro [29] which allows one to test if a collection of events contains a bad event.

In a follow-up work, Aaronson–Rothblum [4] introduce a new *Quantum Private Multiplicative Weights* (QPMW) algorithm for shadow tomography that uses $\widetilde{O}(\log^2(m)\log^2(d)/\epsilon^8)$ copies. The QPMW algorithm features an online search routine that replaces the binary search method used in the gentle search lemma (which, despite the name, is not gentle per se) with a linear search that is actually gentle, in the sense that each quantum measurement in the search causes a small amount of damage to the measured state, so the copies of ρ that are prepared initially can be *reused.*

Inspired by the gentle linear search in the QPMW algorithm of [4], we introduce an online *quantum threshold search* routine with improved efficiency:

THEOREM. *Given m quantum events $A_1, \dots, A_m$, probability thresholds $\theta_1, \dots, \theta_m \in [0,1]$, and measurement access to n copies of an unknown d-dimensional quantum state ρ, there exists an algorithm that, with high probability, either finds an event A_j with $\mathrm{tr}(\rho A_j) > \theta_j - \epsilon$ or correctly outputs "$\mathrm{tr}(\rho A_i) \le \theta_i$ for all $i \in [m]$," using $\widetilde{O}(\log^2(m)/\epsilon^2)$ copies of ρ.*

Furthermore, this algorithm is online*: the event-threshold pairs (A_i, θ_i) can be presented one-by-one to the algorithm and chosen adversarially. Upon being presented with a pair (A_j, θ_j), the algorithm will either "pass" or halt and output "$\mathrm{tr}(\rho A_j) > \theta_j - \epsilon$." If the algorithm passes on all inputs, then it shall output "$\mathrm{tr}(\rho A_i) \le \theta_i$ for all $i \in [m]$."*

Combining this algorithm with the online state learning algorithm of Aaronson–Chen–Hazan–Kale–Nayak [3], we obtain our shadow tomography result above.

Using our quantum threshold search result, we also give improved bounds for the following hypothesis selection problem:

THEOREM. *Given m known quantum hypothesis states $\sigma_1, \dots, \sigma_m$ and measurement access to n copies of an unknown d-dimensional quantum state ρ, there exists an algorithm that, with high probability, selects a hypothesis σ_k such that*

$$d_{\mathrm{tr}}(\rho, \sigma_k) \le 3.01 \cdot \min_i \{d_{\mathrm{tr}}(\rho, \sigma_i)\} + \epsilon,$$

using $\min\{\widetilde{O}(\log^2(m)\log(d)/\epsilon^4), \widetilde{O}(\log^3(m)/\epsilon^2)\}$ *copies of ρ.*

1.3.1. Related work. Aaronson introduced the shadow tomography problem in [1] and proved that $O(\log^4(m)\log(d)/\epsilon^4)$ copies are sufficient to solve the problem. As previously stated, Aaronson's algorithm combines a quantum state learning algorithm with a search routine called the gentle search lemma. The state learning algorithm used in [1] is improved further in a follow-up work of Aaronson–Chen–Hazan–Kale–Nayak [3] (see Theorem 5.3.1). The gentle search lemma is replaced with a gentle linear search in the work of Aaronson–Rothblum [4] using techniques inspired by differential privacy. The algorithm for shadow tomography presented in [4], which uses $\widetilde{O}(\log^2(m)\log^2(d)/\epsilon^8)$ copies, improves the dependence on m at the cost of a worse dependence on the other parameters, d and ϵ. In [11], Badescu–O'Donnell obtain an algorithm for shadow tomography, detailed in Chapter 5, that uses $\widetilde{O}(\log^2(m)\log(d)/\epsilon^4)$ copies of the unknown state. A different algorithm for quantum threshold search with a matching bound of $\widetilde{O}(\log^2(m)/\epsilon^2)$ was obtained subsequently by Bene Watts–Bostanci [58]. Their technique does not require injecting noise into measurements, as we do, but their algorithm is not online.

To the best of my knowledge, our result remains at the time of this writing the best upper bound on the copy complexity of general shadow tomography. The best known lower bound for general shadow tomography is $\Omega(\log(m)/\epsilon^2)$, proved in Aaronson's first paper on the subject [1].

Multiple variations of the shadow tomography problem arise by either limiting the types of quantum events given or by constraining the possible states of the unknown quantum system.

If the Hilbert–Schmidt norm of the events is bounded by a universal constant, i.e. there exists a universal constant C such that $\mathrm{tr}(A_i^2) \leq C$ for all $i = 1, \ldots, m$, then [34, 60] give an algorithm for shadow tomography that only uses $O(\log(m)/\epsilon^2)$ copies with a matching the lower bound.

The algorithms of [1, 4, 11] rely on quantum measurements that are applied to all n copies of the unknown quantum state ρ *simultaneously*, viz. the algorithms measure the multipartite state $\rho^{\otimes n}$ directly. If one only allows measurements on single copies of ρ at a time, then Chen–Cotler–Huang–Li [13] proved that any algorithm subject to this restriction requires $\Omega(m, d)$ copies of ρ in the worst case.

CHAPTER 2

Preliminaries

This chapter presents the technical foundations for the thesis. Section 2.1 and Section 2.2 cover the basics of quantum theory, including states, measurements, observables, and quantum operations. Section 2.3 explores the relationship between probability theory and quantum mechanics, while Section 2.4 and Section 2.5 introduce classical and quantum measures of divergence. Section 2.6 reviews key concepts from representation theory and Section 2.7 defines the complexity model used throughout this work. Finally, Section 2.8 establishes auxiliary results needed in later chapters.

2.1. Operators on finite-dimensional Hilbert spaces

This section provides a brief introduction to finite-dimensional complex Hilbert spaces $\mathcal{H}$ and the associated algebras of linear operators $\mathrm{B}(\mathcal{H})$ which are used in von Neumann's formulation of quantum mechanics [55]. The main purpose of this review is to establish our notation and emphasize a basis-independent approach; for a more comprehensive introduction, see e.g. [32, 37, 50].

A complex ***inner product space*** $\mathcal{H}$ is a vector space over $\mathbb{C}$ with an ***inner product*** $\langle \_, \_ \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$, i.e. a sesquilinear form[1] $(x, y) \mapsto \langle x, y\rangle \in \mathbb{C}$ such that, for all $x, y, z \in \mathcal{H}$,

(i) $\langle z, ax + y\rangle = a\langle z, x\rangle + \langle z, y\rangle$ for all $a \in \mathbb{C}$;
(ii) $\langle y, x\rangle = \overline{\langle x, y\rangle}$;
(iii) $\langle x, x\rangle \geq 0$;
(iv) $\langle x, x\rangle = 0$ if and only if $x = 0$.

Given an inner product $\langle \_, \_ \rangle$, the map $x \mapsto \|x\| = \sqrt{\langle x, x\rangle}$ defines a norm on $\mathcal{H}$. If $\mathcal{H}$ is complete (i.e. every Cauchy sequence in $\mathcal{H}$ converges) with respect to the metric induced by this norm, $d(x, y) = \|x - y\|$, then $\mathcal{H}$ is a ***Hilbert space***.

If the form $\langle \_, \_ \rangle$ does not satisfy the condition that $\langle x, x\rangle = 0$ holds only for $x = 0$, then $\langle \_, \_ \rangle$ is a ***pre-inner product*** and $\mathcal{H}$ is a ***pre-Hilbert space***. In that scenario, $\langle \_, \_ \rangle$ defines an inner product on the quotient $\mathcal{H}/N$ with $N = \{x \in \mathcal{H} \mid \langle x, x\rangle = 0\}$ and the completion of $\mathcal{H}/N$ is a Hilbert space.

$\dim(\mathcal{H})$ denotes the dimension of $\mathcal{H}$ as a vector space over $\mathbb{C}$.

Remark 2.1.1. Henceforth, the Hilbert spaces considered in this thesis are all assumed to be finite-dimensional.

Given Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, $\mathrm{B}(\mathcal{H}, \mathcal{K})$ denotes the set of linear maps $T : \mathcal{H} \to \mathcal{K}$ from $\mathcal{H}$ to $\mathcal{K}$. The algebra $\mathrm{B}(\mathcal{H}, \mathcal{H})$ of linear operators on $\mathcal{H}$ is denoted by $\mathrm{B}(\mathcal{H})$. If $\mathcal{H}$ is finite-dimensional,

[1] In this work, a ***sesquilinear*** form is taken to be conjugate-linear in the first argument and linear in the second argument, as is common in the physics literature.

then all linear operators $T \in \mathrm{B}(\mathcal{H})$ are bounded, viz. there exists a constant K such that $\|T(x)\| \leq K\|x\|$ for all $x \in \mathcal{H}$.

If $T \in \mathrm{B}(\mathcal{H}, \mathcal{K})$, then there exists a unique operator $T^\dagger \in \mathrm{B}(\mathcal{K}, \mathcal{H})$ such that, for all $x \in \mathcal{H}$ and $y \in \mathcal{K}$,

$$\langle T^\dagger x, y \rangle = \langle x, Ty \rangle.$$

$T^\dagger$ is called the ***adjoint*** of T. If T is represented as a matrix, then $T^\dagger$ is equal to the conjugate-transpose of T: $T^\dagger = \overline{T}^{\mathsf{T}}$.

There exist certain special types of operators in $\mathrm{B}(\mathcal{H})$, described next, that play an important role in this work. An operator $T \in \mathrm{B}(\mathcal{H})$ is ***normal*** if T commutes with its adjoint, $T^\dagger T = TT^\dagger$. T is ***self-adjoint*** if it is equal to its adjoint, viz. $T^\dagger = T$. T is ***positive*** if it is of the form $T = S^\dagger S$ for some operator $S \in \mathrm{B}(\mathcal{H})$; it is easy to check that a positive operator is self-adjoint. A ***projection*** is a self-adjoint operator $\Pi \in \mathrm{B}(\mathcal{H})$ such that $\Pi^2 = \Pi$. $\mathbf{1}_\mathcal{H}$ denotes the ***identity*** operator on $\mathcal{H}$ and $\mathbf{1}_d$ is the identity operator on $\mathbb{C}^d$. When no confusion can arise, the subscript will be dropped. An operator $T \in \mathrm{B}(\mathcal{H})$ is ***invertible*** if there exists $S \in \mathrm{B}(\mathcal{H})$ such that $TS = ST = \mathbf{1}_\mathcal{H}$. An operator $U \in \mathrm{B}(\mathcal{H})$ is ***unitary*** if $U^\dagger U = UU^\dagger = \mathbf{1}_\mathcal{H}$. The set of unitary operators on $\mathcal{H}$ is denoted by $\mathrm{U}(\mathcal{H})$; $\mathrm{U}(\mathcal{H})$ is a subgroup of the group $\mathrm{GL}(\mathcal{H})$ of invertible operators on $\mathcal{H}$.

Given operators $X, Y \in \mathrm{B}(\mathcal{H})$, the relation $X \leq Y$ holds if $Y - X$ is a positive operator. $\leq$ defines a partial order on self-adjoint operators in $\mathrm{B}(\mathcal{H})$. "$X \geq 0$" is shorthand for "X is positive."

Notation 2.1.2. For operators $E \in \mathrm{B}(\mathcal{H})$ with $0 \leq E \leq \mathbf{1}$, let $\overline{E} \in \mathrm{B}(\mathcal{H})$ be the operator defined by $\overline{E} = \mathbf{1} - E$.

The ***spectrum*** $\mathrm{sp}(T)$ of an operator $T \in \mathrm{B}(\mathcal{H})$ is the set $\mathrm{sp}(T) = \{\alpha \in \mathbb{C} \mid T - \alpha\,\mathbf{1}_\mathcal{H}$ is not invertible$\}$. The elements of $\mathrm{sp}(T)$ are eigenvalues of T. It holds that:

(i) $\mathrm{sp}(T) \subseteq \mathbb{R}$ if T is self-adjoint;
(ii) $\mathrm{sp}(T) \subseteq \mathbb{R}_{\geq 0}$ if T is positive;
(iii) $\mathrm{sp}(\Pi) \subseteq \{0, 1\}$ if Π is a projection;
(iv) $\mathrm{sp}(U) \subseteq \{z \in \mathbb{C} \mid |z| = 1\}$ if U is unitary.

The spectral theorem characterizes normal operators as complex linear combinations of mutually orthogonal projections:

Theorem 2.1.3 (spectral theorem). *A normal operator $T \in \mathrm{B}(\mathcal{H})$ on a finite-dimensional Hilbert space $\mathcal{H}$ admits a unique decomposition*

$$T = \alpha_1 \Pi_1 + \cdots + \alpha_k \Pi_k,$$

where $\{\alpha_1, \ldots, \alpha_k\} = \mathrm{sp}(T)$, $\Pi_1, \ldots, \Pi_k \in \mathrm{B}(\mathcal{H})$ are mutually orthogonal projections (i.e. $\Pi_i \Pi_j = 0$ for $i \neq j$) and $\Pi_1 + \cdots + \Pi_k = \mathbf{1}$.

If $p(x) = a_\ell x^\ell + \cdots + a_1 x + a_0$ is a polynomial with coefficients $a_0, \ldots, a_\ell \in \mathbb{C}$, then, for an operator T with decomposition as in Theorem 2.1.3,

$$p(T) = a_\ell T^\ell + \cdots + a_0\,\mathbf{1} = p(\alpha_1)\Pi_1 + \cdots + p(\alpha_k)\Pi_k.$$

Thus, a polynomial $p(T)$ of a normal operator T can be defined and $\mathrm{sp}(p(T)) = p(\mathrm{sp}(T))$. Since $\mathrm{sp}(T)$ is compact, the Stone–Weierstrass theorem allows extending this result from polynomials to continuous functions. In particular, the ***square root*** $\sqrt{T}$ of an operator T is the unique positive operator such that $\left(\sqrt{T}\right)^2 = T$. Thus, the ***absolute value*** map $z \mapsto \sqrt{z \cdot \overline{z}}$ extends naturally to normal operators $|T| = \sqrt{T^\dagger T} = \sqrt{TT^\dagger}$.

Unless $T, S \in \mathrm{B}(\mathcal{H})$ commute, there is no simple relation between $\mathrm{sp}(TS)$, $\mathrm{sp}(T+S)$, $\mathrm{sp}(T)$, and $\mathrm{sp}(S)$. The following result establishes a relation between $\mathrm{sp}(TS)$ and $\mathrm{sp}(ST)$ for *all* operators T and S:

Theorem 2.1.4. *For all operators $T, S \in \mathrm{B}(\mathcal{H})$, $\mathrm{sp}(TS) \cup \{0\} = \mathrm{sp}(ST) \cup \{0\}$.*

Given two Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, there exist two fundamental constructions that allow us to combine $\mathcal{H}$ and $\mathcal{K}$ together into a "composite" Hilbert space: the tensor product $\mathcal{H} \otimes \mathcal{K}$ and the direct sum $\mathcal{H} \oplus \mathcal{K}$.

The ***tensor product*** $\mathcal{H} \otimes \mathcal{K}$ of Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ is a Hilbert space with the inner product $\langle x_1 \otimes y_1, x_2 \otimes y_2 \rangle = \langle x_1, x_2 \rangle \cdot \langle y_1, y_2 \rangle$. Given two operators $T \in \mathrm{B}(\mathcal{H})$ and $S \in \mathrm{B}(\mathcal{K})$, their tensor product $T \otimes S$ is the uniquely determined linear operator on $\mathrm{B}(\mathcal{H} \otimes \mathcal{K})$ satisfying $(T \otimes S)(x \otimes y) = Tx \otimes Sy$ for all $x \in \mathcal{H}$ and $y \in \mathcal{K}$.

The ***direct sum*** $\mathcal{H} \oplus \mathcal{K}$ is also a Hilbert space with inner product $\langle x_1 \oplus y_1, x_2 \oplus y_2 \rangle = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle$.

We adopt Dirac's bra–ket notation for the representation of vectors and their dual linear forms:

Notation 2.1.5. Dirac's bra–ket notation represents vectors in $\mathcal{H}$ and their associated linear functionals as follows:

(i) $|x\rangle$ denotes the vector $x \in \mathcal{H}$;
(ii) $\langle x|$ denotes the linear functional defined by $y \mapsto \langle x, y \rangle$ for all $y \in \mathcal{H}$;
(iii) $\langle x| = |x\rangle^\dagger$ is the adjoint of $|x\rangle$;
(iv) $\langle x|y\rangle$ denotes the inner product $\langle x, y \rangle$;
(v) $|x\rangle\langle y|$ denotes the operator $|z\rangle \mapsto \langle y|z\rangle \cdot |x\rangle$;
(vi) $\langle x|T|y\rangle = \langle x, Ty \rangle$;
(vii) $|x_1, x_2, \ldots, x_n\rangle$ denotes the tensor element $x_1 \otimes x_2 \otimes \cdots \otimes x_n$.

Thus, if $x \in \mathcal{H}$ is a unit vector, then $|x\rangle\langle x|$ is the projection onto the subspace spanned by x. If $\mathcal{H} = \mathbb{C}^d$, the ***standard basis*** of $\mathcal{H}$ is denoted by $|1\rangle, \ldots, |d\rangle$ where $|i\rangle = (\delta_{1i}, \delta_{2i}, \ldots, \delta_{di})$; e.g. $|1\rangle = (1, 0, 0, \ldots, 0)$, $|2\rangle = (0, 1, 0, \ldots, 0)$, $\ldots$, $|d\rangle = (0, 0, \ldots, 0, 1)$.

The ***trace*** of an operator $T \in \mathrm{B}(\mathcal{H})$ is defined by

$$\mathrm{tr}(T) = \sum_{i=1}^{d} \langle e_i|T|e_i\rangle,$$

where $\{e_1, \ldots, e_d\}$ is an orthonormal basis for $\mathcal{H}$. It is easy to check that this definition does not depend on the choice of basis $\{e_1, \ldots, e_d\}$. Thus, if T has the spectral decomposition from Theorem 2.1.3, then

$$\mathrm{tr}(T) = \alpha_1 \cdot d_1 + \cdots + \alpha_k \cdot d_k,$$

where $d_i = \dim(\Pi_i)$ is the dimension of the subspace determined by Π_i for $i = 1, \ldots, k$. The trace is a linear map $\mathrm{tr} : \mathrm{B}(\mathcal{H}) \to \mathbb{C}$ that has the following *cyclic* property:

Proposition 2.1.6. *For all operators $T, S \in \mathrm{B}(\mathcal{H})$, $\mathrm{tr}(ST) = \mathrm{tr}(TS)$.*

The ***Hilbert–Schmidt inner product*** $\langle \_, \_ \rangle$ on $\mathrm{B}(\mathcal{H})$ is defined by $\langle T, S\rangle = \mathrm{tr}(T^\dagger S)$. Thus, the set $\mathrm{B}(\mathcal{H})$ of operators on a Hilbert space $\mathcal{H}$ is a Hilbert space with the Hilbert–Schmidt inner product.

The ***Schatten*** p***-norm*** $\|\_\|_p$ on $\mathrm{B}(\mathcal{H})$ is defined by $\|T\|_p = \mathrm{tr}(|T|^p)^{\frac{1}{p}}$. The Schatten 2-norm is in fact the norm induced by the Hilbert–Schmidt inner product, i.e. $\|T\|_2 = \sqrt{\langle T, T\rangle}$. The Schatten p-norm of a normal operator is equal to the ℓ^p norm of its spectrum. The Schatten p-norm $\|\_\|_p$ is unitarily-invariant, viz.

Fact 2.1.7. For all operators $X \in \mathrm{B}(\mathcal{H})$ and unitary operators $U, V \in \mathrm{U}(\mathcal{H})$, $\|X\|_p = \|UXV\|_p$.

Given a positive operator $A \in \mathrm{B}(\mathcal{H})$, it is possible to define a pre-inner product parameterized by A, denoted $\langle \_, \_ \rangle_A$, on $\mathrm{B}(\mathcal{H})$ as follows:

$$\langle T, S\rangle_A = \langle T\sqrt{A}, S\sqrt{A}\rangle = \mathrm{tr}(AT^\dagger S), \tag{1}$$

where the last equality uses the cyclic property of the trace. This pre-inner product induces a seminorm $\|T\|_A = \sqrt{\langle T, T\rangle_A}$ on $\mathrm{B}(\mathcal{H})$ and satisfies the Cauchy–Schwarz inequality:

Proposition 2.1.8. *Let $A \in \mathrm{B}(\mathcal{H})$ with $A \geq 0$. For all operators $T, S \in \mathrm{B}(\mathcal{H})$,*

$$|\langle T, S\rangle_A| \leq \|T\|_A \|S\|_A.$$

The ***Schur product*** $T \odot S$ of operators $T, S \in \mathrm{B}(\mathbb{C}^d)$ is defined by $\langle i|(T \odot S)|j\rangle = \langle i|T|j\rangle \cdot \langle i|S|j\rangle$ for all $i, j \in [d]$. The ***commutator*** $[T, S]$ is defined by $[T, S] = TS - ST$.

The remainder of this section is concerned with linear maps between operator algebras $\mathrm{B}(\mathcal{H})$ and $\mathrm{B}(\mathcal{K})$.

A linear map $\mathcal{S} : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{K})$ is ***positive*** if $\mathcal{S}$ maps positive operators in $\mathrm{B}(\mathcal{H})$ to positive operators in $\mathrm{B}(\mathcal{K})$, viz. $\mathcal{S}(A) \geq 0$ for all $A \in \mathrm{B}(\mathcal{H})$ with $A \geq 0$. $\mathcal{S}$ is ***completely positive*** (CP) if $\mathcal{S} \otimes \mathbf{1}_d : \mathrm{B}(\mathcal{H}) \otimes \mathrm{B}(\mathbb{C}^d) \to \mathrm{B}(\mathcal{K}) \otimes \mathrm{B}(\mathbb{C}^d)$ is a positive map for all $d \geq 1$.

Example. An important example of a CP map is the partial trace: there exists a unique linear map $\mathrm{Tr}_{\mathcal{K}} : \mathrm{B}(\mathcal{H} \otimes \mathcal{K}) \to \mathrm{B}(\mathcal{H})$, the ***partial trace*** over $\mathcal{K}$, such that $\mathrm{Tr}_{\mathcal{K}}(T \otimes S) = \mathrm{tr}(S) \cdot T$ for all $T \in \mathrm{B}(\mathcal{H})$ and $S \in \mathrm{B}(\mathcal{K})$.

The following result, due to Kraus [38], describes the structure of CP maps:

Theorem 2.1.9 (Kraus)**.** *A completely positive map $\mathcal{S} : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{K})$ admits a representation of the form*

$$\mathcal{S}(X) = E_1 X E_1^\dagger + \cdots + E_k X E_k^\dagger$$

with $E_1, \ldots, E_k \in \mathrm{B}(\mathcal{H}, \mathcal{K})$.

A linear map $\mathcal{S} : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{K})$ is ***unital*** if $\mathcal{S}(\mathbf{1}_{\mathcal{H}}) = \mathbf{1}_{\mathcal{K}}$. The following inequality for positive unital maps is due to Kadison [36]:

Lemma 2.1.10 (Kadison)**.** *Let $\mathcal{S} : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{H})$ be a linear map. If $\mathcal{S}$ is positive and unital, then, for all self-adjoint $\mathcal{O} \in \mathrm{B}(\mathcal{H})$,*

$$\mathcal{S}(\mathcal{O})^2 \leq \mathcal{S}(\mathcal{O}^2).$$

2.2. Quantum theory: states, observables, measurements, and operations

This section reviews the mathematical formulation of quantum mechanics used in this thesis, due to [55]. Let $\mathcal{H} = \mathbb{C}^d$ denote a finite-dimensional Hilbert space.

Definition 2.2.1. A ***quantum state*** ρ on $\mathcal{H}$ is a positive operator $\rho \in \mathrm{B}(\mathcal{H})$ with $\operatorname{tr}(\rho) = 1$. A quantum state of the form $|x\rangle\langle x|$ for $x \in \mathcal{H}$ is called ***pure***.

Thus, by the spectral theorem,

$$\rho = \alpha_1 \Pi_1 + \cdots + \alpha_k \Pi_k,$$

with $\alpha_1, \ldots, \alpha_k \in \mathbb{R}_{\geq 0}$, $\alpha_1 + \cdots + \alpha_k = \operatorname{tr}(\rho) = 1$, and $\Pi_1, \ldots, \Pi_k \in \mathrm{B}(\mathcal{H})$ mutually orthogonal projections summing to $\mathbf{1}$; i.e. ρ is a convex combination of mutually orthogonal projections that sum to the identity.

Example 2.2.2. An important example of a quantum state on $\mathbb{C}^d$ is the ***maximally mixed state***, $\frac{1}{d} \cdot \mathbf{1}$, whose eigenvalues are all equal to $\frac{1}{d}$. The maximally mixed state is the quantum analog of the uniform distribution on $[d]$.

Definition 2.2.3. A ***quantum operation*** $\mathcal{S} : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{K})$ is a completely positive map such that $\mathcal{S}(\mathbf{1}_{\mathcal{H}}) \leq \mathbf{1}_{\mathcal{K}}$. Furthermore, $\mathcal{S}$ is a ***quantum channel*** if $\mathcal{S}(\mathbf{1}_{\mathcal{H}}) = \mathbf{1}_{\mathcal{K}}$.

Quantum operations represent admissible (i.e. physically realizable) transitions between quantum states. If ρ is a quantum state and $\mathcal{S}$ is a quantum operation, then applying the operation $\mathcal{S}$ to ρ transforms ρ into the state

$$\frac{\mathcal{S}(\rho)}{\operatorname{tr}(\mathcal{S}(\rho))}.$$

For example, any operator $E \in \mathrm{B}(\mathcal{H})$ with $E^\dagger E \leq \mathbf{1}$ defines a quantum operation $\rho \mapsto E\rho E^\dagger$. Indeed, by Theorem 2.1.9, every quantum operation $\mathcal{S}$ is of the form

$$\mathcal{S}(\rho) = E_1 \rho E_1^\dagger + \cdots + E_k \rho E_k^\dagger$$

with $E_1, \ldots, E_k \in \mathrm{B}(\mathcal{H})$ such that $E_1^\dagger E_1 + \cdots + E_k^\dagger E_k \leq \mathbf{1}$.

An observer can interact with a quantum state by means of a measurement:

Definition 2.2.4. A ***quantum measurement*** $\mathcal{M}$ is a tuple $\mathcal{M} = (E_1, \ldots, E_k)$ of operators $E_1, \ldots, E_k \in \mathrm{B}(\mathcal{H})$ with $E_1^\dagger E_1 + \cdots + E_k^\dagger E_k = \mathbf{1}$. An observer applying measurement $\mathcal{M}$ to the state ρ registers an outcome $\boldsymbol{i} \in [k]$ with probability $\operatorname{tr}(\rho E_{\boldsymbol{i}}^\dagger E_{\boldsymbol{i}})$; this is similar to sampling from a discrete probability distribution $p = (p_1, \ldots, p_k)$ on the set $[k] = \{1, \ldots, k\}$ with $p_i = \operatorname{tr}(\rho E_i^\dagger E_i)$ for all $i = 1, \ldots, k$.

The fundamental difference between classical distribution sampling and quantum measurements is that sampling does not change the underlying distribution, whereas the state ρ is changed by the quantum measurement $\mathcal{M}$ *depending on the outcome observed* $\boldsymbol{i}$; viz. when the observer registers the outcome $\boldsymbol{i}$, we say that the state ρ *collapses* to

$$\frac{E_{\boldsymbol{i}} \rho E_{\boldsymbol{i}}^\dagger}{\operatorname{tr}(\rho E_{\boldsymbol{i}}^\dagger E_{\boldsymbol{i}})}.$$

Notation 2.2.5. It is helpful to think of the expression in the equation above as the state ρ *conditioned on the outcome* $E_i^\dagger E_i$. Thus, for $E \in \mathrm{B}(\mathcal{H})$ with $E^\dagger E \leq \mathbf{1}$ and $\operatorname{tr}(\rho E^\dagger E) > 0$, we introduce the notation

$$\rho|_E = \frac{E\rho E^\dagger}{\operatorname{tr}(\rho E^\dagger E)}$$

to denote the state ρ conditioned on $E^\dagger E$.

Note that the statistics of $\mathcal{M}$ only depend on $E_i^\dagger E_i$ for $i = 1, \ldots, k$ and on the state ρ. If the observer is not interested in the post-measurement state, then it is more convenient to work with positive operator-valued measures:

Definition 2.2.6. A ***positive operator-valued measure*** (POVM) $\mathcal{E} = (E_1, \ldots, E_k)$ is a tuple of positive operators $E_1, \ldots, E_k \in \mathrm{B}(\mathcal{H})$ such that $E_1 + \cdots + E_k = \mathbf{1}$.

If the operators $E_1, \ldots, E_k$ are all projections, then $\mathcal{E}$ is called a ***projection-valued measure*** (PVM).

A POVM $\mathcal{E}$ and a state ρ determine a probability distribution $p = (p_1, \ldots, p_k) \in \mathbb{R}^k$ with $p_i = \operatorname{tr}(\rho E_i)$:

$$\begin{aligned} p_1 + \cdots + p_k &= \operatorname{tr}(\rho E_1) + \cdots + \operatorname{tr}(\rho E_k) \\ &= \operatorname{tr}(\rho \cdot (E_1 + \cdots + E_k)) = \operatorname{tr}(\rho \cdot \mathbf{1}) = \operatorname{tr}(\rho) = 1. \end{aligned}$$

Definition 2.2.7. A ***quantum event*** is a positive operator $E \in \mathrm{B}(\mathcal{H})$ with $0 \leq E \leq \mathbf{1}$. Quantum events are also known as ***effect***s [38] in the literature. An event E determines a binary POVM $\{E, \overline{E}\}$.

The distribution determined by a binary POVM $\{E, \overline{E}\}$ is Bernoulli(tr(ρE)).

The quantum analog of a real random variable is an observable:

Definition 2.2.8. An ***observable*** $\mathcal{O}$ is a self-adjoint operator in $\mathrm{B}(\mathcal{H})$.

By Theorem 2.1.3, an observable $\mathcal{O}$ admits a decomposition

$$\mathcal{O} = x_1\Pi_1 + \cdots + x_k\Pi_k,$$

where $(\Pi_1, \ldots, \Pi_k)$ is a PVM and $x_i \in \mathbb{R}$ for $i = 1, \ldots, k$, so $\mathcal{O}$ and ρ determine a real random variable $\boldsymbol{X} \in \{x_1, \ldots, x_k\} \subseteq \mathbb{R}$ such that

$$\mathbf{P}[\boldsymbol{X} = x_i] = \operatorname{tr}(\rho\Pi_i).$$

The relationship between probability theory and quantum states and measurements is explored further in the next section. The rest of this section is concerned with multipartite states, i.e. states representing joint physical systems.

If $\rho_\mathfrak{a} \in \mathrm{B}(\mathcal{H}_\mathfrak{a})$ represents the state of a physical system $\mathfrak{a}$ and $\rho_\mathfrak{b} \in \mathrm{B}(\mathcal{H}_\mathfrak{b})$ represents the state of a physical system $\mathfrak{b}$, then the state of the joint physical system $(\mathfrak{a}, \mathfrak{b})$ is represented by $\rho_\mathfrak{a} \otimes \rho_\mathfrak{b} \in \mathrm{B}(\mathcal{H}_\mathfrak{a} \otimes \mathcal{H}_\mathfrak{b})$. A state in $\mathrm{B}(\mathcal{H}_\mathfrak{a} \otimes \mathcal{H}_\mathfrak{b})$ of the form $\rho_\mathfrak{a} \otimes \rho_\mathfrak{b}$ is called a ***product state***.

In quantum mechanics, it is possible for the joint system $(\mathfrak{a}, \mathfrak{b})$ to be in a state $\rho \in \mathrm{B}(\mathcal{H}_\mathfrak{a} \otimes \mathcal{H}_\mathfrak{b})$ which is not a product state or even a convex combination of product states:

Definition 2.2.9. A quantum state $\rho \in \mathrm{B}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$ is ***separable*** if there exist states $\sigma_1, \ldots, \sigma_k \in \mathrm{B}(\mathcal{H}_{\mathfrak{a}})$ and $\tau_1, \ldots, \tau_k \in \mathrm{B}(\mathcal{H}_{\mathfrak{b}})$ such that

$$\rho = a_1 \cdot \sigma_1 \otimes \tau_1 + \cdots + a_k \cdot \sigma_k \otimes \tau_k,$$

where $a_1, \ldots, a_k \in \mathbb{R}_{\geq 0}$ with $a_1 + \cdots + a_k = 1$; in other words, ρ is separable if it can be expressed as a convex combination of product states. The set of separable states on $\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}}$ is denoted by $\mathrm{Sep}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$. $\mathrm{Sep}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$ is a strict subset of the set of quantum states in $\mathrm{B}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$; a state $\rho \in \mathrm{B}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$ which is not separable is called ***entangled***.

Example 2.2.10. The maximally mixed state on $\mathbb{C}^d \otimes \mathbb{C}^d$ is a product state since $\frac{\mathbf{1}}{d^2} = \frac{\mathbf{1}}{d} \otimes \frac{\mathbf{1}}{d}$.

Definition 2.2.11. Let $\mathrm{Sep}_{\pm}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$ denote the ***cylindrical symmetrization*** of the set of separable states $\mathrm{Sep}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}})$:

$$\mathrm{Sep}_{\pm}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}}) = \mathrm{conv}(\mathrm{Sep}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}}) \cup (-\,\mathrm{Sep}(\mathcal{H}_{\mathfrak{a}} \otimes \mathcal{H}_{\mathfrak{b}}))),$$

where $\mathrm{conv}(E)$ denotes the convex hull of the set E.

2.3. Quantum probability

In light of the connection between observables and real random variables, it is reasonable to introduce the following notation:

Definition 2.3.1. The ***expectation*** of an operator $\mathcal{O} \in \mathrm{B}(\mathcal{H})$ with respect to a state $\rho \in \mathrm{B}(\mathcal{H})$ is defined by

$$\mathop{\mathbf{E}}_{\rho}[\mathcal{O}] = \mathrm{tr}(\rho\mathcal{O}).$$

If $\mathcal{O}$ is self-adjoint, i.e. an observable, with spectral decomposition $\mathcal{O} = x_1\Pi_1 + \cdots + x_k\Pi_k$, then $\mathbf{E}_\rho[\mathcal{O}]$ is equal to the expectation $\mathbf{E}[\boldsymbol{X}]$ of the classical real random variable $\boldsymbol{X} \in \{x_1, \ldots, x_k\}$ defined by $\mathbf{P}[\boldsymbol{X} = x_i] = \mathrm{tr}(\rho\Pi_i)$ for all $i \in [k]$.

Since $\mathbf{E}_\rho[\mathbf{1}] = \mathrm{tr}(\rho\,\mathbf{1}) = \mathrm{tr}(\rho) = 1$, $\mathbf{E}_\rho[\mathcal{O}^\dagger] = \overline{\mathbf{E}_\rho[\mathcal{O}]}$, and $\mathbf{E}_\rho[\mathcal{O}^\dagger\mathcal{O}] \geq 0$ for all $\mathcal{O} \in \mathrm{B}(\mathcal{H})$, the map $\mathbf{E}_\rho[\_]$ defines a positive linear functional of norm 1 on $\mathrm{B}(\mathcal{H})$. Moreover, $\mathbf{E}_{\rho\otimes\rho'}[\_]$ has the following tensorization property: for all $\mathcal{O}, \mathcal{O}' \in \mathrm{B}(\mathcal{H})$,

$$\mathop{\mathbf{E}}_{\rho\otimes\rho'}[\mathcal{O} \otimes \mathcal{O}'] = \mathop{\mathbf{E}}_{\rho}[\mathcal{O}] \cdot \mathop{\mathbf{E}}_{\rho'}[\mathcal{O}'].$$

$\mathbf{E}_\rho[\_]$ is also monotone with respect to the partial order $\leq$ on positive operators:

Fact 2.3.2. For all self-adjoint operators $\mathcal{O}_1, \mathcal{O}_2 \in \mathrm{B}(\mathcal{H})$, it holds that $\mathcal{O}_1 \leq \mathcal{O}_2$ if and only if $\mathbf{E}_\rho[\mathcal{O}_1] \leq \mathbf{E}_\rho[\mathcal{O}_2]$ for all states $\rho \in \mathrm{B}(\mathcal{H})$.

Definition 2.3.3. The ***covariance*** of operators $\mathcal{O}_1, \mathcal{O}_2 \in \mathrm{B}(\mathcal{H})$ with respect to a state $\rho \in \mathrm{B}(\mathcal{H})$ is the sesquilinear form defined by

$$\mathop{\mathbf{Cov}}_{\rho}[\mathcal{O}_1^\dagger\mathcal{O}_2] = \mathop{\mathbf{E}}_{\rho}[(\mathcal{O}_1 - \mu_1\,\mathbf{1})^\dagger(\mathcal{O}_2 - \mu_2\,\mathbf{1})] = \langle \mathcal{O}_1 - \mu_1\,\mathbf{1}, \mathcal{O}_2 - \mu_2\,\mathbf{1}\rangle_\rho = \langle \mathcal{O}_1, \mathcal{O}_2\rangle_\rho - \overline{\mu}_1\mu_2,$$

where $\mu_i = \mathbf{E}_\rho[\mathcal{O}_i]$ for $i = 1, 2$ and $\langle \_, \_\rangle_\rho$ is the pre-inner product defined in Equation (1).

Since $\mathbf{Cov}_\rho[\mathbf{1},\_] = \mathbf{Cov}_\rho[\_,\mathbf{1}] = 0$, it follows that $\mathbf{Cov}_\rho$ is translation-invariant in each argument; i.e. $\mathbf{Cov}_\rho[\mathcal{O}_1 + a\,\mathbf{1}, \mathcal{O}_2 + b\,\mathbf{1}] = \mathbf{Cov}_\rho[\mathcal{O}_1, \mathcal{O}_2]$ for all $a, b \in \mathbb{C}$. Furthermore, $\mathbf{Cov}_{\rho\otimes\rho'}[\_,\_]$ satisfies the following tensorization property:

$$\underset{\rho\otimes\rho'}{\mathbf{Cov}}[\mathcal{O}_1 \otimes \mathcal{O}_1', \mathcal{O}_2 \otimes \mathcal{O}_2'] = \underset{\rho}{\mathbf{Cov}}[\mathcal{O}_1, \mathcal{O}_2] \cdot \underset{\rho'}{\mathbf{Cov}}[\mathcal{O}_1', \mathcal{O}_2'],$$

for all operators $\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_1', \mathcal{O}_2' \in \mathrm{B}(\mathcal{H})$. Hence,

$$\underset{\rho\otimes\rho'}{\mathbf{Cov}}[\mathcal{O} \otimes \mathbf{1}, \mathbf{1} \otimes \mathcal{O}'] = 0. \tag{2}$$

The equality above is a quantum analog of the classical fact that the covariance of independent random variables is zero.

Definition 2.3.4. The ***variance*** of operator $\mathcal{O} \in \mathrm{B}(\mathcal{H})$ with respect to a state $\rho \in \mathrm{B}(\mathcal{H})$ is defined by

$$\underset{\rho}{\mathbf{Var}}[\mathcal{O}] = \underset{\rho}{\mathbf{Cov}}[\mathcal{O}, \mathcal{O}].$$

It holds that $\mathbf{Var}_\rho[\mathcal{O}] \geq 0$ for all $\mathcal{O} \in \mathrm{B}(\mathcal{H})$, $\mathbf{Var}_\rho[c\mathcal{O}] = |c|^2\,\mathbf{Var}_\rho[\mathcal{O}]$ for all $c \in \mathbb{C}$, and

$$\underset{\rho}{\mathbf{Var}}\left[\sum_{i=1}^{k} \mathcal{O}_i\right] = \sum_{i=1}^{k} \underset{\rho}{\mathbf{Var}}[\mathcal{O}_i] + \sum_{\substack{i,j=1 \\ i\neq j}}^{k} \underset{\rho}{\mathbf{Cov}}[\mathcal{O}_i, \mathcal{O}_j]$$

for all operators $\mathcal{O}_1, \ldots, \mathcal{O}_k \in \mathrm{B}(\mathcal{H})$.

We end this section with a lemma that will assist us in finding observables with low variance:

Lemma 2.3.5. *Let $\Phi : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{H})$ be a positive unital linear map. If $\mathbf{E}_\rho \circ \Phi = \mathbf{E}_\rho$ then $\mathbf{Var}_\rho[\Phi(\mathcal{O})] \leq \mathbf{Var}_\rho[\mathcal{O}]$ for all observables $\mathcal{O} \in \mathrm{B}(\mathcal{H})$.*

Proof. Let $\mathcal{O} \in \mathrm{B}(\mathcal{H})$ be an observable. Since $\mathbf{E}_\rho[\Phi(\mathcal{O})] = \mathbf{E}_\rho[\mathcal{O}]$, it suffices to show that $\mathbf{E}_\rho[\Phi(\mathcal{O})^2] \leq \mathbf{E}_\rho[\mathcal{O}^2]$. Since $\mathcal{O}$ is self-adjoint and Φ is positive and unital, $\Phi(\mathcal{O})^2 \leq \Phi(\mathcal{O}^2)$, by Lemma 2.1.10. Hence, by Fact 2.3.2, $\mathbf{E}_\rho[\Phi(\mathcal{O})^2] \leq \mathbf{E}_\rho[\Phi(\mathcal{O}^2)]$. Since $\mathbf{E}_\rho \circ \Phi = \mathbf{E}_\rho$, it follows that $\mathbf{E}_\rho[\Phi(\mathcal{O})^2] \leq \mathbf{E}_\rho[\Phi(\mathcal{O}^2)] = \mathbf{E}_\rho[\mathcal{O}^2]$, as needed. ∎

Thus, the class of mean-preserving positive unital maps is variance-nonincreasing.

Remark 2.3.6. Although there are other POVMs that can be associated with an observable $\mathcal{O} \in \mathrm{B}(\mathcal{H})$ apart from its spectral decomposition, the variance of the resulting random variables is at least $\mathbf{Var}_\rho[\mathcal{O}]$. Indeed, suppose $\mathcal{M} = \{E_1, \ldots, E_k\}$ is a POVM and $x_1, \ldots, x_k$ are real coefficients such that $\mathcal{O} = x_1E_1 + \cdots + x_kE_k$. Let $\Phi : \mathrm{B}(\mathbb{C}^k) \to \mathrm{B}(\mathbb{C}^k)$ denote the map defined by $\Phi(A) = A_{11}E_1 + \cdots + A_{kk}E_k$ for all $A \in \mathrm{B}(\mathbb{C}^k)$. Since $\mathcal{M}$ is a POVM, the map Φ is positive and unital. Hence, by Lemma 2.1.10,

$$\mathcal{O}^2 = \Phi(\mathrm{diag}(x_1, \ldots, x_k))^2 \leq \Phi(\mathrm{diag}(x_1, \ldots, x_k)^2) = x_1^2E_1 + \cdots + x_k^2E_k$$

and the result now follows from Fact 2.3.2.

2.4. Distances and divergences between probability distributions

In this section, we introduce different measures of dissimilarity between pairs of classical probability distributions. Fix a positive integer d and let $p, q \in \mathbb{R}^d$ denote arbitrary probability distributions on $[d] = \{1, 2, 3, \dots, d\}$.

Definition 2.4.1. The ***total variation (TV) distance*** $d_{\mathrm{TV}}(p, q)$ is defined by

$$d_{\mathrm{TV}}(p,q) = \frac{1}{2}\|p-q\|_1 = \max_{E \subseteq [d]} \left| \mathop{\mathbf{P}}_{p}[E] - \mathop{\mathbf{P}}_{q}[E] \right|.$$

d_{TV} is a metric, so it satisfies the triangle inequality. Furthermore, since $d_{\mathrm{TV}}(p, q) = |\mathbf{P}_p[E] - \mathbf{P}_q[E]|$ for some $E \subseteq [d]$, it holds that $0 \leq d_{\mathrm{TV}}(p, q) \leq 1$.

Definition 2.4.2. The ℓ^2 ***distance*** $d_{\ell^2}(p, q)$ is defined by

$$d_{\ell^2}(p,q) = \|p-q\|_2 = \sqrt{(p_1-q_1)^2 + \cdots + (p_k-q_k)^2}.$$

By the fact that the p-norm is nonincreasing and by the Cauchy–Schwarz inequality,

$$d_{\ell^2}(p,q) \leq 2d_{\mathrm{TV}}(p,q) \leq 2\sqrt{d}\, d_{\ell^2}(p,q).$$

d_{ℓ^2} is also a metric with $0 \leq d_{\ell^2}(p, q) \leq 2$.

Definition 2.4.3. The ***Bhattacharyya coefficient***[2] $\mathrm{BC}(p, q)$ is defined by

$$\mathrm{BC}(p,q) = \sqrt{p_1 q_1} + \cdots + \sqrt{p_d q_d}.$$

$\mathrm{BC}(p, q)$ has the following geometric interpretation. Let $\sqrt{p} = (\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_d})$ and let $\sqrt{q}$ be similarly defined. Thus, as vectors in $\mathbb{R}^d$, $\|\sqrt{p}\|_2 = \|\sqrt{q}\|_2 = 1$, so $\langle \sqrt{p}, \sqrt{q} \rangle = \mathrm{BC}(p, q)$ is equal to the cosine of the angle determined by the unit vectors $\sqrt{p}$ and $\sqrt{q}$ in d-dimensional space. Hence, $0 \leq \mathrm{BC}(p, q) \leq 1$ with:

(i) $\mathrm{BC}(p, q) = 1$ if and only if $p = q$, and
(ii) $\mathrm{BC}(p, q) = 0$ if and only if $\mathrm{supp}(p) \cap \mathrm{supp}(q) = \varnothing$.

The Bhattacharyya coefficient satisfies the following *tensorization* property:

$$\mathrm{BC}(p \otimes p', q \otimes q') = \mathrm{BC}(p, q) \cdot \mathrm{BC}(p', q'),$$

where the tensor product $p \otimes p'$ is interpreted as the product distribution on $[d] \times [d]$ arising from p and p'.

Definition 2.4.4. The ***Hellinger distance*** $d_{\mathrm{H}}(p, q)$ is defined by

$$d_{\mathrm{H}}(p,q) = \|\sqrt{p} - \sqrt{q}\|_2 = \sqrt{2 - 2\,\mathrm{BC}(p,q)}$$

d_{H} is a metric with $0 \leq d_{\mathrm{H}}(p, q) \leq \sqrt{2}$. The following relation holds between d_{H} and d_{TV} (see e.g. [51, Equation 7.22]):

$$\frac{1}{2}d_{\mathrm{H}}^2(p,q) \leq d_{\mathrm{TV}}(p,q) \leq d_{\mathrm{H}}(p,q) \cdot \sqrt{1 - \frac{d_{\mathrm{H}}^2(p,q)}{4}} \leq 1. \tag{3}$$

[2] Also known as *Hellinger affinity*.

Definition 2.4.5. The ***Kullback–Leibler (KL) divergence*** $d_{\mathrm{KL}}(p,q)$ is defined by

$$d_{\mathrm{KL}}(p,q) = \sum_{\substack{i=1\\ q_i>0}}^{d} p_i \ln(p_i/q_i)$$

with $d_{\mathrm{KL}}(p,q) = +\infty$ if $\operatorname{supp}(p) \not\subseteq \operatorname{supp}(q)$. Pinsker's inequality relates d_{KL} to d_{TV}:

$$d_{\mathrm{KL}}(p,q) \geq 2d_{\mathrm{TV}}^2(p,q).$$

Definition 2.4.6. The χ^2***-divergence*** $d_{\chi^2}(p,q)$ is defined by

$$d_{\chi^2}(p,q) = \sum_{\substack{i=1\\ q_i>0}}^{d} q_i \cdot \left(1 - \frac{p_i}{q_i}\right)^2 = \mathop{\mathbf{E}}_{i\sim q}\left[\left(1 - \frac{p_{\boldsymbol{i}}}{q_{\boldsymbol{i}}}\right)^2\right],$$

where, as with d_{KL}, $d_{\chi^2}(p,q) = +\infty$ if $\operatorname{supp}(p) \not\subseteq \operatorname{supp}(q)$. Neither d_{χ^2} nor d_{KL} are metrics.

The following inequalities relate d_{χ^2} to d_{TV} and d_{H}:

$$d_{\mathrm{H}}^2(p,q) \leq d_{\chi^2}(p,q) \tag{4}$$

$$d_{\mathrm{TV}}(p,q) \leq \frac{1}{2}\sqrt{d_{\chi^2}(p,q)}. \tag{5}$$

Although the TV distance is operationally the most meaningful, it is frequently challenging to compute, so bounding the TV distance in terms of ℓ^2 distance, Hellinger distance, or χ^2-divergence often proves more tractable.

Note that all of the distances and divergences reviewed are *permutation invariant*: if p' and q' are distributions on $[d]$ obtained by applying the same permutation $\pi \in \mathfrak{S}_d$ to the outcomes of p and q, then $d_*(p',q') = d_*(p,q)$, where d_* is any of the measures of dissimilarity defined above.

2.5. Distances and divergences between quantum states

This section introduces dissimilarity measures for quantum states. Leveraging the connection to probability theory, these are quantum counterparts to classical divergence measures presented in Section 2.4.

Remark 2.5.1. A straightforward way of porting classical dissimilarity measures to quantum states is as follows [20]: given a POVM $\mathcal{M} = (E_1, \ldots, E_k)$ and a quantum state ρ, let $\Delta_{\mathcal{M},\rho}$ denote the probability distribution on $[k]$ with the probability of outcome $i \in [k]$ being $\operatorname{tr}(\rho E_i)$ for $i = 1, \ldots, k$. Given a classical dissimilarity measure $d_*(p,q)$, we define its quantum counterpart by:

$$d_*(\rho,\sigma) = \sup_{\text{POVM } \mathcal{M}} d_*(\Delta_{\mathcal{M},\rho}, \Delta_{\mathcal{M},\sigma}),$$

where the suppremum is taken over all POVMs compatible with ρ and σ. In this case, we will say that the quantum divergence is a *measured version* of its classical counterpart.

Fix a positive integer d and let ρ and σ denote arbitrary quantum states on $\mathbb{C}^d$.

Definition 2.5.2. The ***trace distance*** $d_{\mathrm{tr}}(\rho,\sigma)$ is defined by

$$d_{\mathrm{tr}}(\rho,\sigma) = \frac{1}{2}\|\rho - \sigma\|_1 = \sup_{0\leq E\leq \mathbf{1}}\left|\mathop{\mathbf{E}}_{\rho}[E] - \mathop{\mathbf{E}}_{\sigma}[E]\right|.$$

d_{tr} is a metric with $0 \leq d_{\mathrm{tr}}(\rho,\sigma) \leq 1$. The following result [30] shows that the supremum in the variational interpretation of trace distance above is attained for some projection Π:

Proposition 2.5.3 (Holevo–Helstrom)**.** *For all states $\rho, \sigma \in \mathrm{B}(\mathcal{H})$, there exists a projection $\Pi \in \mathrm{B}(\mathcal{H})$ such that*

$$|\underset{\rho}{\mathbf{E}}[\Pi] - \underset{\sigma}{\mathbf{E}}[\Pi]| = d_{\mathrm{tr}}(\rho,\sigma).$$

Explicitly, Π can be taken to be the projection onto the subspace of $\mathcal{H}$ generated by all eigenspaces of $\rho - \sigma$ with nonnegative eigenvalues.

Thus, trace distance is the measured version of total variation distance, in the sense of Remark 2.5.1.

Definition 2.5.4. The ***Hilbert–Schmidt (HS) distance*** $d_{\mathrm{HS}}(\rho,\sigma)$ is defined by

$$d_{\mathrm{HS}}(\rho,\sigma) = \|\rho - \sigma\|_2.$$

Note that $d^2_{\mathrm{HS}}(\rho,\sigma) = \langle \rho, \rho\rangle - 2\langle \rho,\sigma\rangle + \langle \sigma,\sigma\rangle = \mathrm{tr}(\rho^2 - 2\rho\sigma + \sigma^2)$; i.e. the squared HS distance is the trace of a polynomial in ρ and σ. Moreover, by the Cauchy–Schwarz inequality,

$$d_{\mathrm{HS}}(\rho,\sigma) \leq 2 d_{\mathrm{tr}}(\rho,\sigma) \leq 2\sqrt{d}\, d_{\mathrm{HS}}(\rho,\sigma). \tag{6}$$

Definition 2.5.5. The ***fidelity*** $\mathrm{F}(\rho,\sigma)$ is defined by

$$\mathrm{F}(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \mathrm{tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right) = \mathrm{tr}(\sqrt{\rho\sigma}),$$

where the last equality follows from [8].

Fidelity is the measured version of the Bhattacharyya coefficient. Moreover, if $\alpha = (\alpha_1, \ldots, \alpha_d)$ and $\beta = (\beta_1, \ldots, \beta_d)$ are the spectra of ρ and σ, respectively, then $\mathrm{F}(\rho,\sigma) = \mathrm{BC}(\alpha,\beta)$ [20].

Definition 2.5.6. The ***Bures distance*** $d_{\mathrm{B}}(\rho,\sigma)$ is defined by

$$d_{\mathrm{B}}(\rho,\sigma) = \sqrt{2 - 2\,\mathrm{F}(\rho,\sigma)}.$$

The squared Bures distance is the measured version of squared Hellinger distance [20]. d_{B} is a metric with $0 \leq d_{\mathrm{B}}(\rho,\sigma) \leq \sqrt{2}$. Moreover, the following inequality relates d_{B} to d_{tr} (cf. Equation (3)):

$$\frac{1}{2} d^2_{\mathrm{B}}(\rho,\sigma) \leq d_{\mathrm{tr}}(\rho,\sigma) \leq d_{\mathrm{B}}(\rho,\sigma). \tag{7}$$

The last dissimilarity measure we will define is Bures χ^2-divergence, the measured version of classical χ^2-divergence [10]. Recall the definition of classical χ^2-divergence (cf. Definition 2.4.6):

$$d_{\chi^2}(p,q) = \sum_{\substack{i=1 \\ q_i > 0}}^{d} \frac{(p_i - q_i)^2}{q_i}.$$

In this formula, one divides by the probability q_i when $q_i > 0$. In order to define a quantum version of χ^2-divergence, we want to somehow "divide" by a quantum state σ. If σ is invertible, then division by σ corresponds to multiplying by σ^{-1} on the left or the right due to the noncommutativity of matrix multiplication. To avoid introducing one-sidedness in our definition, we define a type of commutative division [10]. To this end, let $\mathcal{R} : \mathrm{B}(\mathcal{H}) \to \mathrm{B}(\mathcal{H})$ be a map on operators defined by

$$\mathcal{R}_\sigma(T) = \frac{1}{2}(\sigma T + T\sigma).$$

In the orthonormal basis in which σ is diagonal, $\mathcal{R}_\sigma(T)$ can be expressed as a Schur product:

$$\mathcal{R}_\sigma(T) = T \odot \frac{1}{2}\left[\beta_i + \beta_j\right]_{i,j=1}^{d}.$$

Thus, $\mathcal{R}_\sigma(T)$ is commutative in the sense that $\mathcal{R}_\sigma(T) = \mathcal{R}_T(\sigma)$, and if $\beta_i + \beta_j \neq 0$ for all $i, j \in [d]$, then $\mathcal{R}_\sigma$ is clearly invertible. Moreover, using the fact that the Schur product of matrices $A, B \in \mathrm{B}(\mathbb{C}^d)$ in the standard basis is given by $A \odot B = \Pi(A \otimes B)\Pi$ where $\Pi = |1,1\rangle + |2,2\rangle + \cdots + |d,d\rangle$, it is straightforward to check that $\mathcal{R}_\sigma(T)$ is a quantum channel.

Since $\mathcal{R}_\sigma$ is an invertible quantum channel, there exists a quantum channel, Ω_σ, the inverse of $\mathcal{R}_\sigma$. In the orthonormal basis in which $\sigma = \mathrm{diag}(\beta_1, \ldots, \beta_d)$ is diagonal, we have

$$\Omega_\rho(T) = T \odot 2\big[(\beta_i + \beta_j)^{-1}\big]_{i,j=1}^{d}. \tag{8}$$

Definition 2.5.7. The ***Bures*** χ^2***-divergence*** $d_{\chi^2}(\rho, \sigma)$ is defined by

$$d_{\chi^2}(\rho, \sigma) = \mathrm{tr}((\rho - \sigma)\,(\Omega_\sigma(\rho - \sigma))).$$

Since d_{B} and d_{χ^2} are the measured versions of their classical counterparts, d_{H} and d_{χ^2}, respectively, it holds that $d_{\mathrm{B}}^2(\rho, \sigma) \leq d_{\chi^2}(\rho, \sigma)$.

All of the dissimilarity measures defined in this section are unitarily invariant, i.e. $d_*(\rho, \sigma) = d_*(U\rho U^\dagger, U\sigma U^\dagger)$ for all $U \in \mathrm{U}(\mathbb{C}^d)$. Furthermore, d_{tr}, d_{B}, and d_{χ^2} satisfy the quantum *data processing inequality*, viz. for all quantum channels $\mathcal{S}$,

$$d_*(\mathcal{S}(\rho), \mathcal{S}(\sigma)) \leq d_*(\rho, \sigma) \tag{9}$$

for $d_* \in \{d_{\mathrm{tr}}, d_{\mathrm{B}}, d_{\chi^2}\}$.

2.6. Representation theory

This section reviews a few facts from the representation theory of groups which are needed in Chapter 3. For additional background, consult e.g. [24, 52, 53].

Let $\mathfrak{S}_n$ denote the symmetric group on the alphabet $[n] = \{1, 2, 3, \ldots, n\}$ and let $\mathrm{U}(d)$ denote the group of $d \times d$ unitary matrices.

Definition 2.6.1. A *partition* λ is a nonincreasing sequence of nonnegative integers of finite support. If $\lambda_1 + \lambda_2 + \cdots = n$, then λ is said to be a partition of n, denoted by $\lambda \vdash n$. The size of the support of λ is called the *length* of the partition and is denoted by $\ell(\lambda)$. The *power sum* symmetric polynomial in d variables $p_\lambda(x_1, \ldots, x_d)$ associated to a partition λ of length k is defined by $p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots p_{\lambda_k}$, where $p_r(x_1, \ldots, x_d) = x_1^r + \cdots + x_d^r$ for all $r \geq 0$.

A permutation $\pi \in \mathfrak{S}_n$ decomposes uniquely into a product of disjoint cycles. The sequence of lengths of the cycles in this decomposition is called the ***cycle type*** of the permutation π and is denoted by $\mathrm{cyc}(\pi)$. Sorted in nonincreasing order, $\mathrm{cyc}(\pi)$ is a partition of n. Thus, the partitions of n index the conjugacy classes of $\mathfrak{S}_n$.

Definition 2.6.2. Let $\mathcal{P}$ denote the unitary representation of $\mathfrak{S}_n$ on the n-fold tensor product $(\mathbb{C}^d)^{\otimes n}$ defined by

$$\mathcal{P}(\pi) \, |x_1\rangle \otimes \cdots \otimes |x_n\rangle = |x_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |x_{\pi^{-1}(n)}\rangle,$$

for all $|x_1\rangle, \ldots, |x_n\rangle \in \mathbb{C}^d$ and $\pi \in \mathfrak{S}_n$. Furthermore, let Ad_U be the linear map on observables defined by $\mathrm{Ad}_U(X) = (U^{\otimes n})X(U^{\otimes n})^\dagger$ for all $U \in \mathrm{U}(d)$.

Definition 2.6.3. The *symmetric group algebra* $\mathbb{C}\mathfrak{S}_n$ is the algebra of functions $f : \mathfrak{S}_n \to \mathbb{C}$. The functions $1_\pi : \mathfrak{S}_n \to \mathbb{C}$ with $\pi \in \mathfrak{S}_n$ form a basis of $\mathbb{C}\mathfrak{S}_n$, where 1_π is defined by

$$1_\pi(\tau) = \begin{cases} 1, & \pi = \tau, \\ 0, & \pi \neq \tau. \end{cases}$$

With a slight abuse of notation, we use π to denote the function 1_π and think of elements of $\mathbb{C}\mathfrak{S}_n$ as linear combinations of permutations $\pi \in \mathfrak{S}_n$. Thus, the product in $\mathbb{C}\mathfrak{S}_n$ is obtained by extending the product in $\mathfrak{S}_n$ to a bilinear map. $\mathbb{C}\mathfrak{S}_n$ also admits a conjugate-linear involution $X \mapsto X^\dagger$ defined by $\pi^\dagger = \pi^{-1}$ for all $\pi \in \mathfrak{S}_n$.

The representation $\mathcal{P}$ of $\mathfrak{S}_n$ extends to a $*$-representation of the $*$-algebra $\mathbb{C}\mathfrak{S}_n$ as follows:

$$X = \sum_{\pi \in \mathfrak{S}_n} a_\pi \pi \mapsto \sum_{\pi \in \mathfrak{S}_n} a_\pi \mathcal{P}(\pi) = \mathcal{P}(X).$$

Since the representation $\mathcal{P}$ is unitary, it follows that $\mathcal{P}(X^\dagger) = \mathcal{P}(X)^\dagger$ for all $X \in \mathbb{C}\mathfrak{S}_n$.

The ***center*** of $\mathbb{C}\mathfrak{S}_n$, denoted by $Z(\mathbb{C}\mathfrak{S}_n)$, is the set of elements $X \in \mathbb{C}\mathfrak{S}_n$ that commute with every other element of $\mathbb{C}\mathfrak{S}_n$, viz. $XY = YX$ for all $Y \in \mathbb{C}\mathfrak{S}_n$. For all partitions $\kappa \vdash n$, let $\mathcal{O}_\kappa \in \mathbb{C}\mathfrak{S}_n$ be defined by

$$\mathcal{O}_\kappa = \operatorname*{avg}_{\substack{\pi \in \mathfrak{S}_n \\ \mathrm{cyc}(\pi) = \kappa}} \{\pi\}.$$

In other words, $\mathcal{O}_\kappa$ is the normalized indicator function of the conjugacy class of permutations of cycle type κ. The following elementary result relates the elements $\mathcal{O}_\kappa$ to the center of $\mathbb{C}\mathfrak{S}_n$.

Proposition 2.6.4. *$\{\mathcal{O}_\kappa \mid \kappa \vdash n\}$ is a linear basis for $Z(\mathbb{C}\mathfrak{S}_n)$.*

For a proof, see e.g. [**24**, Proposition 4.3.7]. Since $\mathcal{O}_\kappa^\dagger = \mathcal{O}_\kappa$ for all $\kappa \vdash n$, it follows that $\{\mathcal{O}_\kappa \mid \kappa \vdash n\}$ is also a basis for the real vector space of self-adjoint elements of $Z(\mathbb{C}\mathfrak{S}_n)$.

2.7. The complexity model

The problems considered in this thesis follow the template below:

Problem 2.7.1. Given unrestricted quantum measurement access to n copies $\rho^{\otimes n}$ of an unknown quantum state ρ on $\mathbb{C}^d$, find an algorithm that completes a given task $\mathcal{T}(d, \epsilon, \delta)$ using as few copies of ρ as possible.

The parameters of the task $\mathcal{T}(d, \epsilon, \delta)$ are as follows:

(a) d is the dimension of the Hilbert space underlying ρ;
(b) ϵ is the *proximity parameter* [**23**];
(c) δ is the error probability.

In this work, two main types of tasks are considered:

i. **estimation tasks**: given a function $f : \mathrm{B}(\mathbb{C}^d) \to (X, d)$ from quantum states to a metric space (X, d) of possible estimates (e.g. $X = \mathbb{R}$ or $X = \mathbb{R}^k$), the estimation task determined by f is to:

   produce an estimate $\widehat{y} \in X$ of $f(\rho)$ such that $d(\widehat{y}, f(\rho)) \le \epsilon$ with probability at least $1 - \delta$.

   In the case where $X = \mathbb{R}$, an algorithm to complete an estimation task can be given simply as an observable $\mathcal{O}$ on $(\mathbb{C}^d)^{\otimes n}$. Given the spectral decomposition

$$\mathcal{O} = x_1\Pi_1 + x_2\Pi_2 + \cdots + x_k\Pi_k,$$

   the algorithm measures $\rho^{\otimes n}$ using $\{\Pi_1, \ldots, \Pi_k\}$ and outputs estimate $x_{\boldsymbol{i}}$ upon observing outcome $\boldsymbol{i} \in [k]$.

   The observable $\mathcal{O}$ must satisfy the following concentration inequality:

$$\sum_{\substack{1 \le i \le k \\ d(f(\rho), x_i) \le \epsilon}} \mathop{\mathbf{E}}_{\rho^{\otimes n}} [\Pi_i] \ge 1 - \delta.$$

ii. **property testing tasks**: given a property $\mathcal{P} \subseteq \mathrm{B}(\mathbb{C}^d)$, which is a subset of the set of quantum states, and a dissimilarity measure d_* on quantum states (cf. Section 2.5), the property testing task is to:

   decide correctly if $\rho \in \mathcal{P}$ or $d_*(\rho, \mathcal{P}) = \inf_{\sigma \in \mathcal{P}} d_*(\rho, \sigma) \ge \epsilon$ with probability at least $1 - \delta$.

   If the algorithm decides that $\rho \in \mathcal{P}$, we say that it *accepts* ρ. Otherwise, we say that the algorithm *rejects* ρ.

   An algorithm to complete an estimation task can be given as a quantum event E on $(\mathbb{C}^d)^{\otimes n}$ satisfying:

$$\rho \in \mathcal{P} \implies \mathop{\mathbf{E}}_{\rho^{\otimes n}} [E] \ge 1 - \delta,$$
$$d_*(\rho, \mathcal{P}) \ge \epsilon \implies \mathop{\mathbf{E}}_{\rho^{\otimes n}} [E] \le \delta.$$

To complete such an estimation or property testing task, an algorithm needs to measure a number n of copies $\rho^{\otimes n}$ of the unknown state ρ. The problem is to design algorithms that minimize n as a function of the parameters d, ϵ, and δ, *ignoring constant factors*.

For property testing tasks, the error probability δ can always be easily improved:

Remark 2.7.2. Given an algorithm $\mathcal{A}$ that completes a property testing task $\mathcal{T}(d, \epsilon, 1/3)$ using n copies, one can run $\mathcal{A}$ independently N times, using $n \cdot N$ copies of ρ in total, and expect that $2N/3$ of the outputs (which are either ACCEPT or REJECT) to be correct. Thus, picking the majority output yields an answer that, by standard amplification techniques, is correct with probability at least $1 - O(e^{-N})$. Therefore, $\mathcal{A}$ can be easily converted to an algorithm $\mathcal{A}'$ that completes the task $\mathcal{T}(d, \epsilon, \delta)$ using $O(n \cdot \log(1/\delta))$ copies of ρ for any $\delta > 0$.

2.8. Miscellaneous lemmas

In this section, we prove a few key lemmas needed in the following chapters. Let $\mathcal{H}$ denote a finite-dimensional Hilbert space.

2.8.1. A special version of Chebyshev's inequality. Our main property testing results in Chapter 3 depend on a special version of Chebyshev's inequality proved below:

Lemma 2.8.1. *Fix $\mu > 0$ and let $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}$ denote a sequence of random variables with $\mathbf{E}[\boldsymbol{X}^{(i)}] = \mu$ for all $i = 1, \ldots, n$. Let $b(x)$ and $v(x)$ be functions such that:*

i. $b(x)$ and $v(x)$ are nondecreasing on $[0, \infty)$;
ii. $b(x)/x^2$ and $v(x)/x^2$ are nonincreasing on $[0, \infty)$.

If the following upper bound holds,

$$\mathbf{Var}[\boldsymbol{X}^{(n)}] \leq O\left(\frac{b(\mu)}{n^2} + \frac{v(\mu)}{n}\right), \tag{10}$$

where the constant implicit in the $O(\_)$ notation does not depend on n, then there exists a test that distinguishes with high probability between the case $\mu \leq 0.99\theta$ and $\mu > \theta$ provided n is sufficiently large, viz.

$$n \geq C \max\left\{\sqrt{\frac{b(\theta)}{\theta^2}}, \frac{v(\theta)}{\theta}\right\}. \tag{11}$$

Proof. The test checks if $\boldsymbol{X}^{(n)}$ is above or below a certain threshold:

$$\boldsymbol{X}^{(n)} \leq 0.995\theta \implies \text{output "}\mu \leq 0.99\theta\text{"}$$
$$\boldsymbol{X}^{(n)} > 0.995\theta \implies \text{output "}\mu > \theta\text{"}.$$

To prove correctness, we show that by adjusting the constant C in the lower bound Equation (11), the standard deviation $\mathbf{stddev}[\boldsymbol{X}^{(n)}]$ can be made arbitrarily small. Thus, if μ is sufficiently distant from the testing threshold 0.995θ, then the probability of an error is low by Chebyshev's inequality.

Suppose $\mu \leq 0.99\theta$. Since $b(x)$ and $v(x)$ are nondecreasing,

$$\frac{b(\mu)}{n^2} + \frac{v(\mu)}{n} \leq \frac{b(\theta)}{n^2} + \frac{v(\theta)}{n}.$$

By Equation (11), $n \geq C\sqrt{b(\theta)}/\theta$ and $n \geq Cv(\theta)/\theta$, so

$$\frac{b(\mu)}{n^2} + \frac{v(\mu)}{n} \leq \frac{b(\theta)}{n^2} + \frac{v(\theta)}{n} \leq \frac{\theta^2}{C^2} + \frac{\theta}{C}.$$

Thus, by Equation (10),

$$\mathbf{Var}[\boldsymbol{X}^{(n)}] \leq O\left(\frac{1}{C^2} + \frac{1}{C}\right)\theta^2.$$

Hence, for C sufficiently large, $\mathbf{stddev}[\boldsymbol{X}^{(n)}] \leq 0.001\theta$. Since $\mu \leq 0.99\theta$, the output of the test is only incorrect if $\boldsymbol{X}^{(n)} - \mu > 0.005\theta \geq 5 \cdot \mathbf{stddev}[\boldsymbol{X}^{(n)}]$. By Chebyshev's inequality, the probability of being at least five standard deviations away from the mean is at most $1/25$.

Suppose $\mu > \theta$. Since $b(x)/x^2$ and $v(x)/x^2$ are nonincreasing on $[0, \infty)$, it holds that

$$\frac{b(\mu)}{\mu^2} \leq \frac{b(\theta)}{\theta^2} \text{ and } \frac{v(\mu)}{\mu^2} \leq \frac{v(\theta)}{\theta^2}.$$

Thus, by Equation (11),

$$n \geq C \max\left\{\sqrt{\frac{b(\mu)}{\mu^2}}, \frac{v(\mu)}{\mu^2}\right\}.$$

Hence, by Equation (10),

$$\mathbf{Var}[\boldsymbol{X}^{(n)}] \leq O\left(\frac{b(\mu)}{n^2} + \frac{v(\mu)}{n}\right) \leq O\left(\frac{1}{C^2} + \frac{1}{C}\right)\mu^2.$$

Therefore, for C sufficiently large, $\mathbf{stddev}[\boldsymbol{X}^{(n)}] \leq 0.001\mu$. Since $\mu > \theta$ by assumption, it holds that $0.995\mu > 0.995\theta$, so the output of the test is only incorrect if $\mu - \boldsymbol{X}^{(n)} > 0.005\mu \geq 5 \cdot \mathbf{stddev}[\boldsymbol{X}^{(n)}]$. By Chebyshev's inequality, this occurs with probability at most $1/25$. ■

2.8.2. Lifting quantum events to projections. Due to their idempotent property, projections are simpler to work with than quantum events. The following result, a special case of Naimark's dilation theorem [43], shows that every quantum event can be lifted to a projection at the expense of adding a qubit component (i.e. a quantum state on $\mathbb{C}^2$) to the quantum system.

Theorem 2.8.2 (Naimark)**.** *For any quantum event $A \in \mathrm{B}(\mathbb{C}^d)$, there exists a projection $\Pi \in \mathrm{B}(\mathbb{C}^d \otimes \mathbb{C}^2)$ such that, for any quantum state $\rho \in \mathrm{B}(\mathbb{C}^d)$,*

$$\mathop{\mathbf{E}}_{\rho\otimes|0\rangle\langle 0|}[\Pi] = \mathop{\mathbf{E}}_{\rho}[A].$$

2.8.3. A simpler formula for fidelity. Below we give a simpler formula for the fidelity $\mathrm{F}(\rho, \sigma)$ when σ is a conditioned version of ρ (such results are sometimes known under the name "gentle measurement"; see e.g. [57, Corollary 3.15]):

Lemma 2.8.3. *For all positive operators $X, Y \in \mathrm{B}(\mathcal{H})$, $\mathrm{F}(X, YXY) = \langle X, Y\rangle$.*

Proof. Since $\sqrt{X}Y\sqrt{X} \geq 0$ and

$$\sqrt{X}YXY\sqrt{X} = (\sqrt{X}Y\sqrt{X}) \cdot (\sqrt{X}Y\sqrt{X}) = (\sqrt{X}Y\sqrt{X})^2,$$

it follows that

$$\begin{aligned}\mathrm{F}(X, YXY) &= \mathrm{tr}\left(\sqrt{\sqrt{X}YXY\sqrt{X}}\right) = \mathrm{tr}\left(\sqrt{(\sqrt{X}Y\sqrt{X})^2}\right) \\ &= \mathrm{tr}\left(\sqrt{X}Y\sqrt{X}\right) = \mathrm{tr}(XY) = \langle X, Y\rangle.\end{aligned}$$

■

Corollary 2.8.4. *Let $A \in \mathrm{B}(\mathcal{H})$ be a quantum event. For all quantum states $\rho \in \mathrm{B}(\mathcal{H})$,*

$$\mathrm{F}(\rho, \rho|_{\sqrt{A}}) = \frac{\mathbf{E}_\rho[\sqrt{A}]}{\sqrt{\mathbf{E}_\rho[A]}}.$$

Moreover, if $A = \Pi$ is a projection, then

$$\mathrm{F}(\rho, \rho|_\Pi) = \sqrt{\mathop{\mathbf{E}}_{\rho}[\Pi]}.$$

Proof. By Lemma 2.8.3,

$$\mathrm{F}(\rho, \rho|_{\sqrt{A}}) = \mathrm{F}\left(\rho, \frac{\sqrt{A}\rho\sqrt{A}}{\mathbf{E}_\rho[A]}\right) = \left\langle \rho, \frac{\sqrt{A}}{\sqrt{\mathbf{E}_\rho[A]}} \right\rangle = \frac{\mathbf{E}_\rho[\sqrt{A}]}{\sqrt{\mathbf{E}_\rho[A]}}.$$

If $A = \Pi$ is a projection, then $\sqrt{A} = \sqrt{\Pi} = \Pi$. ■

Below we give a further formula for $\mathrm{F}(\rho, \rho|_{\sqrt{A}})$ in terms of the Bhattacharyya coefficient using the spectral decomposition of A. This result may be obtained as a special case of a theorem of Fuchs and Caves [20].

Proposition 2.8.5. *Let $\rho \in \mathrm{B}(\mathcal{H})$ be a quantum state and let $A \in \mathrm{B}(\mathcal{H})$ be a quantum event with spectral decomposition $A = \sum_{i=1}^k \lambda_i \Pi_i$ as in Theorem 2.1.3.*

If p is the probability distribution on $[k]$ determined by the measurement $\mathcal{M} = (\Pi_1, \ldots, \Pi_k)$ on ρ and q is the one determined by $\mathcal{M}$ on $\rho|_{\sqrt{A}}$, then

$$\mathrm{F}(\rho, \rho|_{\sqrt{A}}) = \mathrm{BC}(p, q).$$

Proof. By definition, $p_i = \mathbf{E}_\rho[\Pi_i]$ and

$$q_i = \mathop{\mathbf{E}}_{\rho|_{\sqrt{A}}}[\Pi] = \frac{\mathrm{tr}(\sqrt{A}\rho\sqrt{A}\Pi)}{\mathbf{E}_\rho[A]} = \frac{\mathrm{tr}(\lambda_i \rho \Pi)}{\mathbf{E}_\rho[A]} = \frac{\lambda_i p_i}{\mathbf{E}_\rho[A]},$$

for all $i = 1, \ldots, k$. Hence, by Corollary 2.8.4,

$$\begin{aligned}
\mathrm{F}(\rho, \rho|_{\sqrt{A}}) &= \frac{\mathbf{E}_\rho[\sqrt{A}]}{\sqrt{\mathbf{E}_\rho[A]}} \\
&= \frac{\mathbf{E}_\rho\left[\sum_{i=1}^k \sqrt{\lambda_i}\Pi_i\right]}{\sqrt{\mathbf{E}_\rho[A]}} \\
&= \frac{\sum_{i=1}^k \sqrt{\lambda_i}\,\mathbf{E}_\rho[\Pi_i]}{\sqrt{\mathbf{E}_\rho[A]}} \\
&= \frac{\sum_{i=1}^k \sqrt{\lambda_i} p_i}{\sqrt{\mathbf{E}_\rho[A]}} = \sum_{i=1}^k \sqrt{\frac{\lambda_i p_i}{\mathbf{E}_\rho[A]}} \cdot \sqrt{p_i} = \mathrm{BC}(p, q).
\end{aligned}$$ ■

2.8.4. Naive expectation estimation. The following lemma provides a simple method for estimating the expected value of a quantum event E with respect to a quantum state ρ using n copies of ρ.

Lemma 2.8.6. *Let $E \in \mathrm{B}(\mathcal{H})$ be a quantum event and let $0 < \epsilon, \delta < \frac{1}{2}$. Then there exists $n = O(\log(1/\delta)/\epsilon^2)$ (not depending on E) and a measurement $\mathcal{M} = (A_0, \ldots, A_n)$ such that, for any quantum state $\rho \in \mathrm{B}(\mathcal{H})$,*

$$\mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \mathrm{tr}(\rho E)\right| > \epsilon\right] \le \delta,$$

where $\boldsymbol{k} \in \{0, \ldots, n\}$ is the random outcome of the measurement $\mathcal{M}$ applied to the state $\rho^{\otimes n}$.

Moreover, for any parameters $0 \le \tau, c \le 1$, there exists a quantum event B such that

$$|\mathrm{tr}(\rho E) - \tau| > c + \epsilon \implies \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] \ge 1 - \delta \text{ and}$$
$$|\mathrm{tr}(\rho E) - \tau| \le c - \epsilon \implies \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] \le \delta.$$

Additionally, if E is a projector, then so is B.

Proof. Let $E_1 = E$ and $E_0 = \mathbf{1} - E$. For all $x \in \{0,1\}^n$, let $E_x \in \mathrm{B}(\mathcal{H})^{\otimes n}$ be defined by $E_x = E_{x_1} \otimes E_{x_2} \otimes \cdots \otimes E_{x_n}$. For $k = 0, \ldots, n$, let $A_k \in \mathrm{B}(\mathcal{H})^{\otimes n}$ be the quantum event defined by

$$A_k = \sum_{\substack{x \in \{0,1\}^n \\ |x| = k}} E_x.$$

Let $\mathcal{M}$ be the measurement defined by $\mathcal{M} = \{A_0, \ldots, A_n\}$.

Thus, if $\boldsymbol{k} \in \{0, \ldots, n\}$ is the random outcome of measuring $\rho^{\otimes n}$ according to $\mathcal{M}$, then $\boldsymbol{k}$ is distributed as $\mathrm{Binomial}(n, \mathrm{tr}(\rho E))$. Hence, if $n = O(\log(1/\delta)/\epsilon^2)$, then, by Hoeffding's inequality [31],

$$\mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \mathrm{tr}(\rho E)\right| \ge \epsilon\right] \le 2\exp(-2n\epsilon^2) \le \delta.$$

Let parameters $\tau, c \in [0,1]$ be given and let the function $f : [0,1] \to \{0,1\}$ be defined by

$$f(t) = \begin{cases} 1, & |t - \tau| \ge c, \\ 0, & \text{otherwise.} \end{cases}$$

Finally, let the quantum event B be defined by

$$B = \sum_{k=0}^{n} f(k/n) A_k.$$

Thus, if $\boldsymbol{k} \sim \mathrm{Binomial}(n, \mathrm{tr}(\rho E))$, then

$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] = \sum_{k=0}^{n} \mathbf{P}[\boldsymbol{k} = k] \cdot f(k/n) = \mathbf{E}[f(\boldsymbol{k}/n)] = \mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \tau\right| \ge c\right].$$

If $c + \epsilon \le |\mathrm{tr}(\rho E) - \tau|$, then $|\mathrm{tr}(\rho E) - \boldsymbol{k}/n| < \epsilon$ implies $|\boldsymbol{k}/n - \tau| \ge c$. Hence,

$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] = \mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \tau\right| \ge c\right] \ge \mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \mathrm{tr}(\rho E)\right| < \epsilon\right] \ge 1 - \delta.$$

If $c - \epsilon \ge |\mathrm{tr}(\rho E) - \tau|$, then $|\mathrm{tr}(\rho E) - \boldsymbol{k}/n| < \epsilon$ implies $|\boldsymbol{k}/n - \tau| < c$. Hence,

$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[\overline{B}] = \mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \tau\right| < c\right] \ge \mathbf{P}\left[\left|\frac{\boldsymbol{k}}{n} - \mathrm{tr}(\rho E)\right| < \epsilon\right] \ge 1 - \delta.$$

If E is a projector, then A_k is a projector and $A_k A_\ell = A_\ell A_k = 0$ for all $k, \ell \in \{0, \ldots, n\}$. Since B is a sum of orthogonal projectors A_k with $k \in \{0, \ldots, n\}$, it follows that B is a projector. ■

2.8.5. The damage lemma. The following result is part of the "Damage Lemma" of Aaronson and Rothblum [4, Lemma 17]. Since the original proof of the "Damage Lemma" was found to be incorrect [39], we provide a slightly different proof by induction below:

Lemma 2.8.7. *Let $S_1, \dots, S_m$ be arbitrary quantum operations on d-dimensional quantum states. Let ρ be a quantum state on $\mathbb{C}^d$ with $p_i = \operatorname{tr}(S_i(\rho)) > 0$ for all $i \in [m]$. It holds that*

$$|\operatorname{tr}(S_m(\cdots S_1(\rho))) - p_1 \cdots p_m| \leq 2 \cdot \sum_{k=1}^{m-1} p_1 \cdots p_k \cdot d_{\operatorname{tr}}\left(\frac{S_k(\rho)}{\operatorname{tr}(S_k(\rho))}, \rho \right).$$

Proof. For all $k \in [m]$, let $p_{[k]} = p_1 \cdots p_k$ and $\sigma_k = S_k(\rho) / \operatorname{tr}(S_k(\rho))$. For all self-adjoint matrices X, $|\operatorname{tr}(X)| \leq \|X\|_1$ and $\|S(X)\|_1 \leq \|X\|_1$ for all quantum operations S. Hence,

$$\begin{aligned}
|\operatorname{tr}(S_m(\cdots S_1(\rho))) - p_{[m]}| &= |\operatorname{tr}(S_m(\cdots S_1(\rho))) - p_{[m-1]} \operatorname{tr}(S_m(\rho))| \\
&= |\operatorname{tr}(S_m(\cdots S_1(\rho)) - p_{[m-1]} S_m(\rho))| \\
&= |\operatorname{tr}(S_m(S_{m-1}(\cdots S_1(\rho)) - p_{[m-1]}\rho))| \\
&\leq \|S_m(S_{m-1}(\cdots S_1(\rho)) - p_{[m-1]}\rho)\|_1 \\
&\leq \|S_{m-1}(\cdots S_1(\rho)) - p_{[m-1]}\rho\|_1 \\
&\leq \|S_{m-1}(\cdots S_1(\rho)) - p_{[m-1]}\sigma_{m-1}\|_1 + \|p_{[m-1]}\sigma_{m-1} - p_{[m-1]}\rho\|_1 \\
&= \|S_{m-1}(\cdots S_1(\rho)) - p_{[m-2]} S_{m-1}(\rho)\|_1 + 2p_{[m-1]} d_{\operatorname{tr}}(\sigma_{m-1}, \rho) \\
&\leq \|S_{m-2}(\cdots S_1(\rho)) - p_{[m-2]}\rho\|_1 + 2p_{[m-1]} d_{\operatorname{tr}}(\sigma_{m-1}, \rho).
\end{aligned}$$

Note that $\|S_1(\rho) - p_1\rho\|_1 = p_1\|\sigma_1 - \rho\|_1 = 2p_{[1]} d_{\operatorname{tr}}(\sigma_1, \rho)$. Therefore, by induction,

$$\begin{aligned}
|\operatorname{tr}(S_m(\cdots S_1(\rho))) - p_{[m]}| &\leq \|S_{m-2}(\cdots S_1(\rho)) - p_{[m-2]}\rho\|_1 + 2p_{[m-1]} d_{\operatorname{tr}}(\sigma_{m-1}, \rho) \\
&\leq 2 \cdot \sum_{k=1}^{m-2} p_{[k]} \cdot d_{\operatorname{tr}}(\sigma_k, \rho) + 2p_{[m-1]} d_{\operatorname{tr}}(\sigma_{m-1}, \rho) \\
&\leq 2 \cdot \sum_{k=1}^{m-1} p_{[k]} \cdot d_{\operatorname{tr}}(\sigma_k, \rho).
\end{aligned}$$

■

Lemma 2.8.7 compares the probability $\operatorname{tr}(S_1(\rho)) \cdots \operatorname{tr}(S_m(\rho))$ that the operations $S_1, \dots, S_m$ accept the same state ρ independently with the probability $\operatorname{tr}(S_m(\cdots S_1(\rho)))$ that all $S_1, \dots, S_m$ accept when applied sequentially to the initial state ρ.

2.8.6. Quantum union bound-style results. The following inequality, which appears in the proof of [44, Theorem 1.3], will be used in Chapter 5 to show that when $S_1, \dots, S_m$ are applied sequentially to the initial state ρ, the probability of observing $S_1, \dots, S_{t-1}$ accept and S_t reject for certain "good" values of $t \in [m]$ is bounded below by a positive constant for specific ρ and $S_1, \dots, S_m$.

Lemma 2.8.8. *Let ρ be a mixed quantum state and let $A_1, \dots, A_m$ denote quantum events on $\mathbb{C}^d$ with $\mathbf{E}_\rho[A_i] > 0$ for all $i \in [m]$. Let $p_0 = 1$, $q_0 = 1$, $\rho_0 = \rho$, $p_i = 1 - \mathbf{E}_\rho[A_i]$, and $\rho_i = \rho_{i-1}|_{\sqrt{A_i}}$ for all $i \in [m]$.*

Suppose the measurements $(A_1, \overline{A}_1), \ldots, (A_m, \overline{A}_m)$ are applied to ρ sequentially; for all $t \in [m]$, let q_t denote the probability of observing outcomes $A_1, \ldots, A_t$ and let s_t denote the probability of observing outcomes $A_1, \ldots, A_{t-1}, \overline{A}_t$. It holds that

$$1 \leq \sqrt{q_m}\,\mathrm{F}(\rho, \rho_m) + \sum_{i=1}^{m} \sqrt{s_i}\sqrt{p_i}.$$

Proof. Since $1 = q_0\,\mathrm{F}(\rho, \rho_0)$ and $q_i = q_{i-1} \cdot \mathbf{E}_{\rho_{i-1}}[A_i]$ for all $i \in [m]$,

$$\begin{aligned} 1 - \sqrt{q_m}\,\mathrm{F}(\rho, \rho_m) &= \sum_{i=1}^{m} (\sqrt{q_{i-1}}\,\mathrm{F}(\rho, \rho_{i-1}) - \sqrt{q_i}\,\mathrm{F}(\rho, \rho_i)) \\ &= \sum_{i=1}^{m} \left(\sqrt{q_{i-1}}\,\mathrm{F}(\rho, \rho_{i-1}) - \sqrt{q_{i-1}} \sqrt{\mathop{\mathbf{E}}_{\rho_{i-1}}[A_i]\,\mathrm{F}(\rho, \rho_i)} \right) \\ &= \sum_{i=1}^{m} \sqrt{q_{i-1}} \left(\mathrm{F}(\rho, \rho_{i-1}) - \sqrt{\mathop{\mathbf{E}}_{\rho_{i-1}}[A_i]\,\mathrm{F}(\rho, \rho_i)} \right). \end{aligned}$$

By [44, Lemma 2.1] and the inequality $\mathbf{1} - \sqrt{A_i} \leq \overline{A}_i$,

$$\mathrm{F}(\rho, \rho_{i-1}) - \sqrt{\mathop{\mathbf{E}}_{\rho_{i-1}}[A_i]\,\mathrm{F}(\rho, \rho_i)} \leq \sqrt{\mathop{\mathbf{E}}_{\rho}[\mathbf{1} - \sqrt{A_i}]} \sqrt{\mathop{\mathbf{E}}_{\rho_{i-1}}[\mathbf{1} - \sqrt{A_i}]} \leq \sqrt{\mathop{\mathbf{E}}_{\rho}[\overline{A}_i]} \sqrt{\mathop{\mathbf{E}}_{\rho_{i-1}}[\overline{A}_i]}.$$

Hence,

$$1 - \sqrt{q_m}\,\mathrm{F}(\rho, \rho_m) \leq \sum_{i=1}^{m} \sqrt{q_{i-1}} \sqrt{\mathop{\mathbf{E}}_{\rho}[\overline{A}_i]} \sqrt{\mathop{\mathbf{E}}_{\rho_{i-1}}[\overline{A}_i]} \leq \sum_{i=1}^{m} \sqrt{s_i}\sqrt{p_i}.$$ ■

Finally, for the "unique decoding" part of our Hypothesis Selection routine (cf. Chapter 5), we will use a related result, Gao's *quantum union bound* [21]:

Lemma 2.8.9. *Let $\Pi_1, \ldots, \Pi_m \in \mathrm{B}(\mathcal{H})$ be projections. For any quantum state $\rho \in \mathrm{B}(\mathcal{H})$,*

$$\mathop{\mathbf{E}}_{\rho}[(\Pi_1 \cdots \Pi_m)(\Pi_1 \cdots \Pi_m)^\dagger] \geq 1 - 4 \sum_{i=1}^{m} \mathop{\mathbf{E}}_{\rho}[\overline{\Pi}_i].$$

Corollary 2.8.10. *In the setting of Lemma 2.8.9, suppose $\mathbf{E}_\rho[\Pi_i] \geq 1 - \epsilon$ for all $1 \leq i \leq m$. If an algorithm sequentially measures ρ with $(\Pi_1, \overline{\Pi}_1), \ldots, (\Pi_m, \overline{\Pi}_m)$, then the probability that the measurement outcomes are precisely $\Pi_1, \ldots, \Pi_m$ is at least $1 - 4\epsilon m$.*

CHAPTER 3

Quantum state certification

In this chapter, we consider property testing and estimation problems related to the *quantum state certification* problem: given measurement access to copies of an unknown quantum state ρ and either a mathematical description (e.g. the density matrix) or measurement access to copies of another (possibly unknown) quantum state σ, the state certification task is to, with high probability, distinguish between $\rho = \sigma$ or $d_*(\rho, \sigma) \geq \epsilon$ for a given quantum dissimilarity measure d_* (cf. Section 2.5).

We show that $O(1/\epsilon^2)$ copies are sufficient to complete the state certification task with respect to Hilbert–Schmidt distance:

Theorem 3.0.1. *There is an algorithm that, given $n = O(1/\epsilon^2)$ copies each of unknown mixed quantum states $\rho, \sigma \in \mathrm{B}(\mathbb{C}^d)$, with high probability distinguishes between $d_{\mathrm{HS}}(\rho, \sigma) \leq 0.99\epsilon$ or $d_{\mathrm{HS}}(\rho, \sigma) > \epsilon$.*

As a consequence of Theorem 3.0.1, we obtain an algorithm for state certification with respect to trace distance using $O(d/\epsilon^2)$ copies:

Corollary 3.0.2. *There is an algorithm that, given $n = O(d/\epsilon^2)$ copies each of unknown mixed quantum states $\rho, \sigma \in \mathrm{B}(\mathbb{C}^d)$, with high probability distinguishes between $d_{\mathrm{tr}}(\rho, \sigma) = 0$ or $d_{\mathrm{tr}}(\rho, \sigma) > \epsilon$.*

Furthermore, if at least one of the states measured is close to a state of rank k, then the number of copies needed is $O(k/\epsilon^2)$, an improvement if k is significantly smaller than d:

Theorem 3.0.3. *If either ρ or σ is close to having rank k, in the sense that the sum of its largest k eigenvalues is at least $1 - \delta$, then there is an algorithm that, given $n = O(k/\epsilon^2)$ copies of each $\rho, \sigma \in \mathrm{B}(\mathbb{C}^d)$, with high probability distinguishes between $d_{\mathrm{HS}}(\rho, \sigma) \leq 0.58\epsilon/\sqrt{k}$ or $d_{\mathrm{tr}}(\rho, \sigma) > \epsilon + \delta$.*

Additionally, we show that state certification with respect to Bures χ^2-divergence can be completed with $O(d/\epsilon^2)$ copies:

Theorem 3.0.4. *Let $\sigma \in \mathrm{B}(\mathbb{C}^d)$ be a known quantum state with smallest eigenvalue at least $c\epsilon^2/d$ for some $c > 0$. There is an algorithm that, given $n = O(d/\epsilon^2)$ copies of ρ, with high probability distinguishes between $d_{\chi^2}(\rho, \sigma) \leq 0.99\epsilon^2$ and $d_{\chi^2}(\rho, \sigma) > \epsilon^2$.*

As a consequence of Theorem 3.0.4, we obtain a state certification algorithm with respect to fidelity:

Corollary 3.0.5. *Let $\sigma \in \mathrm{B}(\mathbb{C}^d)$ be a known quantum state. There is an algorithm that, given $n = O(d/\epsilon)$ copies of ρ, with high probability distinguishes between $\mathrm{F}(\rho, \sigma) < 1 - \epsilon$ and $d_{\chi^2}(\rho, \sigma) \leq 0.49\epsilon$.*

This chapter draws on material originally published in [12].

3.1. Efficient quantum estimators

The connection between observables and random variables presented in Section 2.3 allows us to import notions from classical statistics into the quantum setting. In this section, this connection is used to define (unbiased) quantum estimators and introduce the notion of statistical efficiency of a quantum estimator. These notions are used to formulate a structure theorem for efficient quantum estimators in situations where the statistic of interest is unitarily invariant.

Let V be a finite-dimensional Hilbert space. Let S denote a set of quantum states on V and let $f : S \to \mathbb{R}$ be a statistic on S. The set S serves to restrict an estimation problem to a particular class of quantum states. S will be gradually restricted, as needed, from an arbitrary set of quantum states to a set of multipartite quantum states of the form $\rho^{\otimes n}$ or $\rho^{\otimes m} \otimes \sigma^{\otimes n}$, where ρ and σ are quantum states on $\mathbb{C}^d$.

Definition 3.1.1. An *estimator* for f is an observable $\mathcal{O} \in \mathrm{End}(V)$ such that $\mathbf{E}_\varrho[\mathcal{O}] = f(\varrho)$ for all $\varrho \in S$. An estimator $\mathcal{O}$ is *efficient* if $\mathbf{Var}_\varrho[\mathcal{O}] \leq \mathbf{Var}_\varrho[\mathcal{O}']$ for all estimators $\mathcal{O}' \in \mathrm{End}(V)$ for f.

Henceforth, fix $V = (\mathbb{C}^d)^{\otimes n}$ and let S denote the set of states of the form $\rho_1 \otimes \cdots \otimes \rho_n$, where $\rho_1, \ldots, \rho_n$ are quantum states on $\mathbb{C}^d$.

Definition 3.1.2. A statistic $f : S \to \mathbb{R}$ is *unitarily invariant* if $f \circ \mathrm{Ad}_U = f$ for all $U \in \mathrm{U}(d)$. An observable $\mathcal{O} \in \mathrm{End}(V)$ is *unitarily invariant* if $\mathrm{Ad}_U(\mathcal{O}) = \mathcal{O}$ for all $U \in \mathrm{U}(d)$.

Let Φ be the map on observables $\mathcal{O} \in \mathrm{End}(V)$ defined by

$$\Phi(\mathcal{O}) = \int_{\mathrm{U}(d)} \mathrm{Ad}_U(\mathcal{O})\, dU,$$

where dU denotes the Haar measure on $\mathrm{U}(d)$. Note that Φ preserves self-adjointness and, hence, maps observables to observables.

Proposition 3.1.3. *If $\mathcal{O}$ is an estimator for a unitarily invariant statistic f, then $\Phi(\mathcal{O})$ is also an estimator for f, and $\mathbf{Var}_\varrho[\Phi(\mathcal{O})] \leq \mathbf{Var}_\varrho[\mathcal{O}]$ for all $\varrho \in S$.*

Proof. The map Φ is positive and unital. Since f is unitarily invariant,

$$\mathop{\mathbf{E}}_\varrho[\Phi(\mathcal{O})] = \int_{\mathrm{U}(d)} \mathrm{tr}(\mathrm{Ad}_{U^\dagger}(\varrho)\mathcal{O})\, dU = \int_{\mathrm{U}(d)} f(\mathrm{Ad}_{U^\dagger}(\varrho))\, dU = \int_{\mathrm{U}(d)} f(\varrho)\, dU = \mathop{\mathbf{E}}_\varrho[\mathcal{O}].$$

Hence, by Lemma 2.3.5, $\mathbf{Var}_\varrho[\Phi(\mathcal{O})] \leq \mathbf{Var}_\varrho[\mathcal{O}]$. ■

The following result relates the image of the map Φ to the symmetric group algebra $\mathbb{C}\mathfrak{S}_n$ and the representation $\mathcal{P}$, using the Schur–Weyl duality theorem. For a proof, see e.g. [14, Proposition 2.2].

Proposition 3.1.4. *The map Φ is a projection into $\mathcal{P}(\mathbb{C}\mathfrak{S}_n)$.*

Thus, by Proposition 3.1.3, if $\mathcal{O}$ is an efficient estimator for a unitarily invariant statistic f, then $\Phi(\mathcal{O})$ is also an efficient estimator for f. Hence, the next corollary follows immediately from Proposition 3.1.4.

Corollary 3.1.5. *To find an efficient estimator for a unitarily invariant statistic $f : S \to \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X \in \mathbb{C}\mathfrak{S}_n$.*

In light of Corollary 3.1.5, we introduce the following (abuse of) notation:

Notation 3.1.6. Let $\mathbf{E}_\varrho$ be extended to a map on elements $X \in \mathbb{C}\mathfrak{S}_n$ defined by $\mathbf{E}_\varrho[X] = \mathbf{E}_\varrho[\mathcal{P}(X)]$. Thus, $\mathbf{E}_\varrho$, $\mathbf{Cov}_\varrho$, and $\mathbf{Var}_\varrho$ are defined directly on elements of $\mathbb{C}\mathfrak{S}_n$ via the representation $\mathcal{P}$.

If $\gamma = (i_1\ i_2\ \cdots\ i_\ell) \in \mathfrak{S}_n$, let tr_γ be defined by $\mathrm{tr}_\gamma(\varrho) = \mathrm{tr}(\rho_{i_1}\rho_{i_2}\cdots\rho_{i_\ell})$. The following proposition establishes a formula for the expectation $\mathbf{E}_\varrho[\pi]$ of a permutation $\pi \in \mathfrak{S}_n$ with respect to a state $\varrho \in S$. (Caution: π is not in general an observable.)

Proposition 3.1.7. *Let $\pi \in \mathfrak{S}_n$ be an arbitrary permutation. If $\pi = \gamma_1 \ldots \gamma_k$ is a decomposition of π into disjoint cycles, including cycles of length 1, then*

$$\mathop{\mathbf{E}}_{\varrho}[\pi^{-1}] = \prod_{i=1}^{k} \mathrm{tr}_{\gamma_i}(\varrho).$$

Proof. In light of the tensorization property of $\mathbf{E}_\varrho$ and the fact that ϱ is an n-partite quantum state, the problem reduces immediately to the case when π is an n-cycle. Without loss of generality, suppose $\pi = (1\ 2\ \ldots\ n)$. Thus,

$$\begin{aligned}\mathrm{tr}(\rho_1\rho_2\cdots\rho_n) &= \sum_{v\in[d]^n} \langle v_1|\rho_1|v_2\rangle \cdots \langle v_n|\rho_n|v_1\rangle \\ &= \sum_{v\in[d]^n} \langle v_1|\rho_1|\pi(v)_1\rangle \cdots \langle v_n|\rho_n|\pi(v)_n\rangle \\ &= \sum_{v\in[d]^n} \langle v|\varrho|\pi(v)\rangle = \mathrm{tr}(\varrho\,\mathcal{P}(\pi^{-1})) = \mathop{\mathbf{E}}_{\varrho}[\pi^{-1}]. \qquad \blacksquare\end{aligned}$$

Remark 3.1.8. In describing the cycle type of a permutation $\pi \in \mathfrak{S}_n$, it is common to omit mentioning 1-cycles. Conveniently, this would have no effect in Proposition 3.1.7, since $\mathrm{tr}(\rho_i) = 1$ anyway for all i.

Definition 3.1.9. The group Γ of *permutation invariants* of the set of states S is defined by

$$\Gamma = \{\pi \in \mathfrak{S}_n \mid \forall \varrho \in S,\ \forall X \in \mathbb{C}\mathfrak{S}_n,\ \mathop{\mathbf{E}}_{\varrho}[\pi^{-1}X\pi] = \mathop{\mathbf{E}}_{\varrho}[X]\}.$$

Note that the definition of Γ depends on S. For all $X \in \mathbb{C}\mathfrak{S}_n$, let $X^\Gamma \in \mathbb{C}\mathfrak{S}_n$ be defined by

$$X^\Gamma = \frac{1}{|\Gamma|}\sum_{\pi\in\Gamma} \pi^{-1}X\pi.$$

Thus, $X^\Gamma\tau = \tau X^\Gamma$ for all $\tau \in \Gamma$ and $X \in \mathbb{C}\mathfrak{S}_n$.

Proposition 3.1.10. *For all self-adjoint elements $\mathcal{O} \in \mathbb{C}\mathfrak{S}_n$, $\mathbf{Var}_\varrho[\mathcal{O}^\Gamma] \le \mathbf{Var}_\varrho[\mathcal{O}]$.*

Proof. The map $\mathcal{O} \mapsto \mathcal{O}^\Gamma$ is positive and unital. Moreover, $\mathbf{E}_\varrho[\mathcal{O}^\Gamma] = \mathbf{E}_\varrho[\mathcal{O}]$ for all $\mathcal{O} \in \mathbb{C}\mathfrak{S}_n$. Hence, by Lemma 2.3.5, $\mathbf{Var}_\varrho[\mathcal{O}^\Gamma] \le \mathbf{Var}_\varrho[\mathcal{O}]$. $\blacksquare$

Corollary 3.1.11. *To find an efficient estimator for a unitarily invariant statistic $f : S \to \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X \in \mathbb{C}\mathfrak{S}_n$ and $X\tau = \tau X$ for all $\tau \in \Gamma$.*

The group Γ acts on $\mathfrak{S}_n$ by conjugation, viz. $\tau \in \Gamma$ acts on $\mathfrak{S}_n$ by $\pi \mapsto \tau^{-1}\pi\tau$. This action partitions the group $\mathfrak{S}_n$ into disjoint orbits: $\mathfrak{S}_n = O_1 \cup \cdots \cup O_\ell$, where two permutations π_1 and π_2 belong to the same orbit O_i for $i \in [\ell]$ if and only if there exists $\tau \in \Gamma$ such that $\tau^{-1}\pi_1\tau = \pi_2$. It is easy to see that an element $X \in \mathbb{C}\mathfrak{S}_{m+n}$ commutes with all elements of Γ if and only if X is constant on the orbits $O_1, \ldots, O_\ell$ defined by Γ. Let $\phi_i \in \mathbb{C}\mathfrak{S}_{m+n}$ denote the indicator function of the orbit O_i for $i \in [\ell]$. Thus, the set $\{\phi_1, \ldots, \phi_\ell\}$ forms a linear basis for the elements $X \in \mathbb{C}\mathfrak{S}_n$ that are constant on the orbits $O_1, \ldots, O_\ell$. Therefore, by Corollary 3.1.11, it holds that:

Proposition 3.1.12. *To find an efficient estimator for a unitarily invariant statistic $f : S \to \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X = a_1\phi_1 + \cdots + a_\ell\phi_\ell$, where $a_1, \ldots, a_\ell \in \mathbb{C}$ and $\phi_1, \ldots, \phi_\ell$ are indicator functions on orbits $O_1, \ldots, O_\ell$, respectively, determined by the action of Γ on $\mathfrak{S}_n$.*

3.1.1. Case: $\varrho = \rho^{\otimes n}$.

Let S denote the set of states of the form $\varrho = \rho^{\otimes n}$, where ρ is a quantum state on $\mathbb{C}^d$. Let $\alpha \in \mathbb{R}^d$ denote the spectrum of ρ (taken in some arbitrary order).

When ϱ is a state of the form $\rho^{\otimes n}$, the expectation $\mathbf{E}_\varrho[\pi]$ of $\pi \in \mathfrak{S}_n$ has a particularly simple formula:

Proposition 3.1.13. *For all $\pi \in \mathfrak{S}_n$ with $\mathrm{cyc}(\pi) = \kappa$, $\mathbf{E}_\varrho[\pi] = p_\kappa(\alpha)$.*

Proof. Let ℓ denote the number of disjoint cycles in the decomposition of π. By Proposition 3.1.7,

$$\mathop{\mathbf{E}}_{\varrho}[\pi] = \mathrm{tr}(\rho^{\kappa_1}) \cdots \mathrm{tr}(\rho^{\kappa_\ell}) = p_{\kappa_1}(\alpha) \cdots p_{\kappa_\ell}(\alpha) = p_\kappa(\alpha). \qquad \blacksquare$$

Thus, $\mathbf{E}_\varrho[\pi]$ depends only on the cycle type of π. Since the cycle types of $\pi_1\pi_2$ and $\pi_2\pi_1$ are equal for all $\pi_1, \pi_2 \in \mathfrak{S}_n$, the following result holds:

Proposition 3.1.14. *For all $X, Y \in \mathbb{C}\mathfrak{S}_n$, $\mathbf{E}_\varrho[XY] = \mathbf{E}_\varrho[YX]$.*

Proof. For all $\pi_1, \pi_2 \in \mathfrak{S}_n$, $\mathrm{cyc}(\pi_1\pi_2) = \mathrm{cyc}(\pi_2\pi_1)$. Hence, by Proposition 3.1.13, $\mathbf{E}_\varrho[\pi_1\pi_2] = \mathbf{E}_\varrho[\pi_2\pi_1] = p_\kappa(\alpha)$, where $\kappa = \mathrm{cyc}(\pi_1\pi_2)$. It follows by linearity that $\mathbf{E}_\varrho[XY] = \mathbf{E}_\varrho[YX]$ for all $X, Y \in \mathbb{C}\mathfrak{S}_n$. $\blacksquare$

Thus, we obtain the following strengthening of Corollary 3.1.5:

Proposition 3.1.15. *To find an efficient estimator for a unitarily invariant statistic $f : S \to \mathbb{R}$, it suffices to consider estimators of the form $\mathcal{P}(X)$ with $X \in Z(\mathbb{C}\mathfrak{S}_n)$.*

Proof. By Proposition 3.1.14, $\Gamma = \mathfrak{S}_n$. The statement follows immediately from Corollary 3.1.11. $\blacksquare$

The expectation $\mathbf{E}_\varrho[X]$ of an estimator $X \in Z(\mathbb{C}\mathfrak{S}_n)$ can be expressed as a linear combination of $p_\kappa(\alpha)$ with $\kappa \vdash n$ where, recall, α is the spectrum of ρ. By Proposition 2.6.4, the elements $\mathcal{O}_\kappa \in \mathbb{C}\mathfrak{S}_n$ with $\kappa \vdash n$ form a linear basis for the real vector space of self-adjoint elements of

$Z(\mathbb{C}\mathfrak{S}_n)$. Hence, an estimator $X \in Z(\mathbb{C}\mathfrak{S}_n)$ can be expressed uniquely as a linear combination of the form

$$X = \sum_{\kappa \vdash n} a_\kappa \mathcal{O}_\kappa,$$

where $a_\kappa \in \mathbb{R}$ for all $\kappa \vdash n$. Thus, by Proposition 3.1.13,

$$\mathop{\mathbf{E}}_{\varrho}[X] = \sum_{\kappa \vdash n} a_\kappa \mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_\kappa] = \sum_{\kappa \vdash n} a_\kappa p_\kappa(\alpha).$$

Moreover, an estimator $X \in Z(\mathbb{C}\mathfrak{S}_n)$ is unique, as the following result shows.

Proposition 3.1.16. *If $X_1, X_2 \in Z(\mathbb{C}\mathfrak{S}_n)$ are estimators for $f : S \to \mathbb{R}$, then $X_1 = X_2$.*

Proof. Suppose $X_1 = \sum a_\kappa \mathcal{O}_\kappa$ and $X_2 = \sum b_\kappa \mathcal{O}_\kappa$. Since X_1 and X_2 are estimators for $f : S \to \mathbb{R}$, it follows that

$$\sum_{\kappa \vdash n} a_\kappa p_\kappa(\alpha) = \mathop{\mathbf{E}}_{\varrho}[X_1] = \mathop{\mathbf{E}}_{\varrho}[X_2] = \sum_{\kappa \vdash n} b_\kappa p_\kappa(\alpha).$$

Thus, if $h(\alpha)$ is defined by

$$h(\alpha) = \sum_{\kappa \vdash n} (a_\kappa - b_\kappa) p_\kappa(\alpha),$$

then $h(\alpha) = 0$ for all $\alpha \in \mathbb{R}^d_+$ with $\|\alpha\|_1 = 1$. Note that h is a homogeneous polynomial of degree n in α. Hence, if $x \in \mathbb{R}^d_+$ with $\|x\|_1 > 0$, then

$$h(x) = h\left(\|x\|_1 \cdot \frac{x}{\|x\|_1}\right) = \|x\|_1^n \cdot h\left(\frac{x}{\|x\|_1}\right) = 0.$$

Thus, $h(x) = 0$ for all $x \in \mathbb{R}^d_+$. Since h is a polynomial, it follows that $h \equiv 0$. Therefore, $a_\kappa = b_\kappa$ for all $\kappa \vdash n$, so $X_1 = X_2$. ∎

Therefore, all observables in the center of $\mathbb{C}\mathfrak{S}_n$ are efficient estimators:

Corollary 3.1.17. *If $X \in Z(\mathbb{C}\mathfrak{S}_n)$ is an estimator for $f : S \to \mathbb{R}$, then X is efficient.*

Proof. The result follows from Proposition 3.1.15 and Proposition 3.1.16. ∎

Example 3.1.18. By Corollary 3.1.17, $\mathcal{O}_\kappa$ is an efficient estimator for $f(\varrho) = p_\kappa(\alpha)$. In particular, suppose $\kappa = (k, 1, 1, \dots)$, which we will denote simply as (k) (recalling Remark 3.1.8). Then $\mathcal{O}_{(k)}$ is an efficient estimator of $f(\varrho) = \operatorname{tr}(\rho^k)$.

3.1.2. Case: $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$. Let S denote the set of states of the form $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$, where ρ and σ are quantum states on $\mathbb{C}^d$. Let $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}^d$ denote the spectra of ρ and σ, respectively. The group Γ of permutation invariants of S can be described as follows:

Proposition 3.1.19. *$\Gamma \cong \mathfrak{S}_m \times \mathfrak{S}_n$, where $(\pi_1, \pi_2) \in \Gamma$ embeds in $\mathfrak{S}_{m+n}$ in the natural way, viz. by applying π_1 to $\{1, \dots, m\}$ and applying π_2 to $\{m, \dots, m+n\}$.*

Proof. Let Γ be as in the statement of the proposition and let $\tau \in \Gamma$. The conjugation $\pi \mapsto \tau^{-1}\pi\tau$ applies τ to each index in the cycle decomposition of π. Hence, if τ acts as in the statement of the proposition, then, by Proposition 3.1.7, $\mathbf{E}_\varrho[\pi] = \mathbf{E}_\varrho[\tau^{-1}\pi\tau]$.

Conversely, let $\tau \in \mathfrak{S}_{m+n}$ and suppose there exists an index $i \in \{1, \ldots, m\}$ such that $\tau(i) \in \{m+1, \ldots, m+n\}$. Thus, if $\pi = (1\ i)$, then $\mathbf{E}_{\varrho}[\pi] = \operatorname{tr}(\rho^2)$ and

$$\mathop{\mathbf{E}}_{\varrho}[\tau^{-1}\pi\tau] = \begin{cases} \operatorname{tr}(\rho\sigma), & \tau(1) \in \{1, \ldots, m\}, \\ \operatorname{tr}(\sigma^2), & \tau(1) \in \{m+1, \ldots, m+n\}. \end{cases}$$

Since ρ and σ are arbitrary quantum states, it follows that $\tau \notin \Gamma$. ■

To find an efficient estimator with respect to S, it is sufficient, by Proposition 3.1.12, to consider functions $X \in \mathbb{C}\mathfrak{S}_{m+n}$ which are constant on the orbits defined by the action of Γ on $\mathfrak{S}_{m+n}$.

Notation 3.1.20. Since Γ acts on $\mathfrak{S}_{m+n}$ by conjugation, the orbits of Γ refine the conjugacy classes of $\mathfrak{S}_{m+n}$. An orbit of Γ is uniquely determined by a signature consisting of a cycle type and a map that associates each index in the cycle type with either ρ or σ. For instance, the signature $(\rho\,\sigma)$ identifies the orbit of Γ which consists of all transpositions that exchange an index in $\{1, \ldots, m\}$ with an index in $\{m+1, \ldots, m+n\}$. Note that $(\rho\,\sigma) = (\sigma\,\rho)$. Similarly, $(\rho\,\rho\,\sigma)$ denotes the set of 3-cycles with two indices in $\{1, \ldots, m\}$ and one index in $\{m+1, \ldots, m+n\}$.

If $\mathfrak{s}$ is the signature of an orbit of Γ, let $\mathcal{O}_{\mathfrak{s}} \in \mathbb{C}\mathfrak{S}_{m+n}$ denote the average of all elements in the orbit. For example, $\mathcal{O}_{(\rho\,\sigma)}$ denotes the average of all transpositions in the $(\rho\,\sigma)$ orbit described above.

Example 3.1.21. By Proposition 3.1.7, $\mathcal{O}_{(\rho\,\sigma)}$ is an estimator for $f(\varrho) = \operatorname{tr}(\rho\sigma)$.

Moreover, $\mathcal{O}_{(\rho\,\sigma)}$ satisfies the following uniqueness property:

Proposition 3.1.22. *If $X \in \mathbb{C}\mathfrak{S}_{m+n}$ is an estimator for the statistic $f : S \to \mathbb{R}$ defined by $f(\varrho) = \operatorname{tr}(\rho\sigma)$ and X is of the form presented in Proposition 3.1.12, then $X = \mathcal{O}_{(\rho\,\sigma)}$.*

Proof. In the case when $\rho = \sigma$, X becomes an estimator for $\operatorname{tr}(\rho^2)$. Then, by Proposition 3.1.13, $\mathbf{E}_{\varrho}[X]$ can be expressed as follows:

$$\mathop{\mathbf{E}}_{\varrho}[X] = \sum_{\kappa \vdash m+n} a_\kappa p_\kappa(\alpha),$$

where α is the spectrum of ρ and $a_\kappa \in \mathbb{R}$ for all $\kappa \vdash m+n$. Since $\mathbf{E}_{\varrho}[X] - p_2(\alpha) = 0$ for all $\alpha \in \mathbb{R}^d$ with $\|\alpha\|_1 = 1$, it follows, as in the proof of Proposition 3.1.16, that $a_\kappa = 0$ for all $\kappa \vdash m+n$ with $\kappa \neq (2)$ and $a_{(2)} = 1$. Thus, in general, $X = a\mathcal{O}_{(\rho\,\rho)} + b\mathcal{O}_{(\sigma\,\sigma)} + c\mathcal{O}_{(\rho\,\sigma)}$ with $a + b + c = 1$. Since $\mathbf{E}_{\varrho}[X] = \operatorname{tr}(\rho\sigma)$, it follows that $c = 1$ and $a = b = 0$. ■

A similar argument proves the following:

Proposition 3.1.23. *If $X \in \mathbb{C}\mathfrak{S}_{m+n}$ is an estimator for the statistic $f : S \to \mathbb{R}$ defined by $f(\varrho) = d_{\mathrm{HS}}^2(\rho, \sigma)$ and X is of the form presented in Proposition 3.1.12, then $X = \mathcal{O}_{(\rho\,\rho)} + \mathcal{O}_{(\sigma\,\sigma)} - 2\mathcal{O}_{(\rho\,\sigma)}$.*

Thus, the estimators obtained for $\operatorname{tr}(\rho\sigma)$ and $d_{\mathrm{HS}}^2(\rho, \sigma)$ are efficient:

Corollary 3.1.24. *$\mathcal{O}_{(\rho\,\sigma)}$ is an efficient estimator for $f(\varrho) = \operatorname{tr}(\rho\sigma)$.*

Corollary 3.1.25. *$\mathcal{O}_{(\rho\,\rho)} + \mathcal{O}_{(\sigma\,\sigma)} - 2\mathcal{O}_{(\rho\,\sigma)}$ is an efficient estimator for $f(\varrho) = \mathrm{D}_{\mathrm{HS}}^2(\rho, \sigma)$.*

3.2. Hilbert–Schmidt distance and related estimation

In Corollary 3.1.24 and Corollary 3.1.25, we obtain observables which are efficient quantum estimators for the quantities $\operatorname{tr}(\rho\sigma)$ and $d_{\mathrm{HS}}(\rho, \sigma)$ determined by two quantum states ρ and σ. In this section, we establish upper bounds on the variance of these estimators.

3.2.1. Purity, and testing identity to the maximally mixed state. Let ρ be a quantum state on $\mathbb{C}^d$, let $\varrho = \rho^{\otimes n}$ denote n copies of ρ, and define $f(\varrho) = \operatorname{tr}(\rho^2)$. The quantity $\operatorname{tr}(\rho^2)$ is called the ***purity*** of ρ.

Since $d_{\mathrm{HS}}^2(\rho, \sigma) = \operatorname{tr}(\rho^2) - 2\langle\rho, \sigma\rangle + \operatorname{tr}(\sigma^2)$,

$$d_{\mathrm{HS}}\left(\rho, \frac{\mathbf{1}}{d}\right) = \operatorname{tr}(\rho^2) - \frac{2}{d} + \frac{1}{d} = \operatorname{tr}(\rho^2) - \frac{1}{d}.$$

Thus, up to an additive constant depending on the dimension d, the purity of ρ is equal to the squared Hilbert–Schmidt distance between ρ and the maximally mixed state.

By Corollary 3.1.17, $\mathbf{E}_\varrho[\mathcal{O}_{(2)}] = \operatorname{tr}(\rho^2)$ is an efficient estimator for purity. The following result gives an explicit formula for the variance of $\mathcal{O}_{(2)}$:

Lemma 3.2.1.

$$\underset{\varrho}{\mathbf{Var}}[\mathcal{O}_{(2)}] = \frac{1}{\binom{n}{2}}(1 - p_2(\alpha)^2) + \frac{2(n-2)}{\binom{n}{2}}(p_3(\alpha) - p_2(\alpha)^2).$$

Proof. By definition, $\mathcal{O}_{(2)}$ is the normalized sum of all transpositions in $\mathfrak{S}_n$. The product of two transpositions is either the identity, a 3-cycle, or a product of two disjoint transpositions. Thus, to compute $\mathcal{O}_{(2)}^2$, it suffices to consider the following counting problem: given two uniformly random transpositions τ_1 and τ_2 from $\mathfrak{S}_n$, what is the probability that $\tau_1\tau_2$ is the identity, a 3-cycle, or a product of two disjoint transpositions.

There are, in total, $\binom{n}{2}^2$ choices for τ_1 and τ_2. $\tau_1\tau_2 = \mathrm{id}$ if and only if $\tau_1 = \tau_2$, so $\tau_1\tau_2$ is the identity with probability:

$$\frac{\binom{n}{2}}{\binom{n}{2}^2} = \frac{1}{\binom{n}{2}}.$$

$\tau_1\tau_2$ is a permutation of cycle type $(2, 2)$ if and only if τ_1 and τ_2 are disjoint, which occurs with probability:

$$\frac{\binom{n}{2}\binom{n-2}{2}}{\binom{n}{2}^2} = \frac{\binom{n-2}{2}}{\binom{n}{2}}.$$

Finally, $\tau_1\tau_2$ is a 3-cycle if and only if τ_1 and τ_2 share exactly one index. If $\tau_1 = (i\ j)$, then $\tau_2 = (i\ k)$ or $\tau_2 = (j\ k)$ for some $k \neq i, j$. In each case, there are $n - 2$ choices for k, so the probability that $\tau_1\tau_2$ is a 3-cycle is:

$$\frac{\binom{n}{2} \cdot 2(n-2)}{\binom{n}{2}^2} = \frac{2(n-2)}{\binom{n}{2}}.$$

Hence, since all permutations of cycle type κ for $\kappa \in \{(1), (3), (2, 2)\}$ are obtained as products of two transpositions, it holds that

$$\mathcal{O}_{(2)}^2 = \frac{1}{\binom{n}{2}}\mathbf{1} + \frac{2(n-2)}{\binom{n}{2}}\mathcal{O}_{(3)} + \frac{\binom{n-2}{2}}{\binom{n}{2}}\mathcal{O}_{(2,2)}.$$

Therefore, by Proposition 3.1.13,

$$\begin{aligned}\mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_{(2)}^2] &= \frac{1}{\binom{n}{2}} + \frac{2(n-2)}{\binom{n}{2}}p_3(\alpha) + \frac{\binom{n-2}{2}}{\binom{n}{2}}p_{(2,2)}(\alpha)\\ &= \frac{1}{\binom{n}{2}} + \frac{2(n-2)}{\binom{n}{2}}p_3(\alpha) + \left(1 - \frac{2(n-2)+1}{\binom{n}{2}}\right)p_2(\alpha)^2,\end{aligned}$$

and the lemma follows. ■

Using Lemma 3.2.1, we can prove our main result on state certification with respect to Hilbert–Schmidt distance in the special case when σ is known to be the maximally mixed state (cf. Example 2.2.2). (This result was originally proven, in a slightly more opaque way, in [45, Theorem 4.1].)

Proposition 3.2.2. *There is an algorithm that, given $n = O(1/\epsilon^2)$ copies of the state $\rho \in \mathrm{B}(\mathbb{C}^d)$, with high probability distinguishes between $d_{\mathrm{HS}}(\rho, \frac{\mathbf{1}}{d}) \leq .99\epsilon$ and $d_{\mathrm{HS}}(\rho, \frac{\mathbf{1}}{d}) > \epsilon$.*

Proof. Since $d_{\mathrm{HS}}^2(\rho, \frac{\mathbf{1}}{d}) = \mathrm{tr}(\rho^2) - \frac{1}{d}$, the observable $\mathcal{O}_{(2)} - \frac{\mathbf{1}}{d}$ is an unbiased estimator of the distance $d_{\mathrm{HS}}^2(\rho, \frac{\mathbf{1}}{d})$ between ρ and the maximally mixed state. Let $\alpha \in \mathbb{R}^d$ denote the spectrum of ρ. The variance of this estimator is given by:

$$\begin{aligned}\mathop{\mathbf{Var}}_{\varrho}\left[\mathcal{O}_{(2)} - \frac{\mathbf{1}}{d}\right] &= \mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(2)}]\\ &= \mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_{(2)}^2] - \mathbf{E}[{}_{\varrho}\mathcal{O}_{(2)}]^2\\ &= \mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_{(2)}^2] - p_2(\alpha)^2\\ &= \frac{1}{\binom{n}{2}} + \frac{2(n-2)}{\binom{n}{2}}p_3(\alpha) + \left(1 - \frac{2(n-2)+1}{\binom{n}{2}}\right)p_2(\alpha)^2 - p_2(\alpha)^2\\ &= \frac{1}{\binom{n}{2}} + \frac{2(n-2)}{\binom{n}{2}}p_3(\alpha) - \frac{2(n-2)+1}{\binom{n}{2}}p_2(\alpha)^2\\ &= O\left(\frac{1}{n^2}\right)\cdot(1 - p_2(\alpha)^2) + O\left(\frac{1}{n}\right)\cdot(p_3(\alpha) - p_2(\alpha)^2).\end{aligned}$$

In order to apply Lemma 2.8.1, we would like to bound $p_3(\alpha) - p_2(\alpha)^2$ by $d_{\mathrm{HS}}^2(\rho, \frac{\mathbf{1}}{d})$. To this end, we rewrite $p_3(\alpha) - p_2(\alpha)^2$ in terms of $p_2(\Delta)$ and $p_3(\Delta)$, where $\Delta \in \mathbb{R}^d$ is the difference vector defined by $\Delta_i = \alpha_i - 1/d$ for all $i = 1, \ldots, d$. Note that $d_{\mathrm{HS}}^2(\rho, \frac{\mathbf{1}}{d}) = p_2(\Delta)$. Expanding and simplifying the terms on the LHS below, we obtain:

$$\begin{aligned}p_2(\Delta) &= p_2(\alpha) - \frac{2}{d}p_1(\alpha) + \frac{1}{d} = p_2(\alpha) - \frac{1}{d};\\ p_2(\Delta)^2 &= p_2(\alpha)^2 - \frac{2}{d}p_2(\alpha) + \frac{1}{d^2};\\ p_3(\Delta) &= p_3(\alpha) - \frac{3}{d}p_2(\alpha) + \frac{3}{d^2}p_1(\alpha) - \frac{1}{d^2} = p_3(\alpha) - \frac{3}{d}p_2(\alpha) + \frac{2}{d^2}.\end{aligned}$$

Hence,

$$p_3(\Delta) - p_2(\Delta)^2 = p_3(\alpha) - p_2(\alpha)^2 - \frac{1}{d} p_2(\alpha) + \frac{1}{d^2},$$

so $p_3(\alpha) - p_2(\alpha)^2 = p_3(\Delta) - p_2(\Delta)^2 + \frac{1}{d} p_2(\Delta)$. Therefore,

$$p_3(\alpha) - p_2(\alpha)^2 = p_3(\Delta) - p_2(\Delta)^2 + \frac{1}{d} p_2(\Delta) \le 2p_2(\Delta) = 2d_{\mathrm{HS}}^2\left(\rho, \frac{\mathbf{1}}{d}\right).$$

Finally, we conclude that

$$\operatorname*{\mathbf{Var}}_{\varrho}\left[\mathcal{O}_{(2)} - \frac{\mathbf{1}}{d}\right] = \operatorname*{\mathbf{Var}}_{\varrho}\left[\mathcal{O}_{(2)}\right] \le O\left(\frac{1}{n^2} + \frac{d_{\mathrm{HS}}^2(\rho, \frac{\mathbf{1}}{d})}{n}\right).$$

The result now follows immediately from Lemma 2.8.1. ■

3.2.2. Linear fidelity. Let ρ and σ be quantum states on $\mathbb{C}^d$, let $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$, and define $f(\varrho) = \operatorname{tr}(\rho\sigma)$. The quantity $\operatorname{tr}(\rho\sigma)$ is sometimes called the *overlap* or ***linear fidelity*** between ρ and σ. By Corollary 3.1.24, $\mathcal{O}_{(\rho\,\sigma)}$ is an efficient estimator for the statistic f. The following result gives an explicit formula for the variance of $\mathcal{O}_{(\rho\,\sigma)}$.

Proposition 3.2.3.

$$\operatorname*{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\rho\,\sigma)}] = \frac{1}{mn} + \frac{1-m-n}{mn}\operatorname{tr}(\rho\sigma)^2 + \frac{1}{n}\left(1 - \frac{1}{m}\right)\operatorname{tr}(\rho^2\sigma) + \frac{1}{m}\left(1 - \frac{1}{n}\right)\operatorname{tr}(\rho\sigma^2).$$

Proof. The proof uses a counting argument very similar to the proof of Lemma 3.2.1. Given two transpositions τ_1 and τ_2 of type $(\rho\ \sigma)$, i.e. τ_1 and τ_2 swap one of the m copies of ρ with one of the n copies of σ, their product $\tau_1\tau_2$ is either equal to the identity, a 3-cycle of type $(\rho\ \rho\ \sigma)$ or $(\rho\ \sigma\ \sigma)$, or of type $(\rho\ \sigma)(\rho\ \sigma)$.

The product of two uniformly random transpositions of type $(\rho\ \sigma)$ is:

- the identity with probability $\frac{1}{mn}$;
- of type $(\rho\,\sigma)(\rho\,\sigma)$ with probability $\left(1 - \frac{1}{m}\right)\left(1 - \frac{1}{n}\right)$;
- of type $(\rho\,\rho\,\sigma)$ with probability $\frac{1}{n}(1 - \frac{1}{m})$;
- and of type $(\rho\,\sigma\,\sigma)$ with probability $\frac{1}{m}(1 - \frac{1}{n})$.

Therefore,

$$\mathcal{O}_{(\rho\,\sigma)}^2 = \frac{1}{mn}\mathbf{1} + \left(1 - \frac{1}{m}\right)\left(1 - \frac{1}{n}\right)\mathcal{O}_{(\rho\,\sigma)(\rho\,\sigma)} + \frac{1}{n}\left(1 - \frac{1}{m}\right)\mathcal{O}_{(\rho\,\rho\,\sigma)} + \frac{1}{m}\left(1 - \frac{1}{n}\right)\mathcal{O}_{(\rho\,\sigma\,\sigma)}. \quad ■$$

3.2.3. Squared Hilbert–Schmidt distance. Let ρ and σ be quantum states on $\mathbb{C}^d$, let $\varrho = \rho^{\otimes m} \otimes \sigma^{\otimes n}$, and define $f(\varrho) = d_{\mathrm{HS}}^2(\rho, \sigma) = \operatorname{tr}(\rho^2) + \operatorname{tr}(\sigma^2) - 2\operatorname{tr}(\rho\sigma)$. By Corollary 3.1.25, $\mathcal{O}_{(\rho\,\rho)} + \mathcal{O}_{(\sigma\,\sigma)} - 2\mathcal{O}_{(\rho\,\sigma)}$ is an efficient estimator for the statistic f. In this section, we give an explicit upper bound on the variance of this estimator.

Lemma 3.2.4. $\operatorname*{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\sigma\,\sigma)}] = 0.$

Proof. Note that $\mathcal{O}_{(\rho\,\rho)} = \mathcal{O}_{(2)} \otimes \mathbf{1}$, where $\mathcal{O}_{(2)}$ is defined on the first m components of the tensor product. Similarly, $\mathcal{O}_{(\sigma\,\sigma)} = \mathbf{1} \otimes \mathcal{O}_{(2)}$, where $\mathcal{O}_{(2)}$ is defined on the last n components of the tensor product. Hence, by Equation (2),

$$\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\sigma\,\sigma)}] = \mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(2)} \otimes \mathbf{1}, \mathbf{1} \otimes \mathcal{O}_{(2)}] = 0. \qquad \blacksquare$$

Lemma 3.2.5. $\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\rho\,\sigma)}] = \dfrac{2}{m}\big(\mathrm{tr}(\rho^2\sigma) - \mathrm{tr}(\rho^2)\,\mathrm{tr}(\rho\sigma)\big)$.

Proof. A permutation of type $(\rho\,\rho)(\rho\,\sigma)$ or $(\rho\,\rho\,\sigma)$ is uniquely determined by a product of two transpositions of types $(\rho\,\rho)$ and $(\rho\,\sigma)$. Hence,

$$\mathcal{O}_{(\rho\,\rho)}\mathcal{O}_{(\rho\,\sigma)} = \frac{2}{m}\mathcal{O}_{(\rho\,\rho\,\sigma)} + \left(1 - \frac{2}{m}\right)\mathcal{O}_{(\rho\,\rho)(\rho\,\sigma)}.$$

Therefore,

$$\begin{aligned}\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\rho\,\sigma)}] &= \mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}\mathcal{O}_{(\rho\,\sigma)}] - \mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}]\,\mathop{\mathbf{E}}_{\varrho}[\mathcal{O}_{(\rho\,\sigma)}] \\ &= \frac{2}{m}\,\mathrm{tr}(\rho^2\sigma) + \left(1 - \frac{2}{m}\right)\mathrm{tr}(\rho^2)\,\mathrm{tr}(\rho\sigma) - \mathrm{tr}(\rho^2)\,\mathrm{tr}(\rho\sigma) \\ &= \frac{2}{m}\,\mathrm{tr}(\rho^2\sigma) - \frac{2}{m}\,\mathrm{tr}(\rho^2)\,\mathrm{tr}(\rho\sigma). \qquad \blacksquare\end{aligned}$$

Proposition 3.2.6. *If $m = n$, then*

$$\mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)} + \mathcal{O}_{(\sigma\,\sigma)} - 2\mathcal{O}_{(\rho\,\sigma)}] = O\!\left(\frac{1}{n^2} + \frac{d_{\mathrm{HS}}^2(\rho,\sigma)}{n}\right).$$

Proof. Let $\mathcal{V} = \mathbf{Var}_{\varrho}[\mathcal{O}_{(\rho\,\rho)} + \mathcal{O}_{(\sigma\,\sigma)} - 2\mathcal{O}_{(\rho\,\sigma)}]$ denote the variance of the estimator from Corollary 3.1.25. Since $\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\sigma\,\sigma)} \in \mathbb{C}\Gamma$, $\mathcal{O}_{(\rho\,\rho)}$ and $\mathcal{O}_{(\sigma\,\sigma)}$ commute with each other and with $\mathcal{O}_{(\rho\,\sigma)}$. Hence, by Lemma 3.2.4,

$$\mathcal{V} = \mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}] + \mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\sigma\,\sigma)}] + 4\mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\rho\,\sigma)}] - 4\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\rho\,\sigma)}] - 4\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\sigma\,\sigma)}, \mathcal{O}_{(\rho\,\sigma)}].$$

Using prior results, we have

$$\mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}] + \mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\sigma\,\sigma)}] \le O\!\left(\frac{1}{n^2}\right) + \frac{4}{n}\big(\mathrm{tr}(\rho^3) + \mathrm{tr}(\sigma^3) - \mathrm{tr}(\rho^2)^2 - \mathrm{tr}(\sigma^2)^2\big),$$

$$\begin{aligned}4\mathop{\mathbf{Var}}_{\varrho}[\mathcal{O}_{(\rho\,\sigma)}] &= \frac{4}{n^2} + \frac{4-8n}{n^2}\,\mathrm{tr}(\rho\sigma)^2 + \frac{4n-4}{n^2}\,\mathrm{tr}(\rho^2\sigma) + \frac{4n-4}{n^2}\,\mathrm{tr}(\rho\sigma^2) \\ &\le O\!\left(\frac{1}{n^2}\right) + \frac{4}{n}\big(\mathrm{tr}(\rho^2\sigma) + \mathrm{tr}(\rho\sigma^2) - 2\,\mathrm{tr}(\rho\sigma)^2\big),\end{aligned}$$

and

$$-4\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\rho\,\rho)}, \mathcal{O}_{(\rho\,\sigma)}] - 4\mathop{\mathbf{Cov}}_{\varrho}[\mathcal{O}_{(\sigma\,\sigma)}, \mathcal{O}_{(\rho\,\sigma)}] = -\frac{8}{n}\big(\mathrm{tr}(\rho^2\sigma) + \mathrm{tr}(\rho\sigma^2) - \big(\mathrm{tr}(\rho^2) + \mathrm{tr}(\sigma^2)\big)\,\mathrm{tr}(\rho\sigma)\big).$$

Therefore,

$$\mathcal{V} \le O\!\left(\frac{1}{n^2}\right) + \frac{4}{n}\big(\mathrm{tr}(\rho^3) + \mathrm{tr}(\sigma^3) - \mathrm{tr}(\rho^2)^2 - \mathrm{tr}(\sigma^2)^2 + \mathrm{tr}(\rho^2\sigma) + \mathrm{tr}(\rho\sigma^2) - 2\,\mathrm{tr}(\rho\sigma)^2\big)$$

$$
\begin{aligned}
&\quad -\frac{4}{n}\big(2\operatorname{tr}(\rho^2\sigma)+2\operatorname{tr}(\rho\sigma^2)-2\big(\operatorname{tr}(\rho^2)+\operatorname{tr}(\sigma^2)\big)\operatorname{tr}(\rho\sigma)\big)\\
&= O\left(\frac{1}{n^2}\right)+\frac{4}{n}\big(\operatorname{tr}(\rho^3)+\operatorname{tr}(\sigma^3)-\operatorname{tr}(\rho^2)^2-\operatorname{tr}(\sigma^2)^2-\operatorname{tr}(\rho^2\sigma)-\operatorname{tr}(\rho\sigma^2)-2\operatorname{tr}(\rho\sigma)^2\big)\\
&\qquad +\frac{4}{n}\big(2\big(\operatorname{tr}(\rho^2)+\operatorname{tr}(\sigma^2)\big)\operatorname{tr}(\rho\sigma)\big)\\
&= O\left(\frac{1}{n^2}\right)+\frac{4}{n}\big(\operatorname{tr}((\rho+\sigma)(\rho-\sigma)^2)-(\operatorname{tr}(\rho^2)-\operatorname{tr}(\rho\sigma))^2-(\operatorname{tr}(\sigma^2)-\operatorname{tr}(\rho\sigma))^2\big)\\
&\leq O\left(\frac{1}{n^2}\right)+\frac{4}{n}\operatorname{tr}((\rho+\sigma)(\rho-\sigma)^2)\\
&\leq O\left(\frac{1}{n^2}\right)+\frac{4}{n}\|\rho+\sigma\|_\infty\cdot\operatorname{tr}\big((\rho-\sigma)^2\big)\\
&\leq O\left(\frac{1}{n^2}\right)+O\left(\frac{1}{n}\right)\cdot d_{\mathrm{HS}}^2(\rho,\sigma). \quad\blacksquare
\end{aligned}
$$

3.2.4. Consequences for testing. State certification with respect to Hilbert–Schmidt distance, Theorem 3.0.1, follows immediately from Lemma 2.8.1 and the variance bound in Proposition 3.2.6. State certification with respect to trace distance, Corollary 3.0.2, follows from Theorem 3.0.1 and Equation (6).

In the remainder of this section, we give the proof of Theorem 3.0.3, restated below, which improves Theorem 3.0.1 in the case that one or both of the states ρ and σ is of low rank:

Theorem 3.0.3. *If either ρ or σ is close to having rank k, in the sense that the sum of its largest k eigenvalues is at least $1-\delta$, then there is an algorithm that, given $n = O(k/\epsilon^2)$ copies of each $\rho, \sigma \in \mathrm{B}(\mathbb{C}^d)$, with high probability distinguishes between $d_{\mathrm{HS}}(\rho,\sigma) \leq 0.58\epsilon/\sqrt{k}$ or $d_{\mathrm{tr}}(\rho,\sigma) > \epsilon+\delta$.*

Proof. The algorithm applies the Hilbert–Schmidt tester from Theorem 3.0.1 with proximity parameter $\eta = c\epsilon/\sqrt{k}$ where $c = 2-\sqrt{2}$. The tester with high probability distinguishes between $d_{\mathrm{HS}}(\rho,\sigma) \leq 0.99\eta \leq 0.58\epsilon/\sqrt{k}$ and $d_{\mathrm{HS}}(\rho,\sigma) > \eta$. To complete the proof, it suffices to show that $d_{\mathrm{HS}}(\rho,\sigma) \leq \eta$ implies $d_{\mathrm{tr}}(\rho,\sigma) \leq \delta+\epsilon$.

Since both $d_{\mathrm{HS}}(\rho,\sigma)$ and $d_{\mathrm{tr}}(\rho,\sigma)$ are symmetric and unitarily invariant, we may assume, without loss of generality, that the state σ is close to having rank k and is diagonal in the standard basis, viz. $\sigma = \operatorname{diag}(\beta_1,\ldots,\beta_d)$ with $\beta_1 \geq \cdots \geq \beta_d$ and $\beta_1+\cdots+\beta_k \geq 1-\delta$.

Let ρ_A denote the $d\times d$ matrix obtained from ρ by zeroing out all entries (i,j) with $i > k$ or $j > k$. Let ρ_B denote the $d \times d$ matrix obtained from ρ by zeroing out all entries (i,j) with $i < d-k$ or $j < d-k$. Let $\rho_{\mathrm{off}} = \rho - \rho_A - \rho_B$. Let σ_A, σ_B, and σ_{off} be defined similarly. Note that $\sigma_{\mathrm{off}} = 0$ since σ is diagonal.

Thus, by the triangle inequality,

$$\|\rho-\sigma\|_1 = \|\rho_A-\sigma_A\|_1+\|\rho_B-\sigma_B\|_1+\|\rho_{\mathrm{off}}-\sigma_{\mathrm{off}}\|_1.$$

The rank of $\rho_A-\sigma_A$ is at most k. Since ρ_{off} is the sum of two matrices of rank at most k, the rank of $\rho_{\mathrm{off}}-\sigma_{\mathrm{off}} = \rho_{\mathrm{off}}$ is at most $2k$. Hence, by the Cauchy–Schwarz inequality and the assumption that $d_{\mathrm{HS}}(\rho, \sigma) \leq c\epsilon/\sqrt{k}$,

$$\begin{aligned}\|\rho_A - \sigma_A\|_1 + \|\rho_{\mathrm{off}} - \sigma_{\mathrm{off}}\|_1 &\leq \sqrt{k}\|\rho_A - \sigma_A\|_2 + \sqrt{2k}\|\rho_{\mathrm{off}} - \sigma_{\mathrm{off}}\|_2 \\ &\leq (\sqrt{k} + \sqrt{2k})\|\rho - \sigma\|_2 \leq (1+\sqrt{2})\sqrt{k}c\frac{\epsilon}{\sqrt{k}} = (1+\sqrt{2})c\epsilon,\end{aligned}$$

where the second inequality holds because ρ_A, ρ_{off}, σ_A, and σ_{off} are zero-extended submatrices of ρ and σ, respectively.

To bound $\|\rho_B - \sigma_B\|_1$, we use subadditivity of $\|\_\|_1$:

$$\|\rho_B - \sigma_B\|_1 \leq \|\rho_B\|_1 + \|\sigma_B\|_1 = \mathrm{tr}(\rho_B) + \mathrm{tr}(\sigma_B) = 1 - \mathrm{tr}(\rho_A) + 1 - \mathrm{tr}(\sigma_A).$$

The first equality follows from the fact $\rho_B, \sigma_B \geq 0$ and that $\|X\|_1 = \mathrm{tr}(X)$ for positive operators X. The second equality holds because $1 = \mathrm{tr}(\rho) = \mathrm{tr}(\rho_A) + \mathrm{tr}(\rho_B)$ and similarly for σ. Hence,

$$\begin{aligned}\|\rho_B - \sigma_B\|_1 &\leq 2 - \mathrm{tr}(\rho_A) - \mathrm{tr}(\sigma_A) \\ &= 2 - \mathrm{tr}(\rho_A - \sigma_A + \sigma_A) - \mathrm{tr}(\sigma_A) \\ &= 2 - \mathrm{tr}(\rho_A - \sigma_A) - 2\,\mathrm{tr}(\sigma_A) \\ &= 2 \cdot (1 - \mathrm{tr}(\sigma_A)) + \mathrm{tr}(\rho_A - \sigma_A) \\ &= 2\,\mathrm{tr}(\sigma_B) + \mathrm{tr}(\rho_A - \sigma_A) \\ &\leq 2\delta + \|\rho_A - \sigma_A\|_1 \\ &\leq 2\delta + \sqrt{k}\|\rho_A - \sigma_A\|_2 \\ &\leq 2\delta + \sqrt{k}\|\rho - \sigma\|_2 \\ &\leq 2\delta + \sqrt{k}c\frac{\epsilon}{\sqrt{k}} \\ &= 2\delta + c\epsilon.\end{aligned}$$

Putting everything together, we obtain:

$$\|\rho - \sigma\|_1 \leq (1+\sqrt{2})c\epsilon + 2\delta + c\epsilon = 2\delta + (2+\sqrt{2})c\epsilon = 2\delta + 2\epsilon.$$

Therefore, $d_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1 \leq \delta + \epsilon$, as needed. ■

3.3. Quantum chi-squared estimation

In this section, we define a new unbiased estimator $\mathcal{O}_{\chi^2}$ for the Bures χ^2-divergence $d_{\chi^2}(\rho, \sigma)$ between an unknown state ρ and a known state σ. With the aim of applying Lemma 2.8.1 to obtain a state certification test, we analyze the variance of $\mathcal{O}_{\chi^2}$ and prove that a suitable upper bound holds.

3.3.1. A χ^2 observable. In this section, σ is assumed to be a known quantum state on $\mathbb{C}^d$ of full rank. Recall that the Bures χ^2-divergence $d_{\chi^2}(\rho, \sigma)$ between two quantum states ρ and σ is defined by:

$$d_{\chi^2}(\rho, \sigma) = \mathrm{tr}((\rho - \sigma) \cdot (\Omega_\sigma(\rho - \sigma))).$$

Note that Ω_σ is a quantum operation, so $\Omega_\sigma(\rho - \sigma)$ is an operator in $\mathrm{B}(\mathbb{C}^d)$.

To simplify calculations, we introduce the following notation:

Definition 3.3.1. For operators $X, Y \in \mathrm{B}(\mathbb{C}^d)$, let $\omega_\sigma^{(2)}(X, Y)$ be the bilinear form defined by

$$\omega_\sigma^{(2)}(X, Y) = \mathrm{tr}(\, X \cdot (\Omega_\sigma(Y))\,).$$

$\omega_\sigma^{(2)}$ has the following "contraction" property:

Proposition 3.3.2. *For all $X \in \mathrm{B}(\mathbb{C}^d)$, it holds that $\omega_\sigma^{(2)}(X, \sigma) = \mathrm{tr}(X) = \omega_\sigma^{(2)}(\sigma, X)$.*

Proof. Recall that Ω_σ is defined as the inverse of $\mathcal{R}_\sigma(X) = \frac{1}{2} \cdot (\sigma X + X \sigma)$. Since $\mathcal{R}_\sigma(\mathbf{1}) = \sigma$, it follows that $\Omega_\sigma(\sigma) = \mathbf{1}$. Furthermore, $\mathcal{R}_\sigma(\Omega_\sigma(X)) = X$.

Thus, $\omega_\sigma^{(2)}(X, \sigma) = \mathrm{tr}(X \cdot \Omega_\sigma(\sigma)) = \mathrm{tr}(X \cdot \mathbf{1}) = \mathrm{tr}(X)$ and

$$\begin{aligned} \omega_\sigma^{(2)}(\sigma, X) &= \mathrm{tr}(\sigma \cdot \Omega_\sigma(X)) = \frac{1}{2}\,\mathrm{tr}(\sigma \cdot \Omega_\sigma(X)) + \frac{1}{2}\,\mathrm{tr}(\Omega_\sigma(X) \cdot \sigma) \\ &= \mathrm{tr}(\mathcal{R}_\sigma(\Omega_\sigma(X))) = \mathrm{tr}(X). \end{aligned}$$

■

Expanding the definition of d_{χ^2} and using Proposition 3.3.2 yields a simplified formula for the divergence:

Proposition 3.3.3. *For any quantum state $\rho \in \mathrm{B}(\mathbb{C}^d)$,*

$$d_{\chi^2}(\rho, \sigma) = \omega_\sigma^{(2)}(\rho, \rho) - 1 = \mathrm{tr}(\, \rho \cdot (\Omega_\sigma(\rho))\,) - 1.$$

Furthermore, if $\sigma = \mathrm{diag}(\beta_1, \dots, \beta_d)$, then

$$d_{\chi^2}(\rho, \sigma) = \left(\sum_{i,j=1}^{d} \frac{|\rho_{ij}|^2}{\mathrm{avg}\{\beta_i, \beta_j\}} \right) - 1.$$

Proof. By bilinearity of $\omega_\sigma^{(2)}$,

$$\begin{aligned} d_{\chi^2}(\rho, \sigma) &= \mathrm{tr}(\, (\rho - \sigma) \cdot (\Omega_\sigma(\rho - \sigma))\,) = \omega_\sigma^{(2)}((\rho - \sigma), (\rho - \sigma)) \\ &= \omega_\sigma^{(2)}(\rho, \rho) - \omega_\sigma^{(2)}(\rho, \sigma) - \omega_\sigma^{(2)}(\sigma, \rho) + \omega_\sigma^{(2)}(\sigma, \sigma) \\ &= \omega_\sigma^{(2)}(\rho, \rho) - \mathrm{tr}(\rho) - \mathrm{tr}(\rho) + \mathrm{tr}(\sigma) = \omega_\sigma^{(2)}(\rho, \rho) - 1, \end{aligned}$$

where the last two equalities follow from Proposition 3.3.2 and the fact that $\mathrm{tr}(\rho) = \mathrm{tr}(\sigma) = 1$.

Recall from Equation (8) that if $\sigma = \mathrm{diag}(\beta_1, \dots, \beta_d)$, then $\Omega_\sigma(X) = X \odot 2[(\beta_i + \beta_j)^{-1}]_{i,j=1}^d$. Let $B = [(\beta_i + \beta_j)^{-1}]_{i,j=1}^d$. Thus,

$$\begin{aligned} \omega_\sigma^{(2)}(\rho, \rho) &= \mathrm{tr}(\rho \cdot (\Omega_\sigma(\rho))) = \mathrm{tr}(\rho \cdot (\rho \odot 2B)) = \sum_{i,j=1}^{d} \rho_{ij} (\rho \odot 2B)_{ji} \\ &= \sum_{i,j=1}^{d} \rho_{ij} \cdot 2(\beta_i + \beta_j)^{-1} \rho_{ji} = \sum_{i,j=1}^{d} \frac{|\rho_{ij}|^2}{\mathrm{avg}\{\beta_i, \beta_j\}}. \end{aligned}$$

■

In light of the above, it is natural to define the following observable:

DEFINITION 3.3.4. The χ^2 ***observable*** associated to a quantum state $\sigma = \operatorname{diag}(\beta_1, \ldots, \beta_d)$ is defined by

$$\mathcal{X}_\sigma = \sum_{i,j=1}^{d} \frac{|ji\rangle\langle ij|}{\operatorname{avg}\{\beta_i, \beta_j\}}.$$

Since

$$\mathop{\mathbf{E}}_{\rho\otimes\rho}[\mathcal{X}_\sigma] = \operatorname{tr}((\rho\otimes\rho)\mathcal{X}_\sigma) = \omega_\sigma^{(2)}(\rho,\rho) = d_{\chi^2}(\rho,\sigma) + 1,$$

$\mathcal{X}_\sigma - \mathbf{1}$ is an unbiased estimator of the Bures χ^2-divergence between an unknown quantum state ρ and the known state σ.

We extend $\mathcal{X}_\sigma$ to an observable on n-copies $\rho^{\otimes n}$ by applying $\mathcal{X}_\sigma$ to all ordered pairs of distinct copies of ρ and averaging the results, similarly to the purity observable from Section 3.2.1:

DEFINITION 3.3.5. Given distinct $i, j \in [n]$, let $\mathcal{X}_\sigma^{(i,j)}$ denote the operator acting on the n-fold tensor product $(\mathbb{C}^d)^{\otimes n}$ by applying $\mathcal{X}_\sigma$ to the i-th and j-th copies of $\mathbb{C}^d$ in the tensor product and acting as the identity on the remaining copies. Note that $\mathcal{X}_\sigma^{(i,j)}$ is implicitly dependent on n.

For $n \geq 2$, the ***averaged*** χ^2 ***observable*** on $(\mathbb{C}^d)^{\otimes n}$ is defined by

$$\mathcal{O}_{\chi^2} = \mathop{\operatorname{avg}}_{i\neq j}\{\mathcal{X}_\sigma^{(i,j)}\} - \mathbf{1},$$

where the average is over all distinct ordered pairs of $i, j \in [n]$.

Clearly, the averaged χ^2 observable is also an unbiased estimator for the Bures χ^2-divergence:

Proposition 3.3.6. *The expectation and variance of the averaged χ^2 observable are given by:*

$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] = d_{\chi^2}(\rho,\sigma), \qquad \mathop{\mathbf{Var}}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] = \mathop{\mathbf{Var}}_{\rho^{\otimes n}}[\mathop{\operatorname{avg}}_{i\neq j}\{\mathcal{X}_\sigma^{(i,j)}\}].$$

The following multilinear form $\omega_\sigma^{(3)}$, which involves terms of the form $\mathcal{X}_\sigma^{(i,j)} \cdot \mathcal{X}_\sigma^{(j,k)}$, will appear in the analysis of the variance of the averaged χ^2 observable.

DEFINITION 3.3.7. For operators $X, Y, Z \in \mathrm{B}(\mathbb{C}^d)$, let $\omega_\sigma^{(3)}(X,Y,Z)$ be the multilinear form defined by

$$\omega_\sigma^{(3)}(X,Y,Z) = \operatorname{tr}(\mathcal{X}_\sigma^{(1,2)} \cdot \mathcal{X}_\sigma^{(2,3)} \cdot (X\otimes Y\otimes Z)).$$

$\omega_\sigma^{(3)}$ satisfies the following "contraction" property:

Proposition 3.3.8. *For all $X, Y \in \mathrm{B}(\mathbb{C}^d)$, it holds that*

$$\omega_\sigma^{(3)}(X,Y,\sigma) = \omega_\sigma^{(2)}(X,Y) = \omega_\sigma^{(3)}(\sigma,X,Y).$$

PROOF. Since

$$\mathcal{X}_\sigma^{(1,2)} \cdot \mathcal{X}_\sigma^{(2,3)} = \sum_{i,j,k=1}^{d} \frac{|ijk\rangle\langle jki|}{\operatorname{avg}\{\beta_i,\beta_j\}\operatorname{avg}\{\beta_i,\beta_k\}},$$

it follows that

$$\omega_\sigma^{(3)}(X,Y,Z) = \sum_{i,j,k=1}^{d} \frac{Z_{ij}Y_{jk}X_{ki}}{\operatorname{avg}\{\beta_i,\beta_j\}\operatorname{avg}\{\beta_i,\beta_k\}}.$$

Thus,

$$\begin{aligned}\omega_\sigma^{(3)}(\sigma, Y, Z) &= \sum_{i,j,k=1}^{d} \frac{Z_{ij} Y_{jk} \sigma_{ki}}{\operatorname{avg}\{\beta_i, \beta_j\} \operatorname{avg}\{\beta_i, \beta_k\}} \\ &= \sum_{j,k=1}^{d} \frac{Z_{kj} Y_{jk} \sigma_{kk}}{\operatorname{avg}\{\beta_k, \beta_j\} \operatorname{avg}\{\beta_k, \beta_k\}} \\ &= \sum_{j,k=1}^{d} \frac{Z_{kj} Y_{jk} \beta_k}{\operatorname{avg}\{\beta_k, \beta_j\} \beta_k} \\ &= \sum_{j,k=1}^{d} \frac{Z_{kj} Y_{jk}}{\operatorname{avg}\{\beta_k, \beta_j\}} \\ &= \omega_\sigma^{(2)}(Y, Z).\end{aligned}$$

Similarly,

$$\begin{aligned}\omega_\sigma^{(3)}(X, Y, \sigma) &= \sum_{i,j,k=1}^{d} \frac{\sigma_{ij} Y_{jk} X_{ki}}{\operatorname{avg}\{\beta_i, \beta_j\} \operatorname{avg}\{\beta_i, \beta_k\}} \\ &= \sum_{j,k=1}^{d} \frac{\sigma_{jj} Y_{jk} X_{kj}}{\operatorname{avg}\{\beta_j, \beta_j\} \operatorname{avg}\{\beta_j, \beta_k\}} \\ &= \sum_{j,k=1}^{d} \frac{Y_{jk} X_{kj}}{\operatorname{avg}\{\beta_k, \beta_j\}} \\ &= \omega_\sigma^{(2)}(X, Y).\end{aligned}$$

■

3.3.2. Analyzing the variance of the averaged χ^2 observable. In this section, an exact formula for the variance of the averaged χ^2 observable is given in Proposition 3.3.9 and an upper bound on the variance is proved in Theorem 3.3.14.

Proposition 3.3.9. *The averaged χ^2-observable has variance*

$$\mathop{\mathbf{Var}}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] = \frac{1}{\binom{n}{2}}(\operatorname{tr}(\rho^{\otimes 2}\mathcal{X}_\sigma^2) - \omega_\sigma^{(2)}(\rho,\rho)^2) + \frac{2(n-2)}{\binom{n}{2}}(\omega_\sigma^{(3)}(\rho,\rho,\rho) - \omega_\sigma^{(2)}(\rho,\rho)^2).$$

Proof. The proof is very similar to the proof of Lemma 3.2.1. Given transpositions $\tau_1, \tau_2 \in \mathfrak{S}_n$, there are three cases to consider: the product $\tau_1\tau_2$ may be the identity, a 3-cycle, or a permutation of cycle type $(2, 2)$.

If $\tau_1\tau_2$ is the identity, then $\tau_1 = \tau_2$, so $\mathcal{X}_\sigma^{\tau_1}\mathcal{X}_\sigma^{\tau_2} = (\mathcal{X}_\sigma^{\tau_1})^2$. Since $\operatorname{tr}(\rho\,\mathbf{1}) = 1$, it follows that

$$\operatorname{tr}(\rho^{\otimes n}\mathcal{X}_\sigma^{\tau_1}\mathcal{X}_\sigma^{\tau_2}) = \operatorname{tr}(\rho^{\otimes n}(\mathcal{X}_\sigma^{\tau_1})^2) = \operatorname{tr}(\rho^{\otimes 2}\mathcal{X}_\sigma^2).$$

If $\tau_1\tau_2$ has cycle type $(2, 2)$, then $\mathcal{X}_\sigma^{\tau_1}$ and $\mathcal{X}_\sigma^{\tau_2}$ act on disjoint components of the tensor product $\rho^{\otimes n}$, so

$$\operatorname{tr}(\rho^{\otimes n}\mathcal{X}_\sigma^{\tau_1}\mathcal{X}_\sigma^{\tau_2}) = \operatorname{tr}(\rho^{\otimes 2}\mathcal{X}_\sigma)^2 = \omega_\sigma^{(2)}(\rho,\rho)^2.$$

If $\tau_1\tau_2$ is a 3-cycle, then, by Definition 3.3.7,

$$\operatorname{tr}(\rho^{\otimes n}\mathcal{X}_\sigma^{\tau_1}\mathcal{X}_\sigma^{\tau_2}) = \omega_\sigma^{(3)}(\rho,\rho,\rho).$$

By the same counting argument used to calculate the variance of the purity observable in Lemma 3.2.1, it holds that

$$\operatorname*{\mathbf{Var}}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] = \frac{1}{\binom{n}{2}}(\operatorname{tr}(\rho^{\otimes 2}\mathcal{X}_\sigma^2) - \omega_\sigma^{(2)}(\rho,\rho)^2) + \frac{2(n-2)}{\binom{n}{2}}(\omega_\sigma^{(3)}(\rho,\rho,\rho) - \omega_\sigma^{(2)}(\rho,\rho)^2). \qquad \blacksquare$$

To bound the variance of $\mathcal{O}_{\chi^2}$, we consider the terms $\operatorname{tr}(\rho^{\otimes 2}\mathcal{X}_\sigma^2) - \omega_\sigma^{(2)}(\rho,\rho)^2$ and $\omega_\sigma^{(3)}(\rho,\rho,\rho) - \omega_\sigma^{(2)}(\rho,\rho)^2$ in Proposition 3.3.9 separately. Let $\Delta = \rho - \sigma$.

Proposition 3.3.10.

$$\omega_\sigma^{(3)}(\rho,\rho,\rho) - \omega_\sigma^{(2)}(\rho,\rho)^2 = \omega_\sigma^{(3)}(\Delta,\Delta,\Delta) + \omega_\sigma^{(3)}(\Delta,\sigma,\Delta) - d_{\chi^2}(\rho,\sigma)^2$$

Proof. Rewriting ρ as $\sigma+\Delta$, expanding by multilinearity of $\omega_\sigma^{(3)}$, and using the contraction properties of $\omega_\sigma^{(3)}$ and $\omega_\sigma^{(2)}$ (Propositions 3.3.2 and 3.3.8), we obtain

$$\begin{aligned}
\omega_\sigma^{(3)}(\rho,\rho,\rho) &= \omega_\sigma^{(3)}(\Delta+\sigma,\Delta+\sigma,\Delta+\sigma)\\
&= \omega_\sigma^{(3)}(\Delta,\Delta,\Delta) + \omega_\sigma^{(3)}(\Delta,\sigma,\Delta) + 2\omega_\sigma^{(2)}(\Delta,\Delta) + 2\omega_\sigma^{(2)}(\sigma,\Delta) + \omega_\sigma^{(2)}(\Delta,\sigma) + \omega_\sigma^{(2)}(\sigma,\sigma)\\
&= \omega_\sigma^{(3)}(\Delta,\Delta,\Delta) + \omega_\sigma^{(3)}(\Delta,\sigma,\Delta) + 2d_{\chi^2}(\rho,\sigma) + 3\operatorname{tr}(\Delta) + 1\\
&= \omega_\sigma^{(3)}(\Delta,\Delta,\Delta) + \omega_\sigma^{(3)}(\Delta,\sigma,\Delta) + 2d_{\chi^2}(\rho,\sigma) + 1,
\end{aligned}$$

where we used the fact that $\operatorname{tr}(\Delta) = \operatorname{tr}(\rho) - \operatorname{tr}(\sigma) = 1 - 1 = 0$ and $d_{\chi^2}(\rho,\sigma) = \omega_\sigma^{(2)}(\Delta,\Delta)$.

Similarly,

$$\begin{aligned}
\omega_\sigma^{(2)}(\rho,\rho)^2 &= (\omega_\sigma^{(2)}(\Delta,\Delta) + 1)^2\\
&= \omega_\sigma^{(2)}(\Delta,\Delta)^2 + 2\omega_\sigma^{(2)}(\Delta,\Delta) + 1\\
&= d_{\chi^2}(\rho,\sigma)^2 + 2d_{\chi^2}(\rho,\sigma) + 1
\end{aligned}$$

Hence,

$$\omega_\sigma^{(3)}(\rho,\rho,\rho) - \omega_\sigma^{(2)}(\rho,\rho)^2 = \omega_\sigma^{(3)}(\Delta,\Delta,\Delta) + \omega_\sigma^{(3)}(\Delta,\sigma,\Delta) - d_{\chi^2}(\rho,\sigma)^2. \qquad \blacksquare$$

We upper bound $\omega_\sigma^{(3)}(\Delta,\Delta,\Delta)$ and $\omega_\sigma^{(3)}(\Delta,\sigma,\Delta)$ separately, and use $-d_{\chi^2}(\rho,\sigma)^2 \le 0$.

Proposition 3.3.11.

$$\omega_\sigma^{(3)}(\Delta,\sigma,\Delta) \le 2d_{\chi^2}(\rho,\sigma).$$

Proof. By definition,

$$\begin{aligned}\omega_\sigma^{(3)}(\Delta, \sigma, \Delta) &= \sum_{i,j,k=1}^{d} \frac{\Delta_{ij}\sigma_{jk}\Delta_{ki}}{\operatorname{avg}\{\beta_i, \beta_j\}\operatorname{avg}\{\beta_i, \beta_k\}} \\ &= \sum_{i,j=1}^{d} \frac{\Delta_{ij}\beta_j\Delta_{ji}}{\operatorname{avg}\{\beta_i, \beta_j\}^2} \\ &= \sum_{i,j=1}^{d} \frac{\beta_j}{\operatorname{avg}\{\beta_i, \beta_j\}} \cdot \frac{|\Delta_{ij}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}} \\ &\le 2 \cdot \sum_{i,j=1}^{d} \frac{|\Delta_{ij}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}} \\ &= 2\omega_\sigma^{(2)}(\Delta, \Delta) \\ &= 2d_{\chi^2}(\rho, \sigma),\end{aligned}$$

where we used the fact that $\sigma = \operatorname{diag}(\beta_1, \dots, \beta_d)$ is a diagonal matrix, $\Delta = \rho - \sigma$ is self-adjoint, and $\beta_j \le 2\operatorname{avg}\{\beta_i, \beta_j\}$. ■

Proposition 3.3.12. *If the smallest eigenvalue of σ is at least $\delta > 0$, then*

$$\omega_\sigma^{(3)}(\Delta, \Delta, \Delta) \le \sqrt{2d/\delta} \cdot d_{\chi^2}(\rho, \sigma)^{3/2}.$$

Proof. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\omega_\sigma^{(3)}(\Delta, \Delta, \Delta) &= \sum_{i,j,k=1}^{d} \frac{\Delta_{ij}\Delta_{jk}\Delta_{ki}}{\operatorname{avg}\{\beta_i, \beta_j\}\operatorname{avg}\{\beta_i, \beta_k\}} \\ &\le \sqrt{\sum_{i,j,k=1}^{d} \frac{|\Delta_{ij}|^2|\Delta_{ki}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}\operatorname{avg}\{\beta_i, \beta_k\}}} \cdot \sqrt{\sum_{i,j,k=1}^{d} \frac{|\Delta_{jk}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}\operatorname{avg}\{\beta_i, \beta_k\}}}\end{aligned}$$

The sum under the first square root is bounded as follows:

$$\begin{aligned}\sum_{i,j,k=1}^{d} \frac{|\Delta_{ij}|^2|\Delta_{ki}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}\operatorname{avg}\{\beta_i, \beta_k\}} &= \sum_{i=1}^{d}\left(\sum_{j=1}^{d} \frac{|\Delta_{ij}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}}\right)^2 \\ &\le \left(\sum_{i,j=1}^{d} \frac{|\Delta_{ij}|^2}{\operatorname{avg}\{\beta_i, \beta_j\}}\right)^2 \\ &= d_{\chi^2}(\rho, \sigma)^2,\end{aligned}$$

where the inequality holds because all the terms in the summation are nonnegative.

For the sum inside the second square root, the following inequality is used:

$$\operatorname{avg}\{\beta_i, \beta_j\}\operatorname{avg}\{\beta_i, \beta_k\} \ge (\delta/2)\operatorname{avg}\{\beta_j, \beta_k\}$$

which holds if $\delta \leq \beta_i, \beta_j, \beta_k, \leq 1$. Thus,

$$\begin{aligned}\sum_{i,j,k=1}^{d} \frac{|\Delta_{jk}|^2}{\operatorname{avg}\{\beta_i,\beta_j\}\operatorname{avg}\{\beta_i,\beta_k\}} &\leq \sum_{i,j,k=1}^{d} \frac{|\Delta_{jk}|^2}{(\delta/2)\operatorname{avg}\{\beta_j,\beta_k\}} \\ &= d\cdot(2/\delta)\cdot\sum_{j,k=1}^{d}\frac{|\Delta_{jk}|^2}{\operatorname{avg}\{\beta_j,\beta_k\}} \\ &= (2d/\delta)\cdot d_{\chi^2}(\rho,\sigma).\end{aligned}$$

Therefore,

$$\omega_\sigma^{(3)}(\Delta,\Delta,\Delta) \leq \sqrt{2d/\delta}\cdot d_{\chi^2}(\rho,\sigma)^{3/2}. \qquad \blacksquare$$

Proposition 3.3.13. *If the smallest eigenvalue of σ is at least $\delta > 0$, then*

$$\mathop{\mathbf{E}}_{\rho^{\otimes 2}}[\mathcal{X}_\sigma^2] \leq 2d^2 + (2d/\delta)\cdot d_{\chi^2}(\rho,\sigma).$$

Proof. By the AM-GM inequality, $\operatorname{avg}\{\beta_i,\beta_j\} \geq \sqrt{\beta_i\beta_j}$. Hence,

$$\begin{aligned}\mathop{\mathbf{E}}_{\rho^{\otimes 2}}[\mathcal{X}_\sigma^2] = \omega_\sigma^{(2)}(\rho,\rho) &= \sum_{i,j=1}^{d}\frac{\rho_{ii}\rho_{jj}}{\operatorname{avg}\{\beta_i,\beta_j\}^2} = \sum_{i,j=1}^{d}\frac{\rho_{ii}\rho_{jj}}{\beta_i\beta_j} = \left(\sum_{i=1}^{d}\frac{\rho_{ii}}{\beta_i}\right)^2 \\ &= \left(\sum_{i=1}^{d}\frac{\Delta_{ii}+\sigma_{ii}}{\beta_i}\right)^2 = \left(\sum_{i=1}^{d}\frac{\Delta_{ii}+\beta_i}{\beta_i}\right)^2 \leq \left(d+\sum_{i=1}^{d}\frac{|\Delta_{ii}|}{\beta_i}\right)^2.\end{aligned}$$

Since $(a+b)^2 \leq 2a^2+2b^2$ for all $a,b\in\mathbb{R}$,

$$\begin{aligned}\mathop{\mathbf{E}}_{\rho^{\otimes 2}}[\mathcal{X}_\sigma^2] &\leq \left(d+\sum_{i=1}^{d}\frac{|\Delta_{ii}|}{\beta_i}\right)^2 \\ &\leq 2d^2 + 2\left(\sum_{i=1}^{d}\frac{|\Delta_{ii}|}{\beta_i}\right)^2 \\ &\leq 2d^2 + 2\left(\sum_{i=1}^{d}\frac{|\Delta_{ii}|}{\sqrt{\delta}\sqrt{\beta_i}}\right)^2 && (\delta \leq \beta_i \text{ for all } i=1,\dots,d) \\ &= 2d^2 + (2/\delta)\cdot\left(\sum_{i=1}^{d}\frac{|\Delta_{ii}|}{\sqrt{\beta_i}}\right)^2 \\ &\leq 2d^2 + (2d/\delta)\cdot\sum_{i=1}^{d}\frac{|\Delta_{ii}|^2}{\beta_i} && (\text{by the Cauchy–Schwarz inequality}) \\ &\leq 2d^2 + (2d/\delta)\cdot\sum_{i,j=1}^{d}\frac{|\Delta_{ij}|^2}{\operatorname{avg}\{\beta_i,\beta_j\}} \\ &= 2d^2 + (2d/\delta)\cdot d_{\chi^2}(\rho,\sigma). \qquad \blacksquare\end{aligned}$$

Therefore, by Proposition 3.3.13, Proposition 3.3.12, and Proposition 3.3.11,

Theorem 3.3.14. *If the smallest eigenvalue of σ is at least $\delta > 0$, then*

$$\operatorname*{\mathbf{Var}}_{\rho^{\otimes n}}[\mathcal{O}_{\chi^2}] \leq \frac{1}{\binom{n}{2}}\left(2d^2 + (2d/\delta)\cdot d_{\chi^2}(\rho,\sigma)\right) + \frac{2(n-2)}{\binom{n}{2}}\left(\sqrt{2d/\delta}\cdot d_{\chi^2}(\rho,\sigma)^{3/2} + 2d_{\chi^2}(\rho,\sigma)\right)$$
$$= O\left(\frac{d^2}{n^2}\cdot d_{\chi^2}(\rho,\sigma) + \frac{\sqrt{d}}{n}\cdot\left(d_{\chi^2}(\rho,\sigma)^{3/2} + d_{\chi^2}(\rho,\sigma)\right)\right).$$

3.3.3. Consequences for testing. Let $\sigma \in \mathrm{B}(\mathbb{C}^d)$ denote a fixed known quantum state of full rank and consider the task of estimating $d_{\chi^2}(\rho,\sigma)$ given n copies $\rho^{\otimes n}$ of ρ. By unitary invariance of d_{χ^2}, we may assume, without loss of generality, that the matrix representation of σ is diagonal in the standard basis, so the averaged χ^2 observable $\mathcal{O}_{\chi^2}$ defined in Definition 3.3.5 is an unbiased estimator for $d_{\chi^2}(\rho,\sigma)$. Thus, Theorem 3.0.4, restated below, follows from Lemma 2.8.1 and the variance bound Theorem 3.3.14:

Theorem 3.0.4. *Let $\sigma \in \mathrm{B}(\mathbb{C}^d)$ be a known quantum state with smallest eigenvalue at least $c\epsilon^2/d$ for some $c > 0$. There is an algorithm that, given $n = O(d/\epsilon^2)$ copies of ρ, with high probability distinguishes between $d_{\chi^2}(\rho,\sigma) \leq 0.99\epsilon^2$ and $d_{\chi^2}(\rho,\sigma) > \epsilon^2$.*

Corollary 3.0.5, which follows directly from Theorem 3.0.4, is a robust "far-in-fidelity vs. close-in-χ^2-divergence" tester with *no* assumption about σ's eigenvalues. A stronger version of this result is proved below, framed in terms of the squared Bures distance $d_{\mathrm{B}}^2(\rho,\sigma) = 2 - 2\,\mathrm{F}(\rho,\sigma) \leq d_{\chi^2}(\rho,\sigma)$ (see Definition 2.5.7):

Corollary 3.3.15. *Let $\sigma \in \mathrm{B}(\mathbb{C}^d)$ denote a fixed known quantum state. There is an algorithm that, given $n = O(d/\epsilon)$ copies of ρ, with high probability distinguishes between $d_{\chi^2}(\rho,\sigma) \leq 0.49\epsilon$ and $d_{\mathrm{B}}^2(\rho,\sigma) > 0.5\epsilon$.*

Proof. Let Φ_η denote the *depolarizing channel* defined by

$$\Phi_\eta(\rho) = (1-\eta)\rho + \eta\frac{\mathbf{1}}{d}.$$

Let $\rho' = \Phi_{c\epsilon}(\rho)$ and $\sigma' = \Phi_{c\epsilon}(\sigma)$, where $c > 0$ is a small absolute constant to be chosen later.

Since d_{χ^2} satisfies the data processing inequality Equation (9), $d_{\chi^2}(\rho',\sigma') \leq d_{\chi^2}(\rho,\sigma) \leq 0.49\epsilon$. On the other hand, if $d_{\mathrm{B}}^2(\rho,\sigma) > 0.5\epsilon$, then, by the triangle inequality,

$$\sqrt{0.5\epsilon} < d_{\mathrm{B}}(\rho,\sigma) \leq d_{\mathrm{B}}(\rho,\rho') + d_{\mathrm{B}}(\rho',\sigma') + d_{\mathrm{B}}(\sigma',\sigma).$$

Since $d_{\mathrm{B}}^2 \leq 2d_{\mathrm{tr}} \leq 2$ (cf. Equation (7)),

$$d_{\mathrm{B}}^2(\rho,\rho') \leq 2d_{\mathrm{tr}}(\rho,\rho') = \|\rho - \rho'\|_1 = \left\|\rho - (1-c\epsilon)\rho - c\epsilon\frac{\mathbf{1}}{d}\right\|_1 = c\epsilon\left\|\rho - \frac{\mathbf{1}}{d}\right\|_1 \leq 2c\epsilon.$$

By a similar argument, $d_{\mathrm{B}}^2(\sigma,\sigma') \leq 2c\epsilon$. Thus,

$$\sqrt{0.5\epsilon} < d_{\mathrm{B}}(\rho,\sigma) \leq d_{\mathrm{B}}(\rho',\sigma') + 2\sqrt{2c\epsilon}.$$

Let c be sufficiently small such that $\sqrt{0.5\epsilon} - 2\sqrt{2c\epsilon} \geq \sqrt{0.495\epsilon}$. Hence, $d_{\mathrm{B}}(\rho',\sigma') > \sqrt{0.495\epsilon}$, so

$$d_{\chi^2}(\rho',\sigma') \geq d_{\mathrm{B}}^2(\rho',\sigma') > 0.495\epsilon.$$

Therefore,

$$\begin{aligned} d_{\chi^2}(\rho,\sigma) \leq 0.49\epsilon &\implies d_{\chi^2}(\rho',\sigma') \leq 0.49\epsilon, \\ d_{\mathrm{B}}^2(\rho,\sigma) > 0.5\epsilon &\implies d_{\chi^2}(\rho',\sigma') > 0.495\epsilon. \end{aligned}$$

Thus, by applying the depolarizing channel $\Phi_{c\epsilon}$ to $\rho^{\otimes n}$ to obtain $(\rho')^{\otimes n}$ and then using the tester from Theorem 3.0.4 with σ' in place of σ and 0.5ϵ in place of ϵ^2, we can with high probability distinguish between $d_{\chi^2}(\rho,\sigma) \leq 0.49\epsilon$ and $d_{\mathrm{B}}^2(\rho,\sigma) > 0.5\epsilon$. ■

CHAPTER 4

Quantum separability testing

In this chapter, we consider the problem of testing if a bipartite quantum state is separable or ϵ-far from all separable states in trace distance. Specifically, given measurement access to copies of an unknown quantum state $\rho \in \mathrm{B}((\mathbb{C}^d)^{\otimes 2})$, the separability testing task is to, with high probability, distinguish between $\rho \in \mathcal{P}$ or $d_{\mathrm{tr}}(\rho, \mathcal{P}) \geq \epsilon$, where $\mathcal{P} = \mathrm{Sep}((\mathbb{C}^d)^{\otimes 2})$ denotes the set of separable quantum states on $(\mathbb{C}^d)^{\otimes 2}$.

In Section 4.3, we prove that at least $\Omega(d^2/\epsilon^2)$ copies of ρ are necessary for this task:

Theorem 4.0.1. *Let $\mathcal{P} = \mathrm{Sep}((\mathbb{C}^d)^{\otimes 2})$ and suppose $\epsilon = \Omega(1/\sqrt{d})$. If there exists an algorithm that, given n copies of an unknown quantum state $\rho \in \mathrm{B}((\mathbb{C}^d)^{\otimes 2})$, with high probability distinguishes between $\rho \in \mathcal{P}$ or $d_{\mathrm{tr}}(\rho, \mathcal{P}) \geq \epsilon$, then $n = \Omega(d^2/\epsilon^2)$.*

In Section 4.1, we introduce completely positive distributions on $[d]^2$ as classical counterparts to separable quantum states on $(\mathbb{C}^d)^{\otimes 2}$. Drawing an analogy between completely positive distributions and directed weighted graphs on $[d]^2$, we establish a sufficient condition for a arbitrary distribution on $[d]^2$ to be ϵ-far from all completely positive distributions.

Then, in Section 4.2, we study the problem of testing whether a distribution on $[d]^2$ is completely positive. Given sample access to a distribution p, the testing task is to reliably distinguish between two cases: (1) p belongs to the set $\mathcal{Q}$ of all completely positive distributions on $[d]^2$, and (2) p is at least ϵ-far from $\mathcal{Q}$ in total variation distance:

Theorem 4.0.2. *Let $\mathcal{Q}$ denote the set of all completely positive distributions on $[d]^2$. If there exists an algorithm that, given n samples of an unknown distribution p on $[d]^2$, with high probability distinguishes between $p \in \mathcal{Q}$ or $d_{\mathrm{TV}}(p, \mathcal{Q}) \geq \epsilon$, then $n = \Omega(d/\epsilon^2)$.*

Sections of this chapter are derived from the work presented in [7].

4.1. Completely positive distributions

There is a well-developed theory of completely positive and copositive matrices (see e.g. [22, Chapter 7]). In this section, we review some known material.

Let d be a positive integer. We consider distributions over the grid $[d]^2 = [d] \times [d] = \{(1,1), (1,2), \ldots, (d,d)\}$ which we represent as matrices $A \in \mathbb{R}^{d \times d}$ with A_{ij} being the probability of sampling (i,j).

Example 4.1.1. If $p \in \mathbb{R}^d$ is a distribution on $[d] = \{1, \ldots, d\}$ represented as a column vector, then pp^{T} is the natural i.i.d. product probability distribution on $[d] \times [d]$ derived from p, with $p_i p_j$ being the probability of sampling (i,j).

Definition 4.1.2. A matrix $A \in \mathbb{R}^{d \times d}$ is ***completely positive*** (CP) if there exist vectors $v_1, \ldots, v_k \in \mathbb{R}^d_{\geq 0}$ with nonnegative entries such that A can be expressed as a convex combination of their projections $v_1v_1^\mathsf{T}, \ldots, v_kv_k^\mathsf{T}$, viz.

$$A = \sum_{i=1}^{k} c_i v_i v_i^\mathsf{T} \tag{12}$$

for some nonnegative real numbers $c_1, \ldots, c_k \in \mathbb{R}$ with $c_1 + \cdots + c_k = 1$.

A distribution on $[d]^2$ represented as a matrix A is *completely positive* if A is a CP matrix.

Remark 4.1.3. For a CP distribution A, the vectors v_i in Equation (12) may be taken to be probability distributions, since one can replace v_i by $v_i/\|v_i\|_1$ and c_i by $c_i\|v_i\|_1^2$. Thus, CP distributions are precisely the mixtures of i.i.d. distributions.

It follows immediately from Definition 4.1.2 that a CP matrix A satisfies three basic properties:

(i) A is symmetric ($A^\mathsf{T} = A$),
(ii) $A_{ij} \geq 0$ for all $i, j \in [d]$, and
(iii) A is positive semidefinite (PSD), denoted $A \geq 0$.

A matrix satisfying these three properties is called ***doubly nonnegative***. Thus, completely positive matrices are doubly nonnegative. However, if $d \geq 5$, then there exist doubly nonnegative matrices which are not completely positive [40].

Example 4.1.4. Let J denote the $d \times d$ matrix with $J_{ij} = 1$ for all $i, j \in [d]$ and let $\mathrm{Unif}_{d^2} = J/d^2$ denote the uniform distribution on $[d]^2$. Since $\mathrm{Unif}_{d^2} = (\frac{1}{d}, \ldots, \frac{1}{d})(\frac{1}{d}, \ldots, \frac{1}{d})^\mathsf{T}$, the uniform distribution on $[d]^2$ is completely positive.

Let CP_d denote the set of completely positive $d \times d$ matrices and let CPD_d denote its subset of completely positive distributions on $[d]^2$. It is well known that CP_d is a cone and that its dual cone consists of ***copositive*** matrices, i.e. matrices M such that $x^\mathsf{T}Mx \geq 0$ for all nonnegative vectors $x \in \mathbb{R}^d_{\geq 0}$. Thus, by cone duality, if $B \notin \mathrm{CP}_d$ is a non-CP matrix, then there exists a copositive matrix W such that $\mathrm{tr}(AW) \geq 0$ for all $A \in \mathrm{CP}_d$ and $\mathrm{tr}(BW) < 0$. This result yields witnesses certifying nonmembership in CPD_d. However, its usefulness is limited by the fact that it provides no quantitative information about how far a nonmember A is from the set CPD_d.

In what follows, we interpret distributions on $[d]^2$ as weighted directed graphs with self-loops and obtain a sufficient condition for a distribution to be ϵ-far in total variation distance from all completely positive distributions, CPD_d, in terms of the maximum value of a cut in the corresponding graph.

We interpret a distribution A on $[d]^2$ as a weighted directed graph G with vertices $V(G) = [d]$ and edges

$$E(G) = \{(i,j) \in [d]^2 \mid A_{ij} > 0\}.$$

A *cut* $x \in \{\pm 1\}^d$ in G is a bipartition of the vertices $V(G) = E_1 \cup E_2$ with $E_1 = \{i \in [d] \mid x_i < 0\}$ and $E_2 = \{i \in [d] \mid x_i > 0\}$. The total weight of edges cut by this bipartition is

$$\sum_{(i,j)\in[d]^2} \frac{1 - x_i x_j}{2} A_{ij} = \mathop{\mathbf{E}}_{(\boldsymbol{i},\boldsymbol{j})\sim A} \frac{1 - x_{\boldsymbol{i}} x_{\boldsymbol{j}}}{2} = \frac{1}{2} - \frac{1}{2} \mathop{\mathbf{E}}_{(\boldsymbol{i},\boldsymbol{j})\sim A} x_{\boldsymbol{i}} x_{\boldsymbol{j}} = \frac{1}{2} - \frac{1}{2} x^\mathsf{T} A x.$$

In particular, if $A = pp^\intercal$ with $p \in \mathbb{R}^d$, then

$$x^\intercal A x = x^\intercal p p^\intercal x = (x^\intercal p)^2 \geq 0.$$

By Remark 4.1.3, a CP distribution is a convex combination of matrices of the form $pp^\intercal$. Thus, the following holds:

Proposition 4.1.5. *If A is a CP distribution, then the total weight of a cut in the graph represented by A is at most $\frac{1}{2}$.*

This fact allows us to prove the following result which gives a sufficient condition for a distribution to be ϵ-far from all CP distributions in ℓ^1 distance. (The matrix norms in the following are entrywise.)

Proposition 4.1.6. *Let A be a distribution on $[d]^2$. If there exists a cut $x \in \{\pm 1\}^d$ with $x^\intercal A x \leq -\epsilon$, then $\|B - A\|_1 \geq \epsilon$ for all $B \in \mathrm{CPD}_d$.*

Proof. Let $B \in \mathrm{CPD}_d$ be arbitrary. By Hölder's inequality, for all $U \in \mathbb{R}^{d \times d}$ with $\|U\|_\infty = 1$,

$$\|B - A\|_1 \geq \operatorname{tr}(U^\intercal (B - A)) = \operatorname{tr}(U^\intercal B) - \operatorname{tr}(U^\intercal A).$$

Let $U = xx^\intercal$. Since $x^\intercal B x \geq 0$ and $\operatorname{tr}(U^\intercal A) = x^\intercal A x \leq -\epsilon$,

$$\|B - A\|_1 \geq x^\intercal B x - x^\intercal A x \geq \epsilon. \qquad \blacksquare$$

4.2. Testing complete positivity

Let d be a positive integer. If d is odd, we can reduce to the case of $d - 1$ by using distributions that don't involve outcome $d \in [d]$, and the asymptotics of $\Omega(d/\epsilon^2)$ remain unchanged. Hence we may assume, without loss of generality, that d is even.

We begin by defining a family of distributions on $[d]^2$ which are ϵ-far from CPD_d. Let $S \subseteq [d]$ be a subset of size $|S| = \frac{d}{2}$. Thus, $|S^\mathsf{c}| = \frac{d}{2}$ and

$$|S \times S^\mathsf{c} \cup S^\mathsf{c} \times S| = |S \times S^\mathsf{c}| + |S^\mathsf{c} \times S| = \frac{d^2}{2}.$$

Let $\phi_S : [d]^2 \to \mathbb{R}$ be the function defined by

$$\phi_S(x) = \begin{cases} 1 + \epsilon, & x \in S \times S^\mathsf{c} \cup S^\mathsf{c} \times S \\ 1 - \epsilon, & \text{otherwise.} \end{cases}$$

Hence,

$$\operatorname*{avg}_{x \in [d]^2} \phi_S(x) = \frac{1}{d^2}\left(\frac{d^2}{2}(1+\epsilon) + \frac{d^2}{2}(1-\epsilon)\right) = 1.$$

So we may think of ϕ_S as a density function with respect to the uniform distribution on $[d]^2$.

Let $x \in \{\pm 1\}^d$ be defined as follows: for all $i \in [d]$, if $i \in S$, then $x_i = 1$, otherwise $x_i = -1$. Let A^S be the matrix defined by $A^S_{ij} = \phi_S((i,j))/d^2$. Thus, A^S is a symmetric distribution on $[d]^2$ and x is a cut. The total weight of this cut is

$$\frac{d^2}{2} \cdot \frac{1+\epsilon}{d^2} = \frac{1}{2} + \frac{\epsilon}{2}.$$

Therefore, for every subset $S \subseteq [d]$, the distribution A^S is *not* completely positive. Moreover, $x^{\mathsf{T}} A^S x = -\epsilon$, so, by Proposition 4.1.6,

$$\|A^S - B\|_1 \geq \epsilon$$

for every CP distribution B (where the matrix norm is entry-wise). In other words, for every subset $S \subseteq [d]$ with $|S| = \frac{d}{2}$, A^S is a distribution on $[d]^2$ which is ϵ-far in ℓ^1 distance from every CP distribution on $[d]^2$.

Fix $\Omega = [d]^2$ and let $\phi : \Omega^n \to \mathbb{R}$ denote the function defined by

$$\phi(x) = \operatorname*{avg}_{\substack{S \subseteq [d] \\ |S| = d/2}} \phi_S(x_1) \cdots \phi_S(x_n).$$

Let $\mathcal{D}_n$ denote the distribution on Ω^n defined by the density ϕ and let $d_{\chi^2}(\_,\_)$ denote the χ^2-distance between probability distributions, i.e. for distributions $\mathcal{P}$ and $\mathcal{Q}$ on Ω,

$$d_{\chi^2}(\mathcal{P}, \mathcal{Q}) = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathcal{Q}} \left[\left(\frac{\mathcal{P}(\boldsymbol{x})}{\mathcal{Q}(\boldsymbol{x})} - 1 \right)^2 \right].$$

The following proposition will be shown to imply our lower bound:

Proposition 4.2.1. *If $d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) \geq \frac{1}{3}$, then $n = \Omega(d/\epsilon^2)$.*

Proof. Let $\mathcal{H}$ denote the uniform distribution over subsets $S \subseteq [d]$ with $|S| = d/2$. Thus,

$$\begin{aligned} d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) &= \left(\sum_{x \in \Omega^n} \frac{\mathcal{D}_n(x)^2}{\mathrm{Unif}_{d^2}^{\otimes n}(x)} \right) - 1 = \left(\sum_{x \in \Omega^n} \frac{\phi(x)^2}{d^{2n}} \right) - 1 = \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}^{\otimes n}} \phi(\boldsymbol{x})^2 - 1 \\ &= \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}^{\otimes n}} \left[\left(\mathop{\mathbf{E}}_{\boldsymbol{S} \sim \mathcal{H}} \phi_{\boldsymbol{S}}(\boldsymbol{x}_1) \cdots \phi_{\boldsymbol{S}}(\boldsymbol{x}_n) \right)^2 \right] - 1 \\ &= \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}^{\otimes n}} \left[\mathop{\mathbf{E}}_{\boldsymbol{S}, \boldsymbol{S}' \sim \mathcal{H}} \phi_{\boldsymbol{S}}(\boldsymbol{x}_1) \cdots \phi_{\boldsymbol{S}}(\boldsymbol{x}_n) \phi_{\boldsymbol{S}'}(\boldsymbol{x}_1) \cdots \phi_{\boldsymbol{S}'}(\boldsymbol{x}_n) \right] - 1 \\ &= \mathop{\mathbf{E}}_{\boldsymbol{S}, \boldsymbol{S}' \sim \mathcal{H}} \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}^{\otimes n}} \phi_{\boldsymbol{S}}(\boldsymbol{x}_1) \cdots \phi_{\boldsymbol{S}}(\boldsymbol{x}_n) \phi_{\boldsymbol{S}'}(\boldsymbol{x}_1) \cdots \phi_{\boldsymbol{S}'}(\boldsymbol{x}_n) - 1 \\ &= \mathop{\mathbf{E}}_{\boldsymbol{S}, \boldsymbol{S}' \sim \mathcal{H}} \left[\left(\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}} \phi_{\boldsymbol{S}}(\boldsymbol{x}) \phi_{\boldsymbol{S}'}(\boldsymbol{x}) \right)^n \right] - 1. \end{aligned}$$

For a subset $E \subseteq [d]$, let χ_E be the ± 1-valued indicator function defined by $\chi_E(x) = 1$ if $x \in E$ and $\chi_E(x) = -1$ otherwise. Note that $\phi_E(x) = 1 - \chi_E(x_1)\chi_E(x_2)\epsilon$ for all $x \in \Omega$. Hence,

$$\phi_{\boldsymbol{S}}(\boldsymbol{x})\phi_{\boldsymbol{S}'}(\boldsymbol{x}) = 1 - (\chi_{\boldsymbol{S}}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}}(\boldsymbol{x}_2) + \chi_{\boldsymbol{S}'}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_2))\epsilon + \chi_{\boldsymbol{S}}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}}(\boldsymbol{x}_2)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_2)\epsilon^2.$$

For a fixed outcome of $\boldsymbol{S}$ and $\boldsymbol{x}$ uniformly random, $\chi_{\boldsymbol{S}}(\boldsymbol{x}_1)$ and $\chi_{\boldsymbol{S}}(\boldsymbol{x}_2)$ are independent uniform ± 1-valued bits. So, in expectation, the terms involving just ϵ in the expression above drop out. Moreover, $\chi_{\boldsymbol{S}}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_1)$ and $\chi_{\boldsymbol{S}}(\boldsymbol{x}_2)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_2)$ are independent. Hence,

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}} \phi_{\boldsymbol{S}}(\boldsymbol{x})\phi_{\boldsymbol{S}'}(\boldsymbol{x}) = 1 - \epsilon^2 \cdot \left(\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \mathrm{Unif}_{d^2}} \chi_{\boldsymbol{S}}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_1) \right)^2$$

Let $\boldsymbol{r} = |\boldsymbol{S} \cap \boldsymbol{S}'|$, where $\boldsymbol{S}, \boldsymbol{S}' \sim \mathcal{H}$, and let $\boldsymbol{\delta}$ denote the mean of $\chi_{\boldsymbol{S}}(\boldsymbol{x}_1)\chi_{\boldsymbol{S}'}(\boldsymbol{x}_1)$ appearing above. It is easy to check that $\boldsymbol{\delta} = 4\boldsymbol{r}/d - 1$. Thus,

$$\begin{aligned} d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) &\le \mathop{\mathbf{E}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\big[\big(1+\epsilon^2\boldsymbol{\delta}^2\big)^n\big] - 1 \\ &\le \mathop{\mathbf{E}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\big[\exp(\epsilon^2\boldsymbol{\delta}^2)^n\big] - 1 \\ &= \mathop{\mathbf{E}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\big[\exp(n\epsilon^2\boldsymbol{\delta}^2)\big] - 1. \end{aligned}$$

Since $\exp(n\epsilon^2\boldsymbol{\delta}^2) - 1 \ge 0$,

$$\mathop{\mathbf{E}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\big[\exp(n\epsilon^2\boldsymbol{\delta})\big] - 1 = \int_0^\infty \mathop{\mathbf{P}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\big[\exp(n\epsilon^2\boldsymbol{\delta}^2) - 1 \ge t\big]dt.$$

Since $\exp(n\epsilon^2\boldsymbol{\delta}^2) - 1 \ge t$ is equivalent to

$$\boldsymbol{r} \ge \frac{d}{4} + \frac{d}{4}\cdot\left(\frac{\log(1+t)}{n\epsilon^2}\right)^{\frac{1}{2}}$$

it follows that

$$d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) \le \int_0^\infty \mathop{\mathbf{P}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\left[\boldsymbol{r} \ge \frac{d}{4} + \frac{d}{4}\sqrt{f(t)}\right]dt,$$

where $f(t) = \log(1+t)/n\epsilon^2$.

Since $\boldsymbol{r} = |\boldsymbol{S} \cap \boldsymbol{S}'|$ is invariant under permutations of $[d]$, it follows that $\boldsymbol{r}$ is distributed according to the hypergeometric distribution with $d/2$ draws from a set of d elements with $d/2$ successes. If $\boldsymbol{X}$ is a random variable distributed according to the hypergeometric distribution with m draws from a set of N elements with k successes, then (see e.g. [54])

$$\mathbf{P}\left[\frac{\boldsymbol{X}}{m} \ge \frac{k}{N} + s\right] \le \exp(-2s^2m).$$

Hence,

$$\mathop{\mathbf{P}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\left[\boldsymbol{r}\cdot\frac{2}{d} \ge \frac{1}{2} + t\right] = \mathop{\mathbf{P}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\left[\boldsymbol{r} \ge \frac{d}{4} + \frac{dt}{2}\right] \le \exp(-dt^2),$$

whence,

$$\mathop{\mathbf{P}}_{\boldsymbol{S},\boldsymbol{S}'\sim\mathcal{H}}\left[\boldsymbol{r} \ge \frac{d}{4}(\sqrt{f(t)}+1)\right] \le \exp(-df(t)/4).$$

Therefore,

$$\begin{aligned} d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) &\le \int_0^\infty \exp(-df(t)/4)dt \\ &= \int_0^\infty \exp\left(-\frac{d}{4n\epsilon^2}\log(1+t)\right)dt \\ &= \int_0^\infty \left(\frac{1}{1+t}\right)^c dt \\ &= \frac{1}{c-1}, \end{aligned}$$

where $c = d/4n\epsilon^2$. Since $d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) \geq 1/3$, it follows that $c \leq 4$, so $n \geq d/16\epsilon^2$. Therefore, $n = \Omega(d/\epsilon^2)$, as needed. ∎

Let $d_{\mathrm{TV}}(\_,\_)$ denote the total variation distance between probability distributions. Let $p \in \mathrm{CPD}_d$ and let q be a distribution ϵ-far from CPD_d.

A testing algorithm $f : ([d]^2)^n \to \{0,1\}$ for complete positivity determines a probability event $E \subseteq ([d]^2)^n$ satisfying $p^{\otimes n}(E) \geq 2/3$ and $q^{\otimes n}(E) \leq 1/3$. Hence, $\mathrm{Unif}_{d^2}^{\otimes n}(E) \geq 2/3$ and, since $\mathcal{D}_n$ is supported on distributions ϵ-far from CPD_d, $\mathcal{D}_n(E) \leq 1/3$. Therefore, $d_{\mathrm{TV}}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) \geq 1/3$ and the following corollary establishes the lower bound:

Corollary 4.2.2 (Equivalent to Theorem 4.0.2)**.** *If $d_{\mathrm{TV}}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) \geq 1/3$, then $n = \Omega(d/\epsilon^2)$.*

Proof. For all distributions μ and ν, $2d_{\mathrm{TV}}(\mu,\nu)^2 \leq d_{\chi^2}(\mu,\nu)$. Hence,

$$(d/4n\epsilon^2 - 1)^{-1} \geq d_{\chi^2}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n}) \geq 2d_{\mathrm{TV}}(\mathcal{D}_n, \mathrm{Unif}_{d^2}^{\otimes n})^2 \geq \frac{2}{9},$$

where the first inequality is obtained in the proof of Proposition 4.2.1. Therefore, $n = \Omega(d/\epsilon^2)$. ∎

4.3. Testing separability

Let d be a positive integer. As in the previous section, we may assume, without loss of generality, that d is even.

Let $\mathcal{H} = \mathbb{C}^d \otimes \mathbb{C}^d$, let $\mathrm{U}(\mathcal{H})$ denote the set of unitary operators on $\mathcal{H}$, and recall (cf. Definition 2.2.9) that $\mathrm{Sep}(\mathcal{H})$ denotes the set of separable states on $\mathcal{H}$.

We begin by defining a family of quantum states which are with high probability $O(\epsilon)$-far from $\mathrm{Sep}(\mathcal{H})$. For $0 \leq \epsilon \leq 1/2$, let D_ϵ be the diagonal matrix on $\mathcal{H}$ defined by

$$\mathsf{D}_\epsilon = \mathrm{diag}\left(\frac{1+2\epsilon}{d^2}, \dots, \frac{1+2\epsilon}{d^2}, \frac{1-2\epsilon}{d^2}, \dots, \frac{1-2\epsilon}{d^2}\right),$$

where $\mathrm{tr}(\mathsf{D}_\epsilon) = 1$, and let $\mathcal{D}$ denote the family of all quantum states on $\mathcal{H}$ with the same spectrum as D_ϵ, viz. $\mathcal{D} = \{U\mathsf{D}_\epsilon U^\dagger \mid U \in \mathrm{U}(\mathcal{H})\}$.

Our lower bound will rely on the following theorem which follows immediately from [45, Lemma 2.22 and Theorem 4.2]:

Theorem 4.3.1. *$\Omega(d^2/\epsilon^2)$ copies are necessary to test whether a quantum state ϱ on $\mathcal{H}$ is the maximally mixed state or $\varrho \in \mathcal{D}$.*

If $\boldsymbol{U}$ is a random unitary on $\mathcal{H}$ distributed according to the Haar measure, then $\boldsymbol{\varrho} = \boldsymbol{U}\mathsf{D}_\epsilon\boldsymbol{U}^\dagger$ is a random element of $\mathcal{D}$. This induced probability measure is invariant under conjugation by a fixed unitary: for all $V \in \mathrm{U}(\mathcal{H})$, $V\boldsymbol{\varrho}V^\dagger$ has the same distribution as $\boldsymbol{\varrho}$. We want to show the following:

Lemma 4.3.2. *There is a universal constant C_0 such that for all $C_0/\sqrt{d} \leq \epsilon \leq 1/2$, the following holds when $\boldsymbol{\varrho} = \boldsymbol{U}\mathsf{D}_\epsilon\boldsymbol{U}^\dagger$ is a uniformly random state in $\mathcal{D}$:*

$$\mathbf{P}[\forall\sigma \in \mathrm{Sep}(\mathcal{H}),\ \|\boldsymbol{\varrho} - \sigma\|_1 \geq 2\epsilon] \geq \frac{2}{3}.$$

As ϵ tends to zero, the elements of $\mathcal{D}$ get closer to the maximally mixed state and eventually become separable, by the Gurvits–Barnum theorem [27]. Indeed, if $\epsilon \leq 1/(2\sqrt{d^2-1})$, then $\mathcal{D} \subseteq \mathrm{Sep}(\mathcal{H})$. Hence, some assumption on ϵ is necessary for Lemma 4.3.2 to hold.

Lemma 4.3.2 and Theorem 4.3.1 easily imply the desired lower bound:

Theorem 4.3.3 (Equivalent of Theorem 4.0.1)**.** *Let ϱ be a quantum state on $\mathbb{C}^d \otimes \mathbb{C}^d$ and let $\epsilon = \Omega(1/\sqrt{d})$. Testing if ϱ is separable or ϵ-far from* Sep$(\mathcal{H})$ *in trace distance requires $\Omega(d^2/\epsilon^2)$ copies of ϱ.*

Proof. Let $\{E_0, E_1\}$ be a measurement corresponding to a separability testing algorithm using n copies of ϱ. To apply the lower bound in Theorem 4.3.1, we use $\{E_0, E_1\}$ to define an algorithm that decides w.h.p. if a state ϱ is equal to the maximally mixed state $\frac{\mathbf{1}}{d^2}$ or $\varrho \in \mathcal{D}$.

Let $\varrho^{\otimes n}$ be given with either $\varrho \in \mathcal{D}$ or $\varrho = \frac{\mathbf{1}}{d^2}$. Note that, for all $\varrho \in \mathcal{D}$, $d_{\mathrm{tr}}(\varrho, \frac{\mathbf{1}}{d^2}) \geq \epsilon$ holds. Let $\boldsymbol{U}$ be a random unitary. If ϱ is the maximally mixed state, then $V\varrho V^\dagger = \varrho$ for all $V \in \mathrm{U}(\mathcal{H})$, so $(\boldsymbol{U}\varrho\boldsymbol{U}^\dagger)^{\otimes n} = \varrho^{\otimes n}$. Otherwise, $\boldsymbol{U}\varrho\boldsymbol{U}^\dagger$ is a random state in $\mathcal{D}$.

Applying the separability test $\{E_0, E_1\}$ to $\boldsymbol{U}\varrho\boldsymbol{U}^\dagger$, we have that:

(i) if $\boldsymbol{U}\varrho\boldsymbol{U}^\dagger = \varrho = \frac{\mathbf{1}}{d^2}$, then $\boldsymbol{U}\varrho\boldsymbol{U}^\dagger$ is separable, so

$$\mathrm{tr}((\boldsymbol{U}\varrho\boldsymbol{U}^\dagger)^{\otimes n} E_1) = \mathrm{tr}(\varrho^{\otimes n} E_1) \geq \frac{2}{3}.$$

(ii) if $\varrho \in \mathcal{D}$, then the probability of error is

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{U}} \mathrm{tr}((\boldsymbol{U}\varrho\boldsymbol{U}^\dagger)^{\otimes n} E_1) &\leq \mathbf{P}[\boldsymbol{U}\varrho\boldsymbol{U}^\dagger \text{ is } \epsilon\text{-close to } \mathrm{Sep}(\mathcal{H})] + \mathbf{P}[\text{test fails} \mid \boldsymbol{U}\varrho\boldsymbol{U}^\dagger \text{ is } \epsilon\text{-far from } \mathrm{Sep}(\mathcal{H})] \\ &\leq \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} = \frac{5}{9},\end{aligned}$$

where the second inequality follows from Lemma 4.3.2.

Thus, using the separability test, we can distinguish w.h.p. between $\varrho = \frac{\mathbf{1}}{d^2}$ and $\varrho \in \mathcal{D}$ using n copies of ϱ. Therefore, by Theorem 4.3.1, $n = \Omega(d^2/\epsilon^2)$. ■

It remains to show that Lemma 4.3.2 holds. Its proof relies on two main facts: first, that Sep$(\mathcal{H})$ is approximated by a polytope with $\exp(O(d))$ vertices which are separable pure states; and, second, that a random element of $\mathcal{D}$ is ϵ-far from a fixed pure state except with probability $\exp(-O(d))$.

The first fact follows from the next lemma which is a rephrasing of [6, Lemma 9.4]:

Lemma 4.3.4. *There exists a constant $C > 0$ such that, for every dimension d, there is a family $\mathcal{N}$ of pure product states on $\mathcal{H}$ (i.e. states of the form $|x \otimes y\rangle\langle x \otimes y|$ with $x, y \in \mathbb{C}^d$) with $|\mathcal{N}| \leq C^d$ satisfying*

$$\mathrm{conv}(\mathcal{N} \cup -\mathcal{N}) \subseteq \mathrm{Sep}_{\pm}(\mathcal{H}) \subseteq 2\,\mathrm{conv}(\mathcal{N} \cup -\mathcal{N}),$$

where Sep$_{\pm}(\mathcal{H})$ *denotes the cyclidrical symmetrization of* Sep$(\mathcal{H})$.

Now, we wish to upper bound the probability that a random element of $\mathcal{D}$ is ϵ-far from a fixed pure state. The following result provides a sufficient condition for a state σ on $\mathcal{H}$ to be ϵ-far from a state $\varrho \in \mathcal{D}$:

Proposition 4.3.5. *Let $\varrho \in \mathcal{D}$ be arbitrary and let $W = \dfrac{\mathbf{1}}{d^2} - \varrho$. For all quantum states σ on $\mathcal{H}$, if $\operatorname{tr}(\sigma W) \geq -\epsilon \|W\|_\infty$, then $\|\varrho - \sigma\|_1 \geq \epsilon$.*

Proof. Note that

$$\begin{aligned} \operatorname{tr}(\varrho W) &= \frac{1}{d^2} - \operatorname{tr}(\varrho^2) = \frac{1}{d^2} - \frac{1+4\epsilon^2}{d^2} = -\frac{4\epsilon^2}{d^2}, \\ \|W\|_\infty &= \left\| \frac{\mathbf{1}}{d^2} - \mathsf{D}_\epsilon \right\|_\infty = \frac{2\epsilon}{d^2}. \end{aligned}$$

By Hölder's inequality for matrices, $\operatorname{tr}((\sigma - \varrho)W) \leq \|\sigma - \varrho\|_1 \cdot \|W\|_\infty$. Hence,

$$\|\sigma - \varrho\|_1 \geq \frac{\operatorname{tr}(\sigma W) - \operatorname{tr}(\varrho W)}{\|W\|_\infty} = 2\epsilon + \frac{\operatorname{tr}(\sigma W)}{\|W\|_\infty}. \qquad \blacksquare$$

When $\sigma = |x\rangle\langle x|$ with $x \in \mathcal{H}$ and $\varrho = U\mathsf{D}_\epsilon U^\dagger$, we have

$$\begin{aligned} \operatorname{tr}(|x\rangle\langle x| W) &= \langle x|W|x\rangle \\ &= \langle x| \left(\frac{\mathbf{1}}{d^2} - U\mathsf{D}_\epsilon U^\dagger \right) |x\rangle \\ &= \langle x|U \left(\frac{\mathbf{1}}{d^2} - \mathsf{D}_\epsilon \right) U^\dagger |x\rangle \\ &= \|W\|_\infty \cdot \langle x|U\mathsf{Z}U^\dagger|x\rangle, \end{aligned} \tag{13}$$

where $\mathsf{Z} = \operatorname{diag}(-1, \ldots, -1, 1, \ldots, 1)$ is just $\mathbf{1}/d^2 - \mathsf{D}_\epsilon$ divided by $\|W\|_\infty$. Hence, $\|\varrho - |x\rangle\langle x|\|_1 \geq \epsilon$ holds if $\langle x|U\mathsf{Z}U^\dagger|x\rangle \geq -\epsilon$.

Since we are interested in the case when $\boldsymbol{\varrho} = \boldsymbol{U}\mathsf{D}_\epsilon \boldsymbol{U}^\dagger$ is random, it suffices to show that $\langle x|\boldsymbol{U}\mathsf{Z}\boldsymbol{U}^\dagger|x\rangle$ concentrates in the interval $[-\epsilon, \epsilon]$. This fact follows easily from the next lemma:

Lemma 4.3.6. *Let k be a positive even integer. If $\boldsymbol{u} \in \mathbb{C}^k$ is a uniformly random unit vector, then, for sufficiently large k,*

$$\mathbf{P}\left[|\langle \boldsymbol{u}|Z|\boldsymbol{u}\rangle| \geq \frac{1}{2} c k^{-1/4} \right] \leq 4 \exp(-\sqrt{k} c^2/8),$$

where $Z = \operatorname{diag}(1, \ldots, 1, -1, \ldots, -1)$ is a $k \times k$ diagonal matrix with $\operatorname{tr}(Z) = 0$ and c may be any positive constant.

Proof. Let $\boldsymbol{u} = (\boldsymbol{a}_1 + i\boldsymbol{b}_1, \ldots, \boldsymbol{a}_k + i\boldsymbol{b}_k) \in \mathbb{C}^k$ be a uniformly random unit vector with $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_k \in \mathbb{R}$ and let $\boldsymbol{v} \in \mathbb{R}^{2k}$ be defined by

$$\boldsymbol{v} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{\frac{k}{2}}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{\frac{k}{2}}, \boldsymbol{a}_{\frac{k}{2}+1}, \ldots, \boldsymbol{a}_k, \boldsymbol{b}_{\frac{k}{2}+1}, \ldots, \boldsymbol{b}_k).$$

Let D be the $2k \times 2k$ diagonal matrix $D = \operatorname{diag}(1, \ldots, 1, -1, \ldots, -1)$ with $\operatorname{tr}(D) = 0$. Thus, $\boldsymbol{v}$ is a uniformly random real unit vector such that $\langle \boldsymbol{v}|D|\boldsymbol{v}\rangle = \langle \boldsymbol{u}|Z|\boldsymbol{u}\rangle$.

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathbb{R}$ be $2k$ standard Gaussian random variables. Let $\boldsymbol{X} = \boldsymbol{x}_1^2 + \cdots + \boldsymbol{x}_k^2$ and $\boldsymbol{Y} = \boldsymbol{y}_1^2 + \cdots + \boldsymbol{y}_k^2$. By the rotational symmetry of multivariate Gaussian random variables, $\boldsymbol{v}$ has the same distribution as

$$\frac{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k)}{\sqrt{\boldsymbol{X} + \boldsymbol{Y}}}.$$

Hence, $\langle \boldsymbol{v}|D|\boldsymbol{v}\rangle$ and $\frac{\boldsymbol{X}-\boldsymbol{Y}}{\boldsymbol{X}+\boldsymbol{Y}}$ have the same distribution. Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent χ^2 random variables with k degrees of freedom each, it holds that (see e.g. [56, Example 2.11])

$$\mathbf{P}\left[\left|\frac{\boldsymbol{X}}{k}-1\right|\geq t\right]\leq 2\exp(-kt^2/8),$$

for all $t\in(0,1)$ and similarly for Y. Hence, for $t=ck^{-1/4}$, we have $\mathbf{P}\big[|\boldsymbol{X}-k|\geq ck^{3/4}\big]\leq 2\exp(-\sqrt{k}c^2/8)$.

If $|\boldsymbol{X}-k|<ck^{3/4}$ and $|\boldsymbol{Y}-k|<ck^{3/4}$, then, for k sufficiently large,

$$|\langle \boldsymbol{v}|D|\boldsymbol{v}\rangle| = \frac{|\boldsymbol{X}-\boldsymbol{Y}|}{\boldsymbol{X}+\boldsymbol{Y}} \leq \frac{2ck^{3/4}}{2k-2ck^{3/4}} = \frac{c}{k^{1/4}-1} < \frac{1}{2}ck^{-1/4}.$$

Hence, $\mathbf{P}[|\langle \boldsymbol{v}|D|\boldsymbol{v}\rangle| < \frac{1}{2}ck^{-1/4}] \geq 1-4\exp(-\sqrt{k}c^2/8)$. ■

If $\boldsymbol{U}$ is a random unitary distributed according to the Haar measure on $\mathrm{U}(\mathcal{H})$ and $x\in\mathcal{H}$ is a fixed unit vector, then $\boldsymbol{u}=\boldsymbol{U}|x\rangle$ is a uniformly random unit vector in $\mathcal{H}$. Hence, we can apply Lemma 4.3.6 to $|\langle\boldsymbol{u}|\mathsf{Z}|\boldsymbol{u}\rangle|$ to get

$$\mathbf{P}[|\langle x|\boldsymbol{U}\mathsf{Z}\boldsymbol{U}^\dagger|x\rangle|\geq\epsilon]\leq 4\exp(-dc^2/8), \tag{14}$$

where c is an arbitrary positive constant and $\epsilon\geq\frac{1}{2}cd^{-1/2}$.

We now have all the elements needed to prove Lemma 4.3.2:

Proof of Lemma 4.3.2. Let $\boldsymbol{\varrho}=\boldsymbol{U}\mathsf{D}_\epsilon\boldsymbol{U}^\dagger$ be a uniformly random element of $\mathcal{D}$ and let $\boldsymbol{W}=\frac{1}{d^2}-\boldsymbol{\varrho}$. Thus, assuming $\epsilon\geq cd^{-1/2}$,

$$\begin{aligned}
&\mathbf{P}[\forall\sigma\in\mathrm{Sep},\ d_{\mathrm{TV}}(\boldsymbol{\varrho},\sigma)\geq\epsilon] && \\
&\quad=\mathbf{P}[\forall\sigma\in\mathrm{Sep},\ \|\boldsymbol{\varrho}-\sigma\|_1\geq 2\epsilon] && \\
&\quad\geq\mathbf{P}[\forall\sigma\in\mathrm{Sep},\ \mathrm{tr}(\sigma\boldsymbol{W})\geq-2\epsilon\|\boldsymbol{W}\|_\infty] && \text{(by Proposition 4.3.5)}\\
&\quad\geq\mathbf{P}[\forall\sigma\in 2\,\mathrm{conv}(\mathcal{N}\cup-\mathcal{N}),\ \mathrm{tr}(\sigma\boldsymbol{W})\geq-2\epsilon\|\boldsymbol{W}\|_\infty] && \text{(by Lemma 4.3.4)}\\
&\quad=\mathbf{P}[\forall|x\rangle\langle x|\in\mathcal{N}\cup-\mathcal{N},\ 2\,\mathrm{tr}(|x\rangle\langle x|\boldsymbol{W})\geq-2\epsilon\|\boldsymbol{W}\|_\infty] && \text{(by convexity)}\\
&\quad=\mathbf{P}\big[\forall|x\rangle\langle x|\in\mathcal{N},\ |\langle x|\boldsymbol{U}\mathsf{Z}\boldsymbol{U}^\dagger|x\rangle|\leq\epsilon\big] && \text{(by Equation (13))}\\
&\quad\geq 1-\sum_{|x\rangle\langle x|\in\mathcal{N}}\mathbf{P}\big[|\langle x|\boldsymbol{U}\mathsf{Z}\boldsymbol{U}^\dagger|x\rangle|>\epsilon\big] && \text{(by the union bound)}\\
&\quad\geq 1-|\mathcal{N}|\cdot 4\exp(-dc^2/8) && \text{(by Equation (14))}\\
&\quad=1-4\exp(d(\log C-c^2/8)) && \text{(since } |\mathcal{N}|=C^d\text{).}
\end{aligned}$$

Hence, if $c=\sqrt{8(\log C+1)}$, then

$$\mathbf{P}[\forall\sigma\in\mathrm{Sep},\ d_{\mathrm{TV}}(\boldsymbol{\varrho},\sigma)\geq\epsilon]\geq 1-4\exp(d(\log C-c^2/8))=1-4\exp(-d)\geq\frac{2}{3},$$

for $d\geq\log 12$. ■

CHAPTER 5

Shadow tomography

In this chapter, we consider the *shadow tomography* estimation problem: given measurement access to copies of an unknown state $\rho \in \mathrm{B}(\mathbb{C}^d)$ and quantum events $A_1, \dots, A_m \in \mathrm{B}(\mathbb{C}^d)$, the shadow tomography task is to with high probability estimate the expected values $\mathbf{E}_\rho[A_1], \mathbf{E}_\rho[A_2], \dots, \mathbf{E}_\rho[A_m]$ to accuracy $\pm\epsilon$.

We show that $n = \widetilde{O}(\log^2(m)\log(d)/\epsilon^4)$ copies are sufficient for shadow tomography:

Theorem 5.0.1. *There is a quantum algorithm that, given parameters $m \in \mathbb{N}$, $0 < \epsilon, \delta < \frac{1}{2}$, and access to unentangled copies of a state $\rho \in \mathrm{B}(\mathbb{C}^d)$, uses*

$$n = \frac{(\log^2 m + \mathrm{L})(\log d)}{\epsilon^4} \cdot O(\mathrm{L}) \qquad (\mathrm{L} = \log(\tfrac{\log d}{\delta\epsilon}))$$

copies of ρ and has the following behavior: when any adversarially or adaptively chosen sequence of observables $A_1, A_2, \dots, A_m \in \mathrm{B}(\mathbb{C}^d)$ with $0 \leq A_i \leq \mathbf{1}$ is presented to the algorithm one-by-one, once A_t is presented, the algorithm responds with an estimate $\widehat{\mu}_i$ of $\mathbf{E}_\rho[A_t] = \mathrm{tr}(\rho A_t)$.

Except with probability at most δ (over the algorithm's measurements), all m estimates satisfy $|\widehat{\mu}_i - \mathbf{E}_\rho[A_t]| \leq \epsilon$.

The shadow tomography algorithm from Theorem 5.0.1 combines an existing quantum state learning algorithm of Aaronson–Chen–Hazan–Kale–Nayak [3] with our novel algorithm for the *quantum threshold search* problem, defined as follows:

Given

- parameters $0 < \epsilon, \delta < \frac{1}{2}$,
- access to unentangled copies of an unknown quantum state $\rho \in \mathrm{B}(\mathbb{C}^d)$,
- a list of d-dimensional observables $0 \leq A_1, \dots, A_m \leq \mathbf{1}$, and
- a list of thresholds $0 \leq \theta_1, \dots, \theta_m \leq 1$,

the quantum threshold search task is to with high probability output correctly

- "$\mathbf{E}_\rho[A_j] > \theta_j - \epsilon$" for some particular j; or else,
- "$\mathbf{E}_\rho[A_i] \leq \theta_i$ for all i".

The output of the algorithm is a sample from a distribution over indices j such that "$\mathbf{E}_\rho[A_j] > \theta_j - \epsilon$" or "$\mathbf{E}_\rho[A_i] \leq \theta_i$ for all i" if no such j exists. The goal is to minimize the number n of copies that are used, while ensuring the probability of a false output statement is at most δ.

We show that $O(\log^2(m)/\epsilon^2)$ copies are sufficient for this problem:

Theorem 5.0.2. *There is an algorithm that performs the quantum threshold search task using*

$$n = \frac{\log^2 m + \mathrm{L}}{\epsilon^2} \cdot O(\mathrm{L}) \qquad (\mathrm{L} = \log(1/\delta))$$

copies of ρ. Furthermore, this algorithm is online *in the sense that:*

1. *The algorithm is initially given only m, ϵ, and δ. It then selects n and obtains $\rho^{\otimes n}$.*
2. *Next, pairs of observables and thresholds $(A_1, \theta_1), (A_2, \theta_2), \ldots$ are presented to the algorithm in sequence. When each (A_t, θ_t) is presented, the algorithm must either "pass", or else halt and output "$\mathbf{E}_\rho[A_t] > \theta_t - \epsilon$."*
3. *If the algorithm passes on all (A_t, θ_t) pairs, then it ends by outputting "$\mathbf{E}_\rho[A_i] \leq \theta_i$ for all i".*

Using our results from Theorem 5.0.1 and Theorem 5.0.2, we additionally give improved copy complexity bounds for the hypothesis selection problem: given m hypothesis states $\sigma_1, \ldots, \sigma_m \in \mathrm{B}(\mathbb{C}^d)$ and measurement access to an unknown state $\rho \in \mathrm{B}(\mathbb{C}^d)$, the hypothesis selection task is to, with high probability, select a hypothesis σ_j such that $d_{\mathrm{tr}}(\rho, \sigma_j) \leq O(\eta) + \epsilon$, where η is the minimum trace distance between ρ and one of the m hypothesis states.

We show that $O(\log^3(m)/\epsilon^2)$ or $O(\log^2(m)\log(d)/\epsilon^4)$ copies are sufficient for this problem:

Theorem 5.0.3. *There is a quantum algorithm that, given m fixed hypothesis states $\sigma_1, \ldots \sigma_m \in \mathrm{B}(\mathbb{C}^d)$, parameters $0 < \epsilon, \delta < \frac{1}{2}$, and access to unentangled copies of a state $\rho \in \mathrm{B}(\mathbb{C}^d)$, uses*

$$n = \min\left\{ \frac{(\log^2 m + \mathrm{L}_1)(\log d)}{\epsilon^4} \cdot O(\mathrm{L}_1), \quad \frac{\log^3 m + \log(\mathrm{L}_2/\delta) \cdot \log m}{\epsilon^2} \cdot O(\mathrm{L}_2 \cdot \log(\mathrm{L}_2/\delta)) \right\}$$

copies of ρ (where $\mathrm{L}_1 = \log(\frac{\log d}{\delta\epsilon})$ and $\mathrm{L}_2 = \log(1/\max\{\eta, \epsilon\})$) and has the following guarantee: except with probability at most δ, it outputs k such that

$$d_{\mathrm{tr}}(\rho, \sigma_k) \leq 3.01\eta + \epsilon, \qquad \textit{where } \eta = \min_i\{d_{\mathrm{tr}}(\rho, \sigma_i)\}.$$

Further, assuming $\eta < \frac{1}{2}(\min_{i \neq j}\{d_{\mathrm{tr}}(\sigma_i, \sigma_j)\} - \epsilon)$ (so there is a unique *σ_i near ρ), one can find the σ_k achieving $d_{\mathrm{tr}}(\rho, \sigma_k) = \eta$ (except with probability at most δ) using only $n = O(\log(m/\delta)/\epsilon^2)$ copies of ρ.*

The chapter starts with the proof of a key "χ^2-stability" result used in the proof of Theorem 5.0.2 involving the probability that a noisy binomial random variable exceeds a given threshold:

Theorem 5.0.4. *Fix a threshold $\theta \in [0, 1]$. Let $\boldsymbol{S} \sim \mathrm{Binomial}(n, p)$ and write $q = 1 - p$. Assume that $\boldsymbol{X}$ is an independent exponential random variable with mean at least $\mathbf{stddev}[\boldsymbol{S}] = \sqrt{pqn}$ (and also at least 1). Let B be the event that $\boldsymbol{S} + \boldsymbol{X} > \theta n$, and assume that $\mathbf{P}[B] < \frac{1}{4}$. Then*

$$d_{\chi^2}((\boldsymbol{S} \mid \overline{B}), \boldsymbol{S}) \lesssim \left(\mathbf{P}[B] \cdot \frac{\mathbf{stddev}[\boldsymbol{S}]}{\mathbf{E}[\boldsymbol{X}]} \right)^2.$$

This chapter draws on material originally published in [11].

5.1. χ^2-stable threshold reporting

Our goal in this section is to prove Theorem 5.0.4 and to show how this classical result applies to quantum states and measurements. We begin with some preparatory facts.

The following is well known [9]:

Proposition 5.1.1. *If $f : \mathbb{R} \to \mathbb{R}$ is 1-Lipschitz, then $\mathbf{Var}[f(\boldsymbol{S})] \leq \mathbf{Var}[\boldsymbol{S}]$ for any random variable $\boldsymbol{S}$.*

Proof. Let $\boldsymbol{S}'$ be an independent copy of $\boldsymbol{S}$. Thus,

$$\mathbf{E}\left[\frac{1}{2}(\boldsymbol{S}-\boldsymbol{S}')^2\right] = \frac{1}{2}\,\mathbf{E}\left[\boldsymbol{S}^2 - 2\cdot\boldsymbol{S}\cdot\boldsymbol{S}' + (\boldsymbol{S}')^2\right] = \frac{1}{2}\cdot\left(2\,\mathbf{E}[\boldsymbol{S}^2] - 2\,\mathbf{E}[\boldsymbol{S}]^2\right) = \mathbf{Var}[\boldsymbol{S}].$$

Similarly, $\mathbf{E}[\frac{1}{2}(f(\boldsymbol{S}) - f(\boldsymbol{S}'))^2] = \mathbf{Var}[f(\boldsymbol{S})]$. Since f is 1-Lipschitz, $\frac{1}{2}(f(\boldsymbol{S}) - f(\boldsymbol{S}'))^2 \leq \frac{1}{2}(\boldsymbol{S}-\boldsymbol{S}')^2$. Therefore, $\mathbf{Var}[f(\boldsymbol{S})] \leq \mathbf{Var}[\boldsymbol{S}]$. ■

We will also need the following inequality:

Lemma 5.1.2. *Fix $p \in [0,1]$ and let $q = 1-p$. If $C = (e-1)^2$, then, for all $\lambda \in [0,1]$,*

$$q + pe^{2\lambda} \leq (1 + Cpq\lambda^2)\cdot(q + pe^{\lambda})^2.$$

Proof. Since $(q+pe^{\lambda})^2 \geq (q+p)^2 = 1$ for $\lambda \geq 0$, it suffices to show that

$$q + pe^{2\lambda} \leq (q+pe^{\lambda})^2 + Cpq\lambda^2 \quad \forall\lambda \in [0,1].$$

Since $p+q=1$,

$$p - p^2 = p(1-p) = (1-q)q = q - q^2.$$

Hence, with $\Lambda = e^{\lambda}$,

$$\begin{aligned}(q+p\Lambda^2) - (q+p\Lambda)^2 &= q + p\Lambda^2 - q^2 - 2qp\Lambda - p^2\Lambda^2 \\ &= \Lambda^2(p-p^2) - 2\Lambda pq + q - q^2 \\ &= pq(\Lambda^2 - 2\Lambda + 1) \\ &= pq(\Lambda-1)^2.\end{aligned}$$

Thus, it suffices to show that, for $\lambda \in [0,1]$,

$$\begin{aligned}pq(e^{\lambda}-1)^2 \leq Cpq\lambda^2 &\iff (e^{\lambda}-1)^2 \leq C\lambda^2 \\ &\iff e^{\lambda} - 1 \leq (e-1)\lambda \\ &\iff e^{\lambda} \leq (1-\lambda) + \lambda e.\end{aligned}$$

The last inequality holds by convexity of e^x, so the result now follows. ■

Given a random variable $\boldsymbol{S}$ and an event B on the same probability space, the following result gives a simpler formula for the χ^2-divergence between the distribution induced by $\boldsymbol{S}$ and the distribution induced by $\boldsymbol{S}$ *conditioned* on the nonoccurrence of the event B. In what follows, the typical mindset is that B is an event that “rarely” occurs, so $\mathbf{P}[\overline{B}]$ is close to 1.

Proposition 5.1.3. *Let $\boldsymbol{S}$ be a discrete random variable and let B be an event on the same probability space with $\mathbf{P}[B] < 1$. For each outcome s of $\boldsymbol{S}$, define $f(s) = \mathbf{P}[B \mid \boldsymbol{S} = s]$. Then*

$$d_{\chi^2}((\boldsymbol{S} \mid \overline{B}), \boldsymbol{S}) = \frac{\mathbf{Var}[f(\boldsymbol{S})]}{\mathbf{P}[\overline{B}]^2}.$$

Proof. By Bayes' theorem,

$$\frac{\mathbf{P}[\boldsymbol{S} = s \mid \overline{B}]}{\mathbf{P}[\boldsymbol{S} = s]} = \frac{(1-f(s))}{\mathbf{P}[\overline{B}]}.$$

Hence,

$$\begin{aligned} d_{\chi^2}((\boldsymbol{S} \mid \overline{B}), \boldsymbol{S}) &= \mathbf{E}\left[\left(1 - \frac{\mathbf{P}[\boldsymbol{S} = s \mid \overline{B}]}{\mathbf{P}[\boldsymbol{S} = s]}\right)^2\right] = \mathbf{E}\left[\left(1 - \frac{1 - f(\boldsymbol{S})}{\mathbf{P}[\overline{B}]}\right)^2\right] \\ &= \frac{1}{\mathbf{P}[\overline{B}]^2}\,\mathbf{E}\big[(f(\boldsymbol{S}) - \mathbf{P}[B])^2\big] = \frac{\mathbf{Var}[f(\boldsymbol{S})]}{\mathbf{P}[\overline{B}]^2}, \end{aligned}$$

where the last step uses $\mathbf{E}[f(\boldsymbol{S})] = \mathbf{P}[B]$. ■

We can now prove Theorem 5.0.4, which we restate for convenience:

Theorem 5.0.4. *Fix a threshold $\theta \in [0, 1]$. Let $\boldsymbol{S} \sim \mathrm{Binomial}(n, p)$ and write $q = 1-p$. Assume that $\boldsymbol{X}$ is an independent exponential random variable with mean at least $\mathbf{stddev}[\boldsymbol{S}] = \sqrt{pqn}$ (and also at least 1). Let B be the event that $\boldsymbol{S} + \boldsymbol{X} > \theta n$, and assume that $\mathbf{P}[B] < \frac{1}{4}$. Then*

$$d_{\chi^2}((\boldsymbol{S} \mid \overline{B}), \boldsymbol{S}) \lesssim \left(\mathbf{P}[B] \cdot \frac{\mathbf{stddev}[\boldsymbol{S}]}{\mathbf{E}[\boldsymbol{X}]}\right)^2.$$

Proof. Write $\lambda = 1/\mathbf{E}[\boldsymbol{X}]$, so $\boldsymbol{X} \sim \mathrm{Exponential}(\lambda)$ and we have the assumptions $\lambda \leq \frac{1}{\sqrt{pqn}}$ and $\lambda \leq 1$. Using Proposition 5.1.3 and $\mathbf{P}[\overline{B}] > \frac{3}{4}$, it suffices to show

$$\mathbf{Var}[f(\boldsymbol{S})] \lesssim \mathbf{P}[B]^2 \cdot pqn\lambda^2,$$

where

$$f(s) = \mathbf{P}[\boldsymbol{X} > \theta n - s] = \min\{1, g(s)\}, \qquad g(s) = \exp(-\lambda(\theta n - s)).$$

Since $y \mapsto \min\{1, y\}$ is 1-Lipschitz, Proposition 5.1.1 tells us that $\mathbf{Var}[f(\boldsymbol{S})] \leq \mathbf{Var}[g(\boldsymbol{S})]$. $\mathbf{Var}[g(\boldsymbol{S})]$ can be computed using the moment-generating function of $\boldsymbol{S} \sim \mathrm{Binomial}(n, p)$, namely $\mathbf{E}[\exp(t\boldsymbol{S})] = (q + pe^t)^n$:

$$\begin{aligned} \mathbf{E}[g(\boldsymbol{S})] &= \mathbf{E}[\exp(-\lambda(\theta n - \boldsymbol{S}))] = \exp(-\lambda\theta n) \cdot (q + pe^{\lambda})^n, \\ \mathbf{E}[g(\boldsymbol{S})^2] &= \mathbf{E}[\exp(-2\lambda(\theta n - \boldsymbol{S}))] = \exp(-2\lambda\theta n) \cdot (q + pe^{2\lambda})^n. \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{Var}[g(\boldsymbol{S})] &= \mathbf{E}[g(\boldsymbol{S})]^2 \cdot \left(\frac{\mathbf{E}[g(\boldsymbol{S})^2]}{\mathbf{E}[g(\boldsymbol{S})]^2} - 1\right) = \mathbf{E}[g(\boldsymbol{S})]^2 \cdot \left(\left(\frac{q + pe^{2\lambda}}{(q + pe^{\lambda})^2}\right)^n - 1\right) \\ &\leq \mathbf{E}[g(\boldsymbol{S})]^2 \cdot \big((1 + 3pq\lambda^2)^n - 1\big) && \text{(by Lemma 5.1.2)} \\ &\lesssim \mathbf{E}[g(\boldsymbol{S})]^2 \cdot pqn\lambda^2 && \left(\text{as } \lambda^2 \leq \frac{1}{pqn}\right) \end{aligned}$$

and it therefore remains to establish

$$\mathbf{E}[g(\boldsymbol{S})] = \exp(-\lambda\theta n) \cdot (q + pe^{\lambda})^n \lesssim \mathbf{P}[B]. \tag{15}$$

Intuitively this holds because $g(s)$ should not be much different than $f(s)$, and $\mathbf{E}[f(\boldsymbol{S})] = \mathbf{P}[B]$ by definition. Formally, we consider two cases: $p \geq \frac{1}{n}$ (intuitively, the main case) and $p \leq \frac{1}{n}$.

Case 1: $p \geq \frac{1}{n}$. In this case we use that $\mathbf{P}[\boldsymbol{S} > pn] \geq \frac{1}{4}$ (see, e.g., [17]), and hence: (i) it must be that $\theta \geq p$, since we are assuming $\mathbf{P}[B] = \mathbf{P}[\boldsymbol{S}+\boldsymbol{X} > \theta n] < \frac{1}{4}$; and, (ii) $\mathbf{P}[B] \geq \mathbf{P}[\boldsymbol{S} > pn] \cdot \mathbf{P}[\boldsymbol{X} \geq (\theta - p)n] \geq \frac{1}{4}\exp(-\lambda(\theta - p)n)$, where the first inequality used independence of $\boldsymbol{S}$ and $\boldsymbol{X}$ and the second inequality used $(\theta - p)n \geq 0$ (by (i)). Thus to establish Inequality (15), it remains to show $\exp(-\lambda\theta n) \cdot (q + pe^{\lambda})^n \lesssim \exp(-\lambda(\theta - p)n)$.

Since $0 < \lambda \leq 1$,

$$e^{\lambda} - 1 = \sum_{i\geq 1} \frac{\lambda^i}{i!} = \lambda + \lambda^2 \sum_{i\geq 2} \frac{\lambda^{i-2}}{i!} \leq \lambda + \lambda^2 \sum_{i\geq 2} \frac{1}{i!} \leq \lambda + \lambda^2 e.$$

By a similar argument, $e^{-\lambda} - 1 \leq -\lambda + \lambda^2 e$. Using these two inequalities and $1 + x \leq e^x$ for $x \in \mathbb{R}$, we obtain

$$\begin{aligned}(q + pe^{\lambda})^n &= (1 + p(e^{\lambda} - 1))^n \leq \exp(p(e^{\lambda} - 1)n) \leq \exp(\lambda pn)\exp(e\lambda^2 \cdot p \cdot n) \qquad \text{and} \\ (q + pe^{\lambda})^n &= \exp(\lambda n)(p + qe^{-\lambda})^n = \exp(\lambda n)(1 + q(e^{-\lambda} - 1))^n \\ &\qquad \leq \exp(\lambda n)\exp(q(e^{-\lambda} - 1)n) \leq \exp(\lambda pn)\exp(e\lambda^2 \cdot q \cdot n).\end{aligned}$$

Hence, $(q + pe^{\lambda})^n \leq \exp(\lambda pn)\exp(e\lambda^2 \cdot \min\{p, q\} \cdot n)$. Since, $\lambda^2 \leq 1/pqn$, by assumption, it follows that $\lambda^2 \min\{p, q\}n \leq 1/\max\{p, q\} \leq 2$, so

$$(q + pe^{\lambda})^n \leq \exp(\lambda pn)\exp(e/\max\{p, q\}) \leq \exp(\lambda pn)\exp(2e).$$

Therefore, $\exp(-\lambda\theta n) \cdot (q + pe^{\lambda})^n \lesssim \exp(-\lambda\theta n)\exp(\lambda pn) = \exp(-\lambda(\theta - p)n)$, as needed.

Case 2: $p \leq \frac{1}{n}$. Since $\lambda \in (0, 1]$, we have $e^{\lambda} \leq 1+2\lambda$. Hence, $q+pe^{\lambda} \leq 1+2p\lambda \leq 1+\frac{2}{n}$, and so $(q+pe^{\lambda})^n \lesssim 1$, meaning that Inequality (15) follows from $\mathbf{P}[B] \geq \mathbf{P}[\boldsymbol{X} > \theta n] = \exp(-\lambda\theta n)$. ■

5.1.1. The quantum version. Having established Theorem 5.0.4, we now show how this result applies to quantum states and measurements. Specifically, we prove that for any quantum event $A \in \mathrm{B}(\mathbb{C}^d)$, there exists a corresponding event $B \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ which exhibits the same statistics as the classical event $\boldsymbol{S} + \boldsymbol{X} > \theta n$ from Theorem 5.0.4 with $\boldsymbol{S} \sim \mathrm{Binomial}(n, \mathrm{tr}(\rho A))$ when $\rho^{\otimes n}$ is measured according to B. Moreover, we also relate the fidelity between the states $\rho^{\otimes n}$ and $\rho^{\otimes n}|_{\sqrt{\mathbf{1}-B}}$ (i.e. the state $\rho^{\otimes n}$ conditioned on the event $\mathbf{1} - B$) to the Bhattacharyya coefficient between $\boldsymbol{S}$ and $(\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \leq \theta n)$ (i.e. $\boldsymbol{S}$ conditioned on the event $\boldsymbol{S} + \boldsymbol{X} \leq \theta n$).

Lemma 5.1.4. *Let $\rho \in \mathrm{B}(\mathbb{C}^d)$ represent an unknown quantum state and let $A \in \mathrm{B}(\mathbb{C}^d)$ be a projection. Let $n \in \mathbb{N}$, let $\lambda > 0$, and let $\theta \in [0, 1]$ be an arbitrary threshold. Let $\boldsymbol{S}$ and $\boldsymbol{X}$ be classical random variables with distributions defined by $\boldsymbol{S} \sim \mathrm{Binomial}(n, \mathbf{E}_{\rho}[A])$ and $\boldsymbol{X} \sim \mathrm{Exponential}(\lambda)$. There exists a quantum event $B \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ such that $\mathbf{E}_{\rho^{\otimes n}}[B] = \mathbf{P}[\boldsymbol{S} + \boldsymbol{X} > \theta n]$ and*

$$\mathrm{F}\Big(\rho^{\otimes n}, \rho^{\otimes n}|_{\sqrt{\mathbf{1}-B}}\Big) = \mathrm{BC}((\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \leq \theta n), \boldsymbol{S}).$$

Proof. Let $\varrho = \rho^{\otimes n}$. Let $A_1 = A$ and $A_0 = \mathbf{1} - A$. For all $x \in \{0, 1\}^n$, let $A_x \in \mathrm{B}(\mathbb{C}^d)^{\otimes n}$ denote the event defined by $A_x = A_{x_1} \otimes A_{x_2} \otimes \cdots \otimes A_{x_n}$. For $k \in \{0, \dots, n\}$, let $E_k \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ be the event defined by

$$E_k = \sum_{\substack{x\in\{0,1\}^n \\ |x|=k}} A_x.$$

Since A is a projection, A_x is also a projection and $A_x A_y = A_y A_x = 0$ for all $x, y \in \{0,1\}^n$ with $x \neq y$. Thus, each E_k is a sum of orthogonal projections, so E_k is a projection as well and $E_k E_\ell = E_\ell E_k = 0$ for all $k, \ell \in \{0, \dots, n\}$ with $k \neq \ell$. Moreover,

$$\sum_{k=0}^{n} E_k = \sum_{x \in \{0,1\}^n} A_x = \mathbf{1}\,.$$

Let $B \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ denote the quantum event defined by

$$B = \sum_{k=0}^{n} \mathbf{P}[\boldsymbol{X} + k > \theta n] \cdot E_k.$$

The statistics of the measurement $\{E_k \mid k = 0, \dots, n\}$ applied to ϱ are distributed as $\mathrm{Binomial}(n, \mathrm{tr}(\rho A))$, so $\mathbf{E}_\varrho[E_k] = \mathbf{P}[\boldsymbol{S} = k]$. Hence,

$$\mathop{\mathbf{E}}_{\varrho}[B] = \sum_{k=0}^{n} \mathbf{P}[\boldsymbol{X} + k > \theta n] \cdot \mathop{\mathbf{E}}_{\varrho}[E_k] = \sum_{k=0}^{n} \mathbf{P}[\boldsymbol{X} + k > \theta n] \cdot \mathbf{P}[\boldsymbol{S} = k] = \mathbf{P}[\boldsymbol{S} + \boldsymbol{X} > \theta n].$$

For all $\ell \in \{0, \dots, n\}$,

$$\sqrt{\mathbf{1} - B} \cdot E_\ell = E_\ell \cdot \sqrt{\mathbf{1} - B} = \sqrt{\mathbf{P}[\boldsymbol{X} + \ell \leq \theta n]} \cdot E_\ell.$$

Hence,

$$\begin{aligned}
\mathrm{tr}\big(\varrho|_{\sqrt{\mathbf{1}-B}} \cdot E_\ell\big) &= \frac{1}{\mathbf{E}_\varrho[\overline{B}]} \cdot \mathrm{tr}(\sqrt{\mathbf{1}-B} \cdot \varrho \cdot \sqrt{\mathbf{1}-B} \cdot E_\ell) \\
&= \frac{1}{\mathbf{E}_\varrho[\overline{B}]} \cdot \mathrm{tr}(E_\ell \cdot \sqrt{\mathbf{1}-B} \cdot \varrho \cdot \sqrt{\mathbf{1}-B} \cdot E_\ell) \\
&= \frac{\mathbf{P}[\boldsymbol{X} + \ell \leq \theta n]}{\mathbf{E}_\varrho[\overline{B}]} \cdot \mathrm{tr}(E_\ell \cdot \varrho \cdot E_\ell) \\
&= \frac{\mathbf{P}[\boldsymbol{X} + \ell \leq \theta n]}{\mathbf{E}_\varrho[\overline{B}]} \cdot \mathop{\mathbf{E}}_{\varrho}[E_\ell] \\
&= \frac{\mathbf{P}[\boldsymbol{X} + \ell \leq \theta n]}{\mathbf{P}[\boldsymbol{S} + \boldsymbol{X} \leq \theta n]} \cdot \mathbf{P}[\boldsymbol{S} = \ell].
\end{aligned}$$

Thus, the measurement $\{E_k \mid k = 0, \dots, n\}$ applied to $\varrho|_{\sqrt{\mathbf{1}-B}}$ yields statistics distributed as $(\boldsymbol{S} \mid \overline{B})$. Therefore, by Proposition 2.8.5,

$$\mathrm{F}\big(\varrho, \varrho|_{\sqrt{\mathbf{1}-B}}\big) = \sum_{k=0}^{n} \sqrt{\mathrm{tr}(\varrho \cdot E_k)} \sqrt{\mathrm{tr}\big(\varrho|_{\sqrt{\mathbf{1}-B}} \cdot E_k\big)} = \mathrm{BC}((\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \leq \theta n), \boldsymbol{S}). \qquad \blacksquare$$

Using Lemma 5.1.4, we obtain the following "quantum version" of Theorem 5.0.4:

Corollary 5.1.5. *Let $\rho \in \mathrm{B}(\mathbb{C}^d)$ represent an unknown quantum state and let $A \in \mathrm{B}(\mathbb{C}^d)$ be a projection. Let $n \in \mathbb{N}$, let $\lambda > 0$, and let $\theta \in [0,1]$ be an arbitrary threshold. Fix $p = \mathbf{E}_\rho[A]$ and let $\boldsymbol{S}$ and $\boldsymbol{X}$ be defined as in Theorem 5.0.4. If p, λ, n, and θ satisfy the* *conditions of Theorem 5.0.4, then there exists a quantum event* $B \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ *such that* $\mathbf{E}_{\rho^{\otimes n}}[B] = \mathbf{P}[\boldsymbol{S} + \boldsymbol{X} > \theta n]$ *and*

$$d_{\mathrm{Bures}}\left(\rho^{\otimes n}, \rho^{\otimes n}\big|_{\sqrt{\mathbf{1}-B}}\right) \lesssim \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] \cdot \frac{\mathbf{stddev}[\boldsymbol{S}]}{\mathbf{E}[\boldsymbol{X}]}.$$

Moreover,

$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] \le \exp(-n\lambda(\theta - (e-1)\mathop{\mathbf{E}}_{\rho}[A])).$$

Proof. Let $\varrho = \rho^{\otimes n}$. By Lemma 5.1.4, there exists a quantum event $B \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ such that $\mathbf{E}_\varrho[B] = \mathbf{P}[\boldsymbol{S} + \boldsymbol{X} > \theta n]$ and $\mathrm{F}(\varrho, \varrho|_{\sqrt{\mathbf{1}-B}}) = \mathrm{BC}((\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \le \theta n), \boldsymbol{S})$. Note that, for all distributions μ and ν, $1 - \mathrm{BC}(\mu, \nu) \le d_{\chi^2}(\mu, \nu)$. Hence, by Lemma 5.1.4 and Theorem 5.0.4, it follows that

$$\begin{aligned}
d_{\mathrm{Bures}}\left(\rho^{\otimes n}, \rho^{\otimes n}\big|_{\sqrt{\mathbf{1}-B}}\right) &= \sqrt{2\left(1 - \mathrm{F}\left(\varrho, \varrho|_{\sqrt{\mathbf{1}-B}}\right)\right)} \\
&= \sqrt{2(1 - \mathrm{BC}((\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \le \theta n), \boldsymbol{S}))} \\
&= d_{\mathrm{H}}((\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \le \theta n), \boldsymbol{S}) \\
&\le \sqrt{d_{\chi^2}((\boldsymbol{S} \mid \boldsymbol{S} + \boldsymbol{X} \le \theta n), \boldsymbol{S})} \\
&\lesssim \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B] \cdot \frac{\mathbf{stddev}[\boldsymbol{S}]}{\mathbf{E}[\boldsymbol{X}]}.
\end{aligned}$$

Since $\mathbf{E}_\varrho[B] = \mathbf{P}[\boldsymbol{S} + \boldsymbol{X} > \theta n]$,

$$\begin{aligned}
\mathop{\mathbf{E}}_{\varrho}[B] &= \mathbf{P}[\boldsymbol{S} + \boldsymbol{X} > \theta n] \\
&\le \mathbf{E}[\exp(-\lambda(\theta n - \boldsymbol{S}))] && (\text{by } \mathbf{P}[\boldsymbol{X} > t] \le \exp(-\lambda t)) \\
&= \exp(-\lambda\theta n)\,\mathbf{E}[\exp(\lambda \boldsymbol{S})] \\
&= \exp(-\lambda\theta n)(1 - p + pe^\lambda)^n && (\mathbf{E}[\exp(\lambda\boldsymbol{S})] \text{ is the m.g.f. of } \boldsymbol{S}) \\
&= \exp(-\lambda\theta n)(1 + p(e^\lambda - 1))^n \\
&\le \exp(-\lambda\theta n)(1 + p(e-1)\lambda)^n && (\text{by } e^x \le 1 + (e-1)x \text{ for } x \in [0,1]) \\
&\le \exp(-\lambda\theta n)\exp((e-1)n\lambda p) && (\text{by } 1 + x \le e^x \text{ for } x \in \mathbb{R}) \\
&= \exp(-n\lambda(\theta - (e-1)p)).
\end{aligned}$$

■

5.2. Threshold search

In this section, we prove Theorem 5.0.2, restated for convenience below:

Theorem 5.0.2. *There is an algorithm that performs the quantum threshold search task using*

$$n = \frac{\log^2 m + \mathrm{L}}{\epsilon^2} \cdot O(\mathrm{L}) \qquad (\mathrm{L} = \log(1/\delta))$$

copies of ρ. *Furthermore, this algorithm is* online *in the sense that:*

1. *The algorithm is initially given only* m, ϵ, *and* δ. *It then selects* n *and obtains* $\rho^{\otimes n}$.

2. *Next, pairs of observables and thresholds* $(A_1, \theta_1), (A_2, \theta_2), \ldots$ *are presented to the algorithm in sequence. When each* (A_t, θ_t) *is presented, the algorithm must either "pass", or else halt and output "*$\mathbf{E}_\rho[A_t] > \theta_t - \epsilon$*."*
3. *If the algorithm passes on all* (A_t, θ_t) *pairs, then it ends by outputting "*$\mathbf{E}_\rho[A_i] \leq \theta_i$ *for all* i*".*

5.2.1. Preliminary reductions. We begin with several reductions that allow us to reduce to the case of projections, and to the case when ϵ, δ, and the θ_i's are all fixed constants:

Reduction to projections. Let $\rho \in \mathrm{B}(\mathbb{C}^d)$ denote the unknown quantum state and let $A_1, \ldots, A_m \in \mathrm{B}(\mathbb{C}^d)$ be the observables in the quantum threshold search problem, which are assumed to be given in an online fashion. If we extend the unknown state ρ to $\rho \otimes |0\rangle\langle 0|$, then by Naimark's Theorem 2.8.2, there exists a projection $\Pi_i \in \mathrm{B}(\mathbb{C}^d \otimes \mathbb{C}^2)$ for each A_i such that $\mathbf{E}_{\rho\otimes|0\rangle\langle 0|}[\Pi_i] = \mathbf{E}_\rho[A_i]$ for all $i = 1, \ldots, m$. Since the state $\rho \otimes |0\rangle\langle 0|$ can be prepared without knowing ρ and this extension increases the dimension of the quantum system only by a constant factor, by replacing ρ with $\rho \otimes |0\rangle\langle 0|$ and each A_i with the corresponding Π_i, it follows that we can assume, without loss of generality, that the observables $A_1, \ldots, A_m$ are projections.

Reduction to $3/4$ **vs.** $1/4$**.** Let $0 < \epsilon < \frac{1}{2}$ be given, and recall that in the threshold search problem the algorithm is presented with a stream of pairs of projections and thresholds, (A_i, θ_i), with the goal of distinguishing the cases $\mathbf{E}_\rho[A_i] > \theta_i$ and $\mathbf{E}_\rho[A_i] \leq \theta_i - \epsilon$. By applying Lemma 2.8.6 with $\tau = 0$, $c = \theta_i - \epsilon/2$, $\delta = 1/4$, and ϵ replaced by $\epsilon/2$, it follows that for some $n_0 = O(1/\epsilon^2)$, each projection A_i may be replaced with a projection $B_i \in \mathrm{B}((\mathbb{C}^d)^{\otimes n_0})$ satisfying

i. if $\mathbf{E}_\rho[A_i] > \theta_i$, then $\mathbf{E}_{\rho^{\otimes n_0}}[B_i] > 3/4$;
ii. if $\mathbf{E}_\rho[A_i] \leq \theta_i - \epsilon$, then $\mathbf{E}_{\rho^{\otimes n_0}}[B_i] \leq 1/4$.

Thus, the general threshold search problem reduces to the "3/4 vs. 1/4" version of threshold search at the expense of an extra factor of $n_0 = O(1/\epsilon^2)$ in the copy complexity. Note that the parameter d increases to d^{n_0}, as well, but – crucially – our Theorem 5.0.2 has no dependence on the dimension parameter.

Reduction to a promise-problem version, with fixed δ**.** So far we have reduced proving Theorem 5.0.2 to proving the following:

Theorem 5.2.1. *There is an algorithm that, given* $m \in \mathbb{N}$ *and* $0 < \delta < \frac{1}{2}$*, first obtains* $n^* = O(\log^2 m + \log(1/\delta)) \cdot \log(1/\delta)$ *copies* $\rho^{\otimes n^*}$ *of an unknown state* $\rho \in \mathrm{B}(\mathbb{C}^d)$*. Next, a sequence of projections* $A_1, \ldots A_m \in \mathbb{C}^{d\times d}$ *is presented to the algorithm (possibly adaptively). After each* A_t*, the algorithm may either* select t*, meaning halt and output the claim "*$\mathbf{E}_\rho[A_t] > 1/4$*", or else* pass *to the next projection. If the algorithm passes on all* m *projections, the algorithm must claim "*$\mathbf{E}_\rho[A_i] \leq 3/4$ *for all* i*". Except with probability at most* δ*, the algorithm's output is correct.*

The most challenging part of the proof is the following similar result, which, operating under the assumption that there exists an observable A_j with $\mathbf{E}_\rho[A_j] \geq 3/4$, with some constant positive probability finds another observable A_t such that $\mathbf{E}_\rho[A_t] \geq 1/3$:

Lemma 5.2.2. *There is an algorithm that, given* $m \in \mathbb{N}$*, first obtains* $n = O(\log^2 m)$ *copies* $\rho^{\otimes n}$ *of an unknown state* $\rho \in \mathrm{B}(\mathbb{C}^d)$*. Next, a sequence of projections* $A_1, \ldots A_m \in \mathrm{B}(\mathbb{C}^d)$*, obeying the promise that* $\mathbf{E}_\rho[A_j] > 3/4$ *for at least one* j*, is presented to the algorithm. After each* A_t*, the algorithm may either halt and* select t*, or else* pass *to the next projection. With probability at least* 0.01*, the algorithm selects a* t *with* $\mathbf{E}_\rho[A_t] \geq 1/3$*.*

One needs a slight bit of care to reduce Theorem 5.2.1 to Lemma 5.2.2 while maintaining the online nature of the algorithm:

Proof of Theorem 5.2.1, assuming Lemma 5.2.2. We will use the algorithm in Lemma 5.2.2 as a kind of "subroutine" for the main theorem. Our first step is to augment this subroutine in the following way:

- Given parameter δ for the main theorem, the subroutine will use a parameter $\delta' = \delta/(C\log(1/\delta))$, where C is a universal constant to be chosen later.
- n is increased from $O(\log^2 m)$ to $n' = O(\log^2 m) + O(\log(1/\delta'))$, where the first $O(\log^2 m)$ copies of ρ are used as usual, and the additional $O(\log(1/\delta'))$ copies are reserved as a "holdout."
- If ever the subroutine is about to halt and select t, it first performs a "failsafe" check: It applies Lemma 2.8.6 with $\tau = 0$, $c = .3$, $\epsilon = .03$, $\delta = \delta'$, and measures with the holdout copies. (Note that $c + \epsilon < 1/3$ and also $c - \epsilon > 1/4$.) If event "B" as defined in Lemma 5.1.4 occurs, the subroutine goes ahead and selects t; otherwise, the algorithm not only passes, but it "aborts", meaning that it automatically passes on all subsequent A_i's without considering them.

We make two observations about this augmented subroutine:

- When run under the promise that $\mathbf{E}_\rho[A_j] > 3/4$ for at least one j, it still selects a t satisfying $\mathbf{E}_\rho[A_t] \geq 1/3$ with probability at least 0.005. This is because the "failsafe" causes an erroneous change of mind with probability at most δ', and we may assume $\delta' \leq 0.005$ for large enough C.
- When run *without* the promise that $\mathbf{E}_\rho[A_j] > 3/4$ for at least one j, the failsafe implies that the probability the algorithm ever selects a t with $\mathbf{E}_\rho[A_t] < 1/4$ is at most δ'.

With the augmented subroutine in hand, we can now give the algorithm that achieves Theorem 5.2.1. The algorithm will obtain $n^* = n' \cdot L$ copies of ρ, where $L = O(\log(1/\delta))$; these are thought of as L "batches", each with of n' copies. As the projections A_i are presented to the algorithm, it will run the augmented subroutine "in parallel" on each batch. If any batch wants to accept a certain A_t, then the overall algorithm halts and outputs "$\mathbf{E}_\rho[A_t] > 1/4$". Otherwise, if all the batches pass on A_t, so too does the overall algorithm. Of course, if the overall algorithm passes on all A_i's, it outputs "$\mathbf{E}_\rho[A_i] \leq 3/4$ for all i".

We now verify the correctness of this algorithm. First, *if* there exists some A_j with $\mathbf{E}_\rho[A_j] > 3/4$, the probability of the algorithm wrongly outputting "$\mathbf{E}_\rho[A_i] \leq 3/4$ for all i" is at most $(1-.005)^L$, which can be made smaller than δ by taking the hidden constant in $L = O(\log(1/\delta))$ suitably large. On the other hand, thanks to the "failsafe" and a union bound, the probability the algorithm ever wrongly outputs "$\mathbf{E}_\rho[A_t] > 1/4$" is at most $L\delta' = L \cdot \delta/(C\log(1/\delta))$, which is again at most δ provided C is sufficiently large. ■

5.2.2. The main algorithm (proof of Lemma 5.2.2). In this subsection, we will prove Lemma 5.2.2. Let $n = n(m)$ and $\lambda = \lambda(m)$ be parameters to be fixed later and let $\theta = 2/3$. As stated in Lemma 5.2.2, we may explicitly assume there exists $i \in [m]$ with $\mathbf{E}_\rho[A_i] \geq 3/4$. For each projection A_i, let B_i denote the event obtained from Lemma 5.1.4. The algorithm proceeds as follows:

> Let ϱ denote the current quantum state, with $\varrho = \rho^{\otimes n}$ initially. Given projection A_i, let B_i be the event obtained from Lemma 5.1.4. Measure the current state ϱ with $(\overline{B}_i, B_i)$ using the canonical implementation. If B_i occurs, halt and select i; otherwise, pass.

Note that the n copies of ρ are only prepared once and reused, and that the current state ϱ collapses to a new state after each measurement.

The algorithm has the following modes of failure:

(FN) the algorithm passes on every observable because the event $\overline{B}_i$ occurs for every $i \in [m]$;
(FP) the algorithm picks an observable A_j with $\mathbf{E}_\rho[A_j] < 1/3$.

We want to show that the algorithm does not make errors of type FP or FN with probability at least 0.01. To this end, we introduce the following notation.

Notation 5.2.3. For $i = 1, \ldots, m$, let:

(1) $\boldsymbol{S}_i$ be a random variable distributed as Binomial$(n, \mathbf{E}_\rho[A_i])$;
(2) $p_i = \mathbf{E}_{\rho^{\otimes n}}[B_i]$ be the probability that B_i would occur if $\rho^{\otimes n}$ were measured with $(\overline{B}_i, B_i)$;
(3) $\varrho_0 = \rho^{\otimes n}$ and let ϱ_i be the quantum state after the ith measurement, *conditioned* on the event $\overline{B}_j$ occurring for all $1 \le j \le i$;
(4) $r_i = \mathbf{E}_{\varrho_{i-1}}[\overline{B}_i]$ be the probability that the event $\overline{B}_i$ occurs assuming all the events $\overline{B}_j$ with $1 \le j \le i-1$ occurred;
(5) $q_i = r_1 \cdots r_i$ be the probability that *all* of the events $\overline{B}_j$ with $1 \le j \le i$ occur;
(6) $s_i = q_{i-1} \cdot \mathbf{E}_{\varrho_{i-1}}[B_i]$ be the probability of observing outcomes $\overline{B}_1, \ldots, \overline{B}_{i-1}, B_i$.

Note that the p_i's refer to a "hypothetical," whereas the r_i's, q_i's, and s_i's concern what actually happens over the course of the algorithm. In particular, q_m is the probability that the algorithm passes on every observable. The following claim shows that, as long as the noise expectation $\mathbf{E}[\boldsymbol{X}] = 1/\lambda$ used in Lemma 5.1.4 is sufficiently large, the probability of a false negative (FN) is bounded above by 4/5:

Claim 5.2.4. *For $\mathbf{E}[\boldsymbol{X}] = \Omega(\sqrt{n})$, there exists $t \in [m]$ such that $q_t \le 4/5$. Moreover, if $t > 1$, then $q_{t-1} \ge 3/4$ and $p_1 + \cdots + p_{t-1} \le 1/4$.*

Proof. By Lemma 5.1.4, $p_i = \mathbf{E}_{\rho^{\otimes n}}[B_i] = \mathbf{P}[\boldsymbol{S}_i + \boldsymbol{X} > \theta n]$. Let $k \in [m]$ be such that $\mathbf{E}_\rho[A_k] \ge 3/4$. Thus, $\boldsymbol{S}_k$ is a binomial random variable with mean at least 3/4. Since $\theta = 2/3 < 3/4$, if n is taken to be a sufficiently large constant,

$$p_k = \mathbf{P}[\boldsymbol{S}_k + \boldsymbol{X} > \theta n] \ge \mathbf{P}[\boldsymbol{S}_k > (2/3)n] \ge 1 - \exp(-1/4).$$

Therefore, there exists a minimal $t \in [m]$ such that $(1-p_1)\cdots(1-p_t) \le \exp(-1/4)$. If $t = 1$, then $q_1 = 1 - p_1 \le \exp(-1/4) \le 4/5$. Otherwise, since t is minimal, it follows that $(1-p_1)\cdots(1-p_{t-1}) \ge \exp(-1/4)$. Hence,

$$\exp(-1/4) \le (1-p_1)\cdots(1-p_{t-1}) \le \exp(-(p_1 + \cdots + p_{t-1})),$$

whence $p_1 + \cdots + p_{t-1} \leq 1/4$. Thus, by Lemma 2.8.7 and Corollary 5.1.5,

$$|(1-p_1)\cdots(1-p_t) - q_t| \leq 2\sum_{i=1}^{t-1} d_{\mathrm{tr}}\left(\rho^{\otimes n}, \rho^{\otimes n}|_{\sqrt{\mathbf{1}-B_i}}\right)$$

$$\lesssim \sum_{i=1}^{t-1} \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B_i] \cdot \frac{\mathbf{stddev}[\boldsymbol{S}_i]}{\mathbf{E}[\boldsymbol{X}]} \leq \frac{\sqrt{n}}{\mathbf{E}[\boldsymbol{X}]} \cdot (p_1 + \ldots + p_{t-1}) \leq \frac{1}{4} \cdot \frac{\sqrt{n}}{\mathbf{E}[\boldsymbol{X}]}.$$

By a similar argument,

$$|(1-p_1)\cdots(1-p_{t-1}) - q_{t-1}| \lesssim \frac{\sqrt{n}}{\mathbf{E}[\boldsymbol{X}]} \cdot (p_1 + \ldots + p_{t-2}) \leq \frac{1}{4} \cdot \frac{\sqrt{n}}{\mathbf{E}[\boldsymbol{X}]}.$$

Therefore, since $3/4 < \exp(-1/4) < 4/5$, we have $q_t \leq 4/5$ and $q_{t-1} \geq 3/4$, for $\mathbf{E}[\boldsymbol{X}] = \Omega(\sqrt{n})$. ∎

Assuming $\mathbf{E}[\boldsymbol{X}] = \Omega(\sqrt{n})$, let $t \in [m]$ be as in Claim 5.2.4. Since $q_m \leq q_t \leq 4/5$, it follows that the probability the algorithm makes an FN error is at most $4/5$. In fact, since $q_t \leq 4/5$, the algorithm will pick an index $i \leq t$ with probability at least $1/5$. Thus, to show that the algorithm succeeds with probability at least 0.01, it suffices to show that w.h.p. the algorithm does not pick an index $i \in \mathcal{B}$, where $\mathcal{B} \subseteq [m]$ is the subset defined by

$$\mathcal{B} = \{i \in [m] \mid 1 \leq i \leq t \text{ and } \mathop{\mathbf{E}}_{\rho}[A_i] < 1/3\}.$$

First, we show that an event B_i with $i \in \mathcal{B}$ is unlikely to occur when the initial state $\rho^{\otimes n}$ is measured according to $(\overline{B}_i, B_i)$:

Claim 5.2.5. *Let $\eta \in (0,1]$, to be specified later. If n is of order $O(\log^2(m/\eta))$, then $p_i \leq (\eta/m)^2$ for all $i \in \mathcal{B}$.*

Proof. By Corollary 5.1.5, for all $i \in [m]$,

$$p_i = \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B_i] \leq \exp(-n\lambda(\theta - (e-1)\mathop{\mathbf{E}}_{\rho}[A_i])).$$

Since $\theta = 2/3$ and $i \in \mathcal{B}$, we have $\mathbf{E}_\rho[A_i] < 1/3$ and $\theta - (e-1)\,\mathbf{E}_\rho[A_i] \geq 0.09$. Since $n\lambda = \Omega(\sqrt{n})$, there exists a constant $C > 0$ such that $n\lambda \geq C\sqrt{n}$. Thus,

$$p_i = \mathop{\mathbf{E}}_{\rho^{\otimes n}}[B_i] \leq \exp(-0.09C\sqrt{n}).$$

Therefore, if $n \geq \log^2((m/\eta)^2)/(0.09C)^2$, then $p_i \leq (\eta/m)^2$. ∎

Next, we show that the algorithm picks an index $i \in [t]$ such that $\mathbf{E}_\rho[A_i] \geq 1/3$ with probability at least 0.03, proving Lemma 5.2.2.

Proof of Lemma 5.2.2. Fix $\eta = 0.01$, so that indeed $n = O(\log^2 m)$ as promised. By Lemma 2.8.8,

$$1 \leq \sqrt{q_t}\,\mathrm{F}(\rho^{\otimes n}, \varrho_t) + \sum_{i=1}^{t} \sqrt{s_i}\sqrt{p_i}.$$

By Claim 5.2.5 and the Cauchy–Schwarz inequality,

$$\sum_{i=1}^{t} \sqrt{s_i}\sqrt{p_i} \leq \frac{\eta}{m} \sum_{i\in\mathcal{B}} \sqrt{s_i} + \sum_{i\notin\mathcal{B}} \sqrt{s_i}\sqrt{p_i} \leq \eta + \sqrt{\sum_{i\notin\mathcal{B}} s_i}\sqrt{\sum_{i\notin\mathcal{B}} p_i},$$

where $i \notin \mathcal{B}$ denotes $i \in [t] \setminus \mathcal{B}$. By Claim 5.2.4, $p_1 + \cdots + p_t \leq 1/4$. Hence,

$$1 - \sqrt{q_t}\,\mathrm{F}(\rho^{\otimes n}, \varrho_t) - \eta \leq \sqrt{\sum_{i\notin\mathcal{B}} s_i}\sqrt{\sum_{i\notin\mathcal{B}} p_i} \leq \frac{1}{2}\sqrt{\sum_{i\notin\mathcal{B}} s_i}.$$

Since $\mathrm{F}(\rho^{\otimes n}, \varrho_t) \leq 1$, $\eta = 0.01$, and, by Claim 5.2.4, $q_t \leq 4/5$, it follows that

$$\frac{1}{2}\sqrt{\sum_{i\notin\mathcal{B}} s_i} \geq 0.99 - \sqrt{4/5} \implies \sum_{i\notin\mathcal{B}} s_i \geq 4 \cdot (0.99 - \sqrt{4/5})^2 \geq 0.03.$$

Since $\sum_{i\notin\mathcal{B}} s_i$ is the probability that the algorithm returns an index $i \in [t]$ with $\mathbf{E}_\rho[A_i] \geq 1/3$, it follows that the algorithm is correct with probability at least 0.03. ■

5.3. Shadow tomography

In this section, we prove our online shadow tomography result, Theorem 5.0.1, using quantum threshold search, Theorem 5.0.2. The reduction from shadow tomography to threshold search is known to follow from a learning algorithm for quantum states due to Aaronson–Chen–Hazan–Kale–Nayak [3]. A detailed proof of this reduction is given below.

Let $\rho \in \mathrm{B}(\mathbb{C}^d)$ be an unknown quantum state. The shadow tomography algorithm is framed as an interaction between a "teacher" and a "student," both of which have measurement access to copies of ρ. The teacher presents a sequence of quantum events $A_1, A_2, \ldots$ to the student. For each event A_t presented by the teacher, the student must output an estimate $\widehat{\mu}_t$ of the expected value $\mu_t = \mathbf{E}_\rho[A_t]$. If the student's estimate $\widehat{\mu}_t$ is too far from μ_t relative to some error tolerance ϵ, the teacher declares a "mistake." When a mistake is declared, the teacher must provide a sufficiently accurate estimate $\mu'_t \approx \mu_t$.

Assuming the existence of an algorithm to emulate the teacher, Aaronson et. al. [3] show that there exists algorithm that can emulate the student that succeeds with probability 1 and makes at most $O(\log(d)/\epsilon^2)$ mistakes:

Theorem 5.3.1 (Aaronson et. al. [3]). *Let ρ be a quantum state on $\mathbb{C}^d$. Let $A_1, \ldots, A_m \in \mathrm{B}(\mathbb{C}^d)$ be a sequence of quantum events, possibly chosen adaptively. Let $\mu_t = \mathbf{E}_\rho[A_t]$ denote the expected value of event A_t. Let $\widehat{\mu}_t$ denote the student's estimate of μ_t.*

Suppose the teacher *satisfies the following properties for each event A_t:*

- *If $|\widehat{\mu}_t - \mu_t| > \epsilon$, the teacher always declares "mistake."*
- *If $|\widehat{\mu}_t - \mu_t| \leq \frac{3}{4}\epsilon$, the teacher always passes.*
- *If $\frac{3}{4}\epsilon < |\widehat{\mu}_t - \mu_t| \leq \epsilon$, the teacher may either pass or declare a mistake.*
- *Whenever the teacher declares "mistake," they must supply a value μ'_t to the student such that $|\mu'_t - \mu_t| \leq \frac{1}{4}\epsilon$.*

Then there is an algorithm for the student *that causes at most $C_0(\log d)/\epsilon^2$ "mistakes" (no matter how many events are presented), where C_0 is a universal constant.*

Note that there is no "δ parameter" above, i.e. algorithm works with probability 1.

The quantum threshold search task can be used to implement a teacher that satisfies the properties stated in Theorem 5.3.1. Thus, to obtain an algorithm for shadow tomography, we will run in parallel the student's algorithm from [3] and our teacher algorithm using threshold search.

As presented in Theorem 5.0.2, the threshold search algorithm halts if the expected value $\mu_t = \mathbf{E}_\rho[A_t]$ *exceeds* a threshold θ_t. To emulate the teacher, we instead need to test if μ_t is sufficiently close to an estimate $\widehat{\mu}_t$. To test for closeness, we essentially use the following logical equivalence:

$$|\mathop{\mathbf{E}}_\rho[A_t] - \widehat{\mu}_t| \leq \epsilon \iff \mathop{\mathbf{E}}_\rho[A_t] \leq \widehat{\mu}_t + \epsilon \text{ and } \mathop{\mathbf{E}}_\rho[\overline{A}_t] \leq 1 - \widehat{\mu}_t + \epsilon.$$

Thus, two threshold tests can be used to determine if μ_t is sufficiently close to $\widehat{\mu}_t$. Recall that the estimate $\widehat{\mu}_t$ is given to the teacher by the student, so it is available for the threshold search routine to use as a threshold value.

When $|\mu_t - \widehat{\mu}_t| > \epsilon$, the teacher is contractually obligated to declare that a mistake has occurred and supply their own, more accurate, estimate μ'_t. To calculate μ'_t, the teacher can directly use Lemma 2.8.6 to naively estimate the expectation of A_t using fresh copies of ρ since the student is guaranteed to not make too many mistakes.

Lemma 5.3.2. *Let ρ be a quantum state on $\mathbb{C}^d$, let $A_1, \ldots, A_m \in \mathrm{B}(\mathbb{C}^d)$ denote quantum events, and let $0 \leq \theta_1, \ldots, \theta_m \leq 1$.*

There is an online algorithm that outputs correctly (except with probability at most δ):

- *"$|\mathbf{E}_\rho[A_j] - \theta_j| > \frac{3}{4}\epsilon$, and in fact $|\mathbf{E}_\rho[A_j] - \mu'_j| \leq \frac{1}{4}\epsilon$", for some particular j and value μ'_j; or else,*
- *"$|\mathbf{E}_\rho[A_i] - \theta_i| \leq \epsilon$ for all i".*

using

$$n = \frac{\log^2 m + \mathrm{L}}{\epsilon^2} \cdot O(\mathrm{L}) \qquad (\mathrm{L} = \log(1/\delta))$$

copies of ρ.

Proof. The algorithm turns each event-threshold pair (A_t, θ_t) given as input into two event-threshold pairs, $(A_t, \theta_t + \epsilon)$ and $(\mathbf{1} - A_t, 1 - \theta_t + \epsilon)$, and applies quantum threshold search with respect to ρ and the resulting $2m$ pairs

$$(A_1, \theta_1), (\mathbf{1} - A_1, 1 - \theta_1 + \epsilon), \ldots, (A_m, \theta_m), (\mathbf{1} - A_m, 1 - \theta_m + \epsilon).$$

with parameters $\epsilon/4$, and $\delta/2$.

Except with probability at most $\delta/2$, the guarantees in Theorem 5.0.2 hold. If threshold search passes on all $2m$ pairs given as input, then $\mathbf{E}_\rho[A_i] \leq \theta_i + \epsilon$ and $\mathbf{E}_\rho[\overline{A}_i] \leq 1 - \theta_i + \epsilon$ for all $i \in [m]$. Otherwise, the algorithm selects either an event A_j such that

$$\mathop{\mathbf{E}}_\rho[A_j] > \theta_j + \epsilon - \frac{1}{4}\epsilon$$

or an event $\mathbf{1} - A_j$ such that

$$\mathop{\mathbf{E}}_\rho[\mathbf{1} - A_j] > 1 - \theta_j + \epsilon - \frac{1}{4}\epsilon.$$

In either case, $|\mathbf{E}_\rho[A_j] - \theta_j| > \frac{3}{4}\epsilon$ holds. In this scenario, the algorithm then uses Lemma 2.8.6 and fresh copies of ρ to output an estimate μ'_t such that $|\mu'_t - \mu_t| < \frac{1}{4}\epsilon$.

Naive expectation estimation, as in Lemma 2.8.6, requires $O(\log(1/\delta)/\epsilon^2)$ fresh copies of ρ where the constant hidden in the $O(\_)$ notation is universal. Since threshold search is applied with parameters $2m$, $\epsilon/4$, and $\delta/2$, the number of copies needed is

$$n = \frac{\log^2(m) + \log(1/\delta)}{\epsilon^2} \cdot O(\log(1/\delta)). \qquad \blacksquare$$

Our online shadow tomography algorithm works by running in parallel the student algorithm of Aaronson et. al. [3] and the custom threshold search algorithm from Lemma 5.3.2 as the teacher:

Proof of Theorem 5.0.1. The algorithm starts by preparing n_0 copies of the quantum state ρ, where n_0 is the number of copies required by Lemma 5.3.2.

When observable A_t is received, the student algorithm from [3] is used to obtain an estimate $\widehat{\mu}_t$, which is presented to the teacher emulated with the custom threshold search algorithm from Lemma 5.3.2.

If the teacher passes, then the student's estimate $\widehat{\mu}_t$ satisfies $|\widehat{\mu}_t - \mu_t| \leq \frac{3}{4}\epsilon \leq \epsilon$, so $\widehat{\mu}_t$ will be the shadow tomography algorithm's estimate of the expected value $\mathbf{E}_\rho[A_t]$. We can then pass to the next observable, reusing the n_0 copies of ρ that we prepared initially.

If the teacher declares "mistake" and supplies a more accurate estimate μ'_t with $|\mu'_t - \mu_t| \leq \frac{1}{4}\epsilon \leq \epsilon$, then μ'_t will be the estimate of μ_t output by our shadow tomography algorithm. When a mistake occurs, the quantum state represented by the initial n_0 copies of ρ is discarded and a new batch of fresh n_0 copies of ρ is prepared for the next observable.

By Lemma 5.3.2, the teacher is guaranteed to satisfy the properties necessary for the student algorithm from [3] to work correctly. Thus, at most $R = \lceil C_0(\log d)/\epsilon^2 \rceil + 1$ mistakes will be declared. Let $\delta_0 = \delta/R$ and suppose that the threshold search algorithm is run each time with parameter δ_0 as its δ. Since the n_0 copies of ρ are discarded after each mistake and the student makes at most R mistakes, all of the invocations of threshold search yield correct outputs except with probability at most $R \cdot \delta_0 = R \cdot \delta/R = \delta$ by the union bound.

Therefore, the number of copies of ρ needed is $n = R \cdot n_0$, i.e.

$$n = \frac{(\log^2(m) + \mathsf{L})(\log d)}{\epsilon^4} \cdot O(\mathsf{L}). \qquad \blacksquare$$

5.4. Hypothesis selection

In this section, we consider the hypothesis selection problem for quantum states: given m quantum states as hypotheses, $\sigma_1, \ldots, \sigma_m \in \mathrm{B}(\mathbb{C}^d)$, and measurement access to copies of an unknown quantum state $\rho \in \mathrm{B}(\mathbb{C}^d)$, the hypothesis selection task is to select a hypothesis σ_j such that $d_{\mathrm{tr}}(\rho, \sigma_j) \leq O(\eta) + \epsilon$, where η is the minimum distance between the unknown state ρ and one of the hypotheses:

$$\eta = \min_{i \in [m]} \{d_{\mathrm{tr}}(\rho, \sigma_i)\}.$$

We show that hypothesis selection can be solved with:

i. $\widetilde{O}(\log^2(m)\log(d)/\epsilon^4)$ copies using the shadow tomography algorithm proved above, Theorem 5.0.1;

ii. $\widetilde{O}(\log^3(m)/\epsilon^2)$ copies using quantum threshold search directly, Theorem 5.0.2;

iii. $O(\log(m)/\epsilon^2)$ copies in the special case where η is smaller than half the minimum distance between any two hypotheses.

Definition 5.4.1. For each pair $(i, j) \in [m]^2$ with $i \neq j$, let $A_{ij} \in \mathrm{B}(\mathbb{C}^d)$ denote the quantum event satisfying

$$d_{\mathrm{tr}}(\sigma_i, \sigma_j) = \mathop{\mathbf{E}}_{\sigma_i}[A_{ij}] - \mathop{\mathbf{E}}_{\sigma_j}[A_{ij}] = \mathrm{tr}((\sigma_i - \sigma_j)A_{ij}).$$

The events A_{ij} exist for all σ_i and σ_j by Proposition 2.5.3. Note that every A_{ij} can be taken to be a projection and that $A_{ji} = \mathbf{1} - A_{ij} = \overline{A_{ij}}$.

Intuitively, if the state ρ is close in trace distance to a hypothesis state σ_k, then the expected value $\mathbf{E}_{\rho}[A_{ij}]$ should be close to the expected value $\mathbf{E}_{\sigma_k}[A_{ij}]$ for all $i, j \in [m]$. Since the hypothesis states are known to the algorithm, the events A_{ij} can be computed without knowing ρ. Thus, using shadow tomography, Theorem 5.0.1, we can obtain estimates $\widehat{\mu}_{ij}$ of the values $\mu_{ij} = \mathbf{E}_{\rho}[A_{ij}]$ for all $i, j \in [m]$ with $i < j$.

Given the estimates $\widehat{\mu}_{ij}$, we consider the m^2 differences $|\mathbf{E}_{\sigma_\ell}[A_{ij}] - \widehat{\mu}_{ij}|$ for each hypothesis state σ_ℓ. Informally, if $\rho \approx \sigma_k$ and $\widehat{\mu}_{ij} \approx \mu_{ij}$, then one expects $|\mathbf{E}_{\sigma_k}[A_{ij}] - \widehat{\mu}_{ij}|$ to be small for all $i, j \in [m]$. Therefore, we will select the hypothesis σ_ℓ that minimizes the maximum possible difference between $\mathbf{E}_{\sigma_\ell}[A_{ij}]$ and $\widehat{\mu}_{ij}$ over all indices $i, j \in [m]$.

Specifically, given parameters $\epsilon > 0$ and $\delta < \frac{1}{2}$ for hypothesis selection, we run our shadow tomography algorithm with parameters $\epsilon/2$, δ, and the $\binom{m}{2}$ quantum events A_{ij} to obtain estimates $\widehat{\mu}_{ij}$ with $|\mathbf{E}_{\rho}[A_{ij}] - \widehat{\mu}_{ij}| \leq \epsilon/2$ for all $i, j \in [m]$. Except with probability at most δ, we obtain estimates $\widehat{\mu}_{ij}$ such that $|\mathbf{E}_{\rho}[A_{ij}] - \widehat{\mu}_{ij}| \leq \epsilon/2$ for all $i, j \in [m]$. We select the hypothesis σ_k that minimizes the expression $\max_{i<j}|\mathbf{E}_{\sigma_k}[A_{ij}] - \widehat{\mu}_{ij}|$:

$$k = \mathop{\mathrm{argmin}}_{\ell \in [m]} \left\{ \max_{i<j} \left\{ |\mathop{\mathbf{E}}_{\sigma_\ell}[A_{ij}] - \widehat{\mu}_{ij}| \right\} \right\}.$$

Let σ_{i^*} for some $i^* \in [m]$ denote one of the hypotheses closest to ρ, i.e. with $d_{\mathrm{tr}}(\rho, \sigma_{i^*}) = \eta$. By definition of σ_k and the guarantees of Theorem 5.0.1, it holds that

$$\begin{aligned}
\max_{i<j} |\mathop{\mathbf{E}}_{\sigma_k}[A_{ij}] - \widehat{\mu}_{ij}| &\leq \max_{i<j} |\mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{ij}] - \widehat{\mu}_{ij}| \\
&\leq \max_{i<j} |\mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{ij}] - \mathop{\mathbf{E}}_{\rho}[A_{ij}] + \mathop{\mathbf{E}}_{\rho}[A_{ij}] - \widehat{\mu}_{ij}| \\
&\leq \max_{i<j} \left\{ |\mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{ij}] - \mathop{\mathbf{E}}_{\rho}[A_{ij}]| + |\mathop{\mathbf{E}}_{\rho}[A_{ij}] - \widehat{\mu}_{ij}| \right\} \\
&\leq d_{\mathrm{tr}}(\rho, \sigma_{i^*}) + \epsilon/2 \\
&= \eta + \epsilon/2.
\end{aligned}$$

Hence, by the triangle inequality,

$$\begin{aligned} d_{\mathrm{tr}}(\rho, \sigma_k) &\leq d_{\mathrm{tr}}(\rho, \sigma_{i^*}) + d_{\mathrm{tr}}(\sigma_{i^*}, \sigma_k) \\ &= \eta + \mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{i^*k}] - \mathop{\mathbf{E}}_{\sigma_k}[A_{i^*k}] \\ &= \eta + \mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{i^*k}] - \widehat{\mu}_{i^*k} + \widehat{\mu}_{i^*k} - \mathop{\mathbf{E}}_{\sigma_k}[A_{i^*k}] \\ &\leq \eta + \max_{i<j} |\mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{ij}] - \widehat{\mu}_{ij}| + \max_{i<j} |\mathop{\mathbf{E}}_{\sigma_k}[A_{ij}] - \widehat{\mu}_{ij}| \\ &\leq \eta + 2 \cdot \max_{i<j} |\mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{ij}] - \widehat{\mu}_{ij}| \\ &\leq \eta + 2 \cdot (\eta + \epsilon/2) \\ &= 3\eta + \epsilon. \end{aligned}$$

It follows that,

Proposition 5.4.2. *The above-described method selects a hypothesis σ_k with $d_{\mathrm{tr}}(\rho, \sigma_k) \leq 3\eta + \epsilon$ (except with probability at most δ), using a number of copies of ρ that is the same as in shadow tomography (up to constant factors).*

Now, we consider the problem of solving hypothesis selection using quantum threshold search directly, i.e. without using the learning algorithm from Theorem 5.3.1. We will need the following result, which is a straightforward special case (cf. Lemma 5.3.2) of the quantum threshold decision algorithm proved in Section 5.5.

Corollary 5.4.3. *Given parameters $0 < \epsilon_0, \delta_0 < \frac{1}{2}$, event-threshold pairs $(A_1, \theta_1), \ldots, (A_{m_0}, \theta_{m_0})$, values $\eta_1, \ldots, \eta_{m_0} > 0$, and measurement access to copies of an unknown state ρ, there exists an algorithm using $n_0 = O(\log(m_0/\delta_0)/{\epsilon_0}^2)$ copies of ρ that, except with probability at most δ_0, correctly outputs:*

- ○ *"there exists j with $|\mathbf{E}_\rho[A_j] - \theta_j| > \eta_j$;" or else,*
- ○ *"$|\mathbf{E}_\rho[A_i] - \theta_i| \leq \eta_i + \epsilon$ for all i."*

Furthermore, the algorithm can be implemented by a projection applied to $\rho^{\otimes n_0}$.

In the remainder of this section, we will use the simpler Hilbert–Schmidt product notation $\langle \rho - \sigma, A_{ij} \rangle$ instead of $\mathbf{E}_\rho[A_{ij}] - \mathbf{E}_\sigma[A_{ij}]$. Using Corollary 5.4.3, we can prove the following:

Proposition 5.4.4. *Fix a threshold value $\nu > 0$. In the setting of hypothesis selection, there exists an algorithm that, except with probability at most δ, correctly ouputs:*

- ○ *one of the hypothesis states σ_ℓ such that $\max_{i<j} |\langle \sigma_\ell - \rho, A_{ij} \rangle| \leq \nu + \epsilon$; or*
- ○ *" $\max_{i<j} |\langle \sigma_\ell - \rho, A_{ij} \rangle| > \nu$ for all $\ell \in [m]$."*

Furthermore, the algorithm requires

$$n = \frac{\log^3(m) + \log(m) \cdot \log(1/\delta)}{\epsilon^2} \cdot O(\log(1/\delta))$$

copies of ρ.

Proof. Clearly, the following equivalence holds

$$\max_{i<j} |\langle \sigma_\ell - \rho, A_{ij} \rangle| \leq \nu \iff \text{there does not exist } A_{ij} \text{ such that } |\langle \sigma_\ell - \rho, A_{ij} \rangle| > \nu.$$

For a given hypothesis σ_ℓ, Corollary 5.4.3 yields a measurement on n_0 copies of ρ that can be used to test if there exists an A_{ij} such that $|\langle \sigma_\ell - \rho, A_{ij}\rangle| > \nu$. Hence, for *each* hypothesis state σ_k with $k \in [m]$, we apply Corollary 5.4.3 with ϵ, $\delta = 1/3$, the $\binom{m}{2}$ observables A_{ij}, $\theta_{ij} = \mathbf{E}_{\sigma_k}[A_{ij}]$, and $\eta_{ij} = \nu$ to obtain a projection B_k on $(\mathbb{C}^d)^{\otimes n_0}$ where $n_0 = O(\log(m)/\epsilon^2)$.

If $\max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| \le \nu \le \nu + \epsilon$, then $\mathbf{E}_{\rho^{\otimes n_0}}[\overline{B}_k] \ge 2/3 > 1/3$ by Corollary 5.4.3. Conversely, if $\mathbf{E}_{\rho^{\otimes n_0}}[\overline{B}_k] > 1/3$, then, since the algorithm errs with probability at most 1/3, it follows that rejection is the correct output, so $\max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| \le \nu + \epsilon$. Therefore,

$$\mathop{\mathbf{E}}_{\rho^{\otimes n_0}}[\overline{B}_k] > 1/3 \iff \max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| \le \nu + \epsilon.$$

To find a hypothesis σ_ℓ satisfying $\max_{i<j}|\langle \sigma_\ell - \rho, A_{ij}\rangle| \le \nu + \epsilon$, we apply threshold search, Theorem 5.0.2, to the projections $\overline{B}_1, \dots, \overline{B}_m$ with thresholds $\theta_1 = \theta_2 = \cdots = \theta_m = 1/2$ and parameters $\epsilon = 1/6$ and $\delta > 0$.

Suppose the output of threshold search is correct. If there exists σ_ℓ such that $\max_{i<j}|\langle \sigma_\ell - \rho, A_{ij}\rangle| \le \nu$, then, by Theorem 5.0.2, the algorithm outputs $k \in [m]$ such that $\mathbf{E}_{\rho^{\otimes n_0}}[\overline{B}_k] > 1/2 - 1/6 = 1/3$. Thus, σ_k satisfies $\max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| \le \nu + \epsilon$.

Otherwise, if $\max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| > \nu$ for all $k \in [m]$, then $\mathbf{E}_{\rho^{\otimes n_0}}[\overline{B}_k] \le 1/3 = 1/2 - 1/6$ for all $k \in [m]$, since the $\overline{B}_k$ test errs with probability at most 1/3. In this case, threshold search will, assuming correctness, pass on all the $\overline{B}_k$'s and output "$\max_{i<j}|\langle \sigma_\ell - \rho, A_{ij}\rangle| > \nu$ for all $\ell \in [m]$."

Since each B_k observable requires $n_0 = O(\log(m)/\epsilon^2)$ copies of ρ, the total number of copies required is $n = n_0 \cdot n_{\mathrm{TS}}(m, 1/6, \delta)$ where $n_{\mathrm{TS}}(\dots)$ is the number of copies required for threshold search, viz.

$$n = \frac{\log^3(m) + \log(m)\cdot\log(1/\delta)}{\epsilon^2} \cdot O(\log(1/\delta)). \qquad \blacksquare$$

Recall that σ_{i^*} denotes the hypothesis state closest to ρ in trace distance. Given σ_ℓ with $\max_{i<j}|\langle \sigma_\ell - \rho, A_{ij}\rangle| \le \nu + \epsilon$, it holds that

$$\begin{aligned} d_{\mathrm{tr}}(\rho, \sigma_\ell) &\le d_{\mathrm{tr}}(\rho, \sigma_{i^*}) + d_{\mathrm{tr}}(\sigma_{i^*}, \sigma_\ell) \\ &= \eta + \mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{i^*\ell}] - \mathop{\mathbf{E}}_{\sigma_k}[A_{i^*\ell}] \\ &= \eta + \mathop{\mathbf{E}}_{\sigma_{i^*}}[A_{i^*\ell}] - \mathop{\mathbf{E}}_{\rho}[A_{i^*\ell}] + \mathop{\mathbf{E}}_{\rho}[A_{i^*\ell}] - \mathop{\mathbf{E}}_{\sigma_k}[A_{i^*\ell}] \\ &\le \eta + \max_{i<j}|\langle \sigma_{i^*} - \rho, A_{ij}\rangle| + \max_{i<j}|\langle \sigma_\ell - \rho, A_{ij}\rangle| \\ &\le 2\eta + \nu + \epsilon. \end{aligned}$$

Thus, if $\max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| < \nu$ for some hypothesis σ_k, then our algorithm from Proposition 5.4.4 can find a hypothesis σ_ℓ with $d_{\mathrm{tr}}(\rho, \sigma_\ell) \le 2\eta + \nu + \epsilon$.

If we assume that η is known to the hypothesis selection algorithm, then we can let $\nu = \eta$, and the hypothesis selection result follows. Otherwise, we iteratively search for η by running the routine above with $\nu = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$, using new copies at each iteration and stopping when threshold search fails to find a hypothesis (i.e. when ν becomes smaller than η) or if $\nu \le \epsilon$. If the algorithm stops because no hypothesis is found, then the last choice of ν satisfies $\nu < \eta$, so the last hypothesis found, say σ_ℓ, will satisfy $d_{\mathrm{tr}}(\rho, \sigma_\ell) \le 2\eta + 2\nu + \epsilon \le 2\eta + 2\eta + \epsilon$. If the algorithm stops because $\nu \le \epsilon$, then $d_{\mathrm{tr}}(\rho, \sigma_\ell) \le 2\eta + 2\epsilon$. In either case, $d_{\mathrm{tr}}(\rho, \sigma_\ell) \le 4\eta + 2\epsilon$. The algorithm stops at round t when $1/2^t \le \max\{2\eta, \epsilon\}$. Hence, $t \ge \log(1/\max\{2\eta,\epsilon\}) = \Theta(\log(1/\max\{\eta,\epsilon\}))$. By tuning the constants, we can make the final guarantee that $d_{\mathrm{tr}}(\rho,\sigma_\ell) \le 3.01\eta + \epsilon$. Therefore,

Proposition 5.4.5. *The above-described method selects σ_ℓ with $d_{\mathrm{tr}}(\rho,\sigma_\ell) \le 3.01\eta + \epsilon$ (except with probability at most δ), using*

$$n = \frac{\log^3 m + \log(\mathrm{L}/\delta)\cdot \log m}{\epsilon^2} \cdot O(\textsc{l}\cdot \log(\textsc{l}/\delta))$$

copies of ρ, where $\textsc{l} = \log(1/\max\{\eta,\epsilon\})$.

We now present an algorithm for hypothesis selection that achieves an improved copy complexity of $O(\log(m/\delta)/\epsilon^2)$ copies in the special case where exactly one correct selection exists.

Let α be the minimum distance between hypothesis states, viz. $\alpha = \min_{i\neq j} d_{\mathrm{tr}}(\sigma_i,\sigma_j)$, and suppose that $\eta < \frac{1}{2}(\alpha - \epsilon)$.

Fix $\nu = \frac{1}{2}(\alpha-\epsilon)$. Since $\eta < \nu$, for all $k \neq i^*$,

$$\begin{aligned}
\max_{i<j} |\langle \sigma_k - \rho, A_{ij}\rangle| &\ge |\langle \sigma_k - \rho, A_{i^*k}\rangle| \\
&= |\langle \sigma_k - \sigma_{i^*} + \sigma_{i^*} - \rho, A_{i^*k}\rangle| \\
&\ge |\langle \sigma_k - \sigma_{i^*}, A_{i^*k}\rangle| - |\langle \sigma_{i^*} - \rho, A_{i^*k}\rangle| \\
&= d_{\mathrm{tr}}(\sigma_k,\sigma_{i^*}) - |\langle \sigma_{i^*} - \rho, A_{i^*k}\rangle| \\
&\ge \alpha - |\langle \sigma_{i^*} - \rho, A_{i^*k}\rangle| \\
&\ge \alpha - \eta \\
&= 2\nu + \epsilon - \eta \\
&> \nu + \epsilon.
\end{aligned}$$

Hence, $\max_{i<j}|\langle \sigma_{i^*} - \rho, A_{ij}\rangle| \le d_{\mathrm{tr}}(\rho,\sigma_{i^*}) = \eta$ and $\max_{i<j}|\langle \sigma_k - \rho, A_{ij}\rangle| > \nu + \epsilon > \eta + \epsilon$ for all $k \neq i^*$.

By the decision-problem version of threshold search, Corollary 5.5.3 from Section 5.5, there exist projections $\overline{B}_1, \dots, \overline{B}_m$ on $(\mathbb{C}^d)^{\otimes n}$ with $n = O(\log(m/\delta)/\epsilon^2)$ such that

$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[\overline{B}_{i^*}] \ge 1 - \delta/(4m)$$
$$\mathop{\mathbf{E}}_{\rho^{\otimes n}}[\overline{B}_k] \le \delta/(4m), \text{ for all } k \neq i^*.$$

Suppose the measurements $B_1, \dots, B_m$ are applied to the state $\rho^{\otimes n}$ sequentially. By the quantum union bound, Corollary 2.8.10, the measurement $\overline{B}_{i^*}$ will accept and all others will reject with probability, over all measurements, at least $1-\delta$ since $4\sum_i \delta/(4m) \le \delta$. We conclude:

Proposition 5.4.6. *Using the assumption $\eta < \frac{1}{2}(\alpha - \epsilon)$, where $\alpha = \min_{i\neq j} d_{\mathrm{tr}}(\sigma_i,\sigma_j)$, the above-described method selects σ_{i^*} (except with probability at most δ), using $n = O(\log(m/\delta)/\epsilon^2)$ copies of ρ.*

Our omnibus Theorem 5.0.3 follows immediately from Proposition 5.4.2, Proposition 5.4.5, and Proposition 5.4.6.

5.5. The quantum threshold decision problem

The decision-problem version of quantum threshold search is as follows: given measurement access to copies of an unknown quantum state $\rho \in \mathrm{B}(\mathbb{C}^d)$, quantum events $A_1, \dots, A_m \in \mathrm{B}(\mathbb{C}^d)$, and threshold values $0 \leq \theta_1, \dots, \theta_m \leq 1$, the task is to, with high probability, distinguish between the following cases:

- $\mathbf{E}_\rho[A_i] \leq \theta_i$ for all $i \in [m]$;
- there exists an event A_j with $\mathbf{E}_\rho[A_j] \geq \theta_j - \epsilon$.

Note that the two possible outputs of the decision problem are not mutually exclusive unless there exists some event A_k with $\mathbf{E}_\rho[A_k] > \theta_k$.

Using a theorem of Harrow–Lin–Montanaro [29, Cor. 11], Aaronson [2] showed that the decision-problem version of quantum threshold search can be solved with $n = O(\log(m)\log(1/\delta)/\epsilon^2)$ copies. In Theorem 5.5.2 below, we prove a new version of the Harrow–Lin–Montanaro theorem, with a mild qualitative improvement, which enables us to reduce the copy complexity of the threshold decision problem slightly, to $n = O(\log(m/\delta)/\epsilon^2)$ (see Corollary 5.5.3).

Recall that a positive operator $\rho \geq 0$ determines a pre-inner product $\langle \_, \_ \rangle_\rho$ which satisfies the Cauchy–Schwarz inequality Equation (1). If one of the two operators in the pre-inner product is positive, the following result holds:

Lemma 5.5.1. *Let $\rho \in \mathrm{B}(\mathbb{C}^d)$ be a quantum state. For all $X, Y \in \mathrm{B}(\mathbb{C}^d)$ with $X \geq 0$,*

$$\langle X, Y \rangle_\rho \leq \sqrt{\mathop{\mathbf{E}}_\rho[X]} \cdot \sqrt{\mathop{\mathbf{E}}_\rho[Y^\dagger X Y]}.$$

Proof. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\langle X, Y \rangle_\rho &= \langle \sqrt{X}, \sqrt{X}Y \rangle_\rho \\ &\leq \sqrt{\langle \sqrt{X}, \sqrt{X} \rangle_\rho} \cdot \sqrt{\langle \sqrt{X}Y, \sqrt{X}Y \rangle_\rho} \\ &= \sqrt{\mathop{\mathbf{E}}_\rho[X]} \cdot \sqrt{\mathop{\mathbf{E}}_\rho[Y^\dagger X Y]}.\end{aligned}$$ ■

Theorem 5.5.2. *Let $A_1, \dots, A_m \in \mathrm{B}(\mathbb{C}^d)$ be quantum events and define $\#A = A_1 + \cdots + A_m$. Given $\nu > 0$, let B be the orthogonal projection onto the span of eigenvectors of $\#A$ with eigenvalue at least ν. For any quantum state $\rho \in \mathrm{B}(\mathbb{C}^d)$,*

$$\max_{i \in [m]} \left\{ \mathop{\mathbf{E}}_\rho[A_i] \right\} - 2\sqrt{\nu} \leq \mathop{\mathbf{E}}_\rho[B] \leq \frac{1}{\nu} \mathop{\mathbf{E}}_\rho[\#A].$$

Proof. By definition of B, $\nu B \leq \#A$. Thus, since $\mathbf{E}_\rho$ is linear and monotonic (Fact 2.3.2)

$$\nu \mathop{\mathbf{E}}_\rho[B] = \mathop{\mathbf{E}}_\rho[\nu B] \leq \mathop{\mathbf{E}}_\rho[\#A].$$

Therefore, $\mathbf{E}_\rho[B] \leq \dfrac{1}{\nu} \mathbf{E}_\rho[\#A]$.

For any $j \in [m]$,

$$\begin{aligned}\mathop{\mathbf{E}}_\rho[\overline{B}] &= \langle \mathbf{1}, \overline{B} \rangle_\rho \\ &= \langle A_j + \overline{A_j}, \overline{B} \rangle_\rho \\ &= \langle A_j, \overline{B} \rangle_\rho + \langle \overline{A_j}, \overline{B} \rangle_\rho\end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{\underset{\rho}{\mathbf{E}}[A_j]} \cdot \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B} A_j \overline{B}]} + \sqrt{\underset{\rho}{\mathbf{E}}[\overline{A_j}]} \cdot \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B}\,\overline{A_j}\,\overline{B}]} && \text{(by Lemma 5.5.1)} \\
&\leq \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B} A_j \overline{B}]} + \sqrt{\underset{\rho}{\mathbf{E}}[\overline{A_j}]} \cdot \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B}\,\overline{B}]} && (\overline{B}\,\overline{A_j}\,\overline{B} \leq \overline{B}\,\mathbf{1}\,\overline{B}) \\
&= \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B} A_j \overline{B}]} + \sqrt{\underset{\rho}{\mathbf{E}}[\overline{A_j}]} \cdot \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B}]} && (\overline{B}^2 = \overline{B}) \\
&\leq \sqrt{\underset{\rho}{\mathbf{E}}[\overline{B} \# A \overline{B}]} + \frac{\mathbf{E}_\rho[\overline{A_j}] + \mathbf{E}_\rho[\overline{B}]}{2} && (A_j \leq \# A \text{ and AM-GM}) \\
&\leq \sqrt{\nu} + \frac{\mathbf{E}_\rho[\overline{A_j}]}{2} + \frac{\mathbf{E}_\rho[\overline{B}]}{2} && (\text{by definition of } B).
\end{aligned}
$$

Therefore, $1 - \mathbf{E}_\rho[B] = \mathbf{E}_\rho[\overline{B}] \leq 2\sqrt{\nu} + \mathbf{E}_\rho[\overline{A_j}] = 2\sqrt{\nu} + 1 - \mathbf{E}_\rho[A_j]$, so

$$\underset{\rho}{\mathbf{E}}[B] \geq \underset{\rho}{\mathbf{E}}[A_j] - 2\sqrt{\nu}.$$

Since the index $j \in [m]$ is arbitrary, the lower bound follows. ■

The algorithm for the decision-problem version of threshold search is similar to Aaronson's (see [2, Lemma 14]):

Corollary 5.5.3. *Given measurement access to copies of an unknown state $\rho \in \mathrm{B}(\mathbb{C}^d)$, quantum events $A_1, \dots, A_m \in \mathrm{B}(\mathbb{C}^d)$, and thresholds $\theta_1, \dots, \theta_m > 0$, there exists an algorithm that, except with probability at most δ, outputs correctly:*

- ○ *"there exists j with $\mathbf{E}_\rho[A_j] > \theta_j - \epsilon$"; or*
- ○ *"$\mathbf{E}_\rho[A_i] \leq \theta_i$ for all i".*

The algorithm requires only $n = O(\log(m/\delta)/\epsilon^2)$ copies of ρ and it can be implemented by applying a projection to $\rho^{\otimes n}$.

Proof. Let $\varrho = \rho^{\otimes n}$ denote the n copies of ρ that the algorithm has measurement access to. For each event A_i with $i \in [m]$, let $A_i' \in \mathrm{B}((\mathbb{C}^d)^{\otimes n})$ denote the amplification of A_i obtained by applying Lemma 2.8.6 with $\delta^3/16m$ as the δ parameter to A_i. Thus, for all $i \in [m]$,

$$
\begin{aligned}
\underset{\rho}{\mathbf{E}}[A_i] > \theta_i &\implies E_\varrho[A_i'] > 1 - \delta^3/(16m), \\
\underset{\rho}{\mathbf{E}}[A_i] \leq \theta_i - \epsilon &\implies E_\varrho[A_i'] \leq \delta^3/(16m).
\end{aligned}
$$

Let B denote the projection onto the span generated by eigenvectors of $A_1' + A_2' + \cdots + A_m'$ with eigenvalue at least $\nu = \delta^2/16$. By Theorem 5.5.2,

$$\max_{i \in [m]} \{\underset{\varrho}{\mathbf{E}}[A_i']\} - \delta/2 \leq \underset{\varrho}{\mathbf{E}}[B] \leq (16/\delta^2) \underset{\varrho}{\mathbf{E}}[\# A].$$

If there exists an $i \in [m]$ such that $\mathbf{E}_\rho[A_i] > \theta_i$, then $\mathbf{E}_\varrho[A_i'] > 1 - \delta^3/(16m) \geq 1 - \delta/2$, so

$$\underset{\varrho}{\mathbf{E}}[B] \geq \max_i \underset{\varrho}{\mathbf{E}}[A_i'] - \delta/2 \geq 1 - \delta/2 - \delta/2 = 1 - \delta.$$

Otherwise, if $\mathbf{E}_\rho[A_i] \le \theta_i$ for all $i \in [m]$, then $\mathbf{E}_\varrho[A'_i] \le \delta^3/(16m)$ for all $i \in [m]$, so

$$\begin{aligned}\mathop{\mathbf{E}}_{\varrho}[B] &\le \frac{16}{\delta^2} \mathop{\mathbf{E}}_{\varrho}[A'_1 + \cdots + A'_m] \\ &= \frac{16}{\delta^2} \cdot \left(\mathop{\mathbf{E}}_{\varrho}[A'_1] + \cdots + \mathop{\mathbf{E}}_{\varrho}[A'_m] \right) \\ &\le \frac{16}{\delta^2} \cdot m \cdot \frac{\delta^3}{16m} \\ &= \delta.\end{aligned}$$

Therefore, the algorithm can measure $\varrho = \rho^{\otimes n}$ with $\{B, \overline{B}\}$ and output "there exists..." if B occurs and "$\mathbf{E}_\rho[A_i] \le \theta_i$ for all i" otherwise. As the analysis above shows, this algorithm is correct except with probability at most δ. ■

Bibliography

[1] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM, jun 2018. 13, 14, 15

[2] Scott Aaronson. Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368–STOC18–394, 2020. 85, 86

[3] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. *Journal of Statistical Mechanics: Theory and Experiment*, (12):124019, 14, 2019. 14, 67, 78, 79, 80

[4] Scott Aaronson and Guy N. Rothblum. Gentle measurement of quantum states and differential privacy. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333. ACM, New York, 2019. 13, 14, 15, 35

[5] Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature Communications*, 6(8498), 2015. 11

[6] Guillaume Aubrun and Stanisław J. Szarek. *Alice and Bob meet Banach*, volume 223 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2017. The interface of asymptotic geometric analysis and quantum information theory. 12, 63

[7] Costin Bădescu and Ryan O'Donnell. Lower bounds for testing complete positivity and quantum separability. In *LATIN 2020: Theoretical Informatics*, pages 375–386. Springer International Publishing, 2020. 11, 57

[8] Andrew J. Baldwin and Jonathan A. Jones. Efficiently computing the uhlmann fidelity for density matrices. *Phys. Rev. A*, 107:012427, Jan 2023. 27

[9] Alain Berlinet. A note on variance reduction. *Statistics & Probability Letters*, 25(4):357–360, 1995. 68

[10] Samuel L. Braunstein and Carlton M. Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72:3439–3443, May 1994. 27, 28

[11] Costin Bădescu and Ryan O'Donnell. Improved quantum data analysis. *TheoretiCS*, 3:Art. 7, 34, 2024. 13, 14, 15, 68

[12] Costin Bădescu, Ryan O'Donnell, and John Wright. Quantum state certification. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 503–514. ACM, New York, 2019. 10, 11, 37

[13] 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 2021*, pages 574–585. IEEE Computer Soc., Los Alamitos, CA, 2022. 15

[14] Benoît Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006. 38

[15] Marcus da Silva, Olivier Landon-Cardinal, and David Poulin. Practical characterization of quantum devices without tomography. *Physical Review Letters*, 107(21):210404, 2011. 11

[16] Persi Diaconis. Finite forms of de Finetti's theorem on exchangeability. *Synthese*, 36(2):271–281, October 1977. 12

[17] Benjamin Doerr. An elementary analysis of the probability that a binomial random variable exceeds its expectation. *Statistics & Probability Letters*, 139:67–74, 2018. 71

[18] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2), Feb 2004. 12

[19] Steven Flammia and Yi-Kai Liu. Direct fidelity estimation from few Pauli measurements. *Physical Review Letters*, 106(23):230501, 2011. 11

[20] Christopher A. Fuchs and Carlton M. Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3(3):345–356, 1995. 26, 27, 33

[21] Jingliang Gao. Quantum union bounds for sequential projective measurements. *Physical Review A*, 92:052331, Nov 2015. 36

[22] Bernd Gärtner and Jiří Matoušek. *Approximation Algorithms and Semidefinite Programming.* Springer Berlin Heidelberg, 2012. 57

[23] Oded Goldreich. *Introduction to property testing.* Cambridge University Press, Cambridge, 2017. 29

[24] Roe Goodman and Nolan Wallach. *Symmetry, representations, and invariants.* Springer, 2009. 28, 29

[25] Otfried Gühne and Géza Tóth. Entanglement detection. *Physics Reports*, 474(1-6):1–75, Apr 2009. 12

[26] Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, Nov 2004. 11

[27] Leonid Gurvits and Howard Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6), Dec 2002. 63

[28] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 63(9):5628–5641, 2017. 10, 11

[29] Aram Harrow, Cedric Yen-Yu Lin, and Ashley Montanaro. Sequential measurements, disturbance and property testing. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1598–1611. SIAM, Philadelphia, PA, 2017. 13, 85

[30] Carl W. Helstrom. *Quantum Detection and Estimation Theory*, volume 123 of *Mathematics in Science and Engineering*. New York, NY: Academic Press, 1976. 27

[31] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 34

[32] Alexander Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* North-Holland Publishing Company, first edition, 1982. 17

[33] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865–942, Jun 2009. 12

[34] 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, Jun 2020. 15

[35] Yu.Ĩ. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003. 11

[36] Richard Kadison. A generalized Schwarz inequality and algebraic invariants for operator algebras. *Annals of Mathematics. Second Series*, 56:494–503, 1952. 20

[37] Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. I*, volume 100 of *Pure and Applied Mathematics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, 1983. Elementary theory. 17

[38] Karl Kraus. *States, effects, and operations*, volume 190 of *Lecture Notes in Physics*. Springer-Verlag, Berlin, 1983. Fundamental notions of quantum theory. 20, 22

[39] Jing Lei. Personal communication, 2022. 35

[40] John E. Maxfield and Henryk Minc. On the matrix equation $X'X = A$. *Proceedings of the Edinburgh Mathematical Society*, 13(02):125, 1962. 58

[41] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. Technical report, arXiv:1310.2035, 2013. 11

[42] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. *Theory of Computing*, 1(1):1–81, 2016. 12

[43] M. A. Naimark. On a representation of additive operator set functions. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 41:359–361, 1943. 32

[44] Ryan O'Donnell and Ramgopal Venkateswaran. The quantum union bound made easy. In *Symposium on Simplicity in Algorithms (SOSA)*, pages 314–320. [Society for Industrial and Applied Mathematics (SIAM)], Philadelphia, PA, 2022. 35, 36

[45] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 529–538. ACM, New York, 2015. 11, 12, 44, 62

[46] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing - STOC 2016*. ACM Press, 2016. 10

[47] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 899–912. ACM, New York, 2016. 11

[48] Ryan O'Donnell and Chirag Wadhwa. Instance-optimal quantum state certification with entangled measurements. 2025. 11

[49] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, Oct 2008. 12, 13

[50] Gert K. Pedersen. *Analysis now*, volume 118 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1989. 17

[51] Yury Polyanskiy and Yihong Wu. *Information theory. From coding to learning*. Cambridge: Cambridge University Press, 2025. 25

[52] Joseph J. Rotman. *An introduction to the theory of groups*, volume 148 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, fourth edition, 1995. 28

[53] Bruce E Sagan. *The symmetric group: representations, combinatorial algorithms, and symmetric functions*. Springer, 2001. 28

[54] Matthew Skala. Hypergeometric tail inequalities: ending the insanity. 2013. Technical report, arXiv:1311.5939. 61

[55] Johann von Neumann. *Mathematische Grundlagen der Quantenmechanik*, volume Band 38 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin-New York, 1968. Unveränderter Nachdruck der ersten Auflage von 1932. 17, 21

[56] Martin J. Wainwright. *High-dimensional statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. A non-asymptotic viewpoint. 65

[57] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, Cambridge, 2018. 32

[58] Adam Bene Watts and John Bostanci. Quantum event learning and gentle random measurements. In *15th Innovations in Theoretical Computer Science Conference*, volume 287 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 97, 22. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2024. 14

[59] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, oct 1982. 9

[60] John Wright. Personal communication, 2016. 15

Index