Coherence in Property Testing: Quantum-Classical Collapses and Separations

Fernando Granha Jeronimo[*] Nir Magrafta[†] Joseph Slote[‡] Pei Wu[§]

October 4, 2024

Working Draft
Please do not distribute

Abstract

Understanding the power and limitations of classical and quantum information, and how they differ, is an important endeavor. On the classical side, property testing of distributions is a fundamental task: a tester, given samples of a distribution over a typically large domain such as $\{0,1\}^n$, is asked to verify properties of the distribution. A key property of interest in this paper is the *support size* both of distributions, a central problem classically [Valiant and Valiant STOC'11], as well, as of quantum states. Classically, even given $2^{n/16}$ samples, no tester can distinguish between distributions of support size $2^{n/8}$ from $2^{n/4}$ with probability better than $2^{-\Theta(n)}$, even with the promise that they are flat distributions.

In the quantum setting, quantum states can be in a coherent superposition of many states of $\{0,1\}^n$, providing a global description of probability distributions. One may ask if coherence can enhance property testing. A natural way to encode a flat distribution is via the *subset states*, $|\phi_S\rangle = 1/\sqrt{|S|}\sum_{i\in S}|i\rangle$. We show that coherence alone is not enough to improve the testability of support size.

1. **Coherence limitations.** Given $2^{n/16}$ copies, no tester can distinguish between subset states of size $2^{n/8}$ from $2^{n/4}$ with probability better than $2^{-\Theta(n)}$.

Our result is more general and establishes the indistinguishability between the subset states and the Haar random states leading to new constructions of pseudorandom and pseudoentangled states, resolving an open problem of [Ji, Liu and Song, CRYPTO'18].

The hardness persists even when allowing multiple public-coin AM provers for a classical tester.

2. **Classical hardness with provers.** Given $2^{O(n)}$ samples from a classical distribution and $2^{O(n)}$ communication with multiple independent AM provers, no classical tester can estimate the support size up to factors $2^{\Omega(n)}$ with probability better than $2^{-\Theta(n)}$. Our hardness result is tight.

[*] University of Illinois Urbana-Champaign. granha@illinois.edu. Supported as a Google Research Fellow.

[†] Weizmann Institute of Science. nir.magrafta@weizmann.ac.il. Supported by the Israel Science Foundation (Grant No. 3426/21), and by the European Union Horizon 2020 Research and Innovation Program via ERC Project REACT (Grant 756482)

[‡] Caltech. jslote@caltech.edu. Supported by Chris Umans Simons Investigator Grant.

[§] Penn State University. pei.wu@psu.edu. Supported by ERC Consolidator Grant VerNisQDevS (101086733).

In contrast, coherent subset state proofs suffice to improve testability exponentially,

3. **Quantum advantage with proofs.** With polynomially many copies and subset state proofs, a tester can approximate the support size of a subset state of arbitrary size.

Some structural assumption on the quantum proofs is required since we show that a general proof cannot (information-theoretically) improve testability of *any* quantum property whatseover. Our results highlight both the power and limitations of coherence in property testing, establishing exponential quantum-classical separations across various parameters. We also show several connections and implications of the study of property testing, in particular, in establishing quantum-to-quantum state transformation lower bounds.

Contents

1 Introduction

Testing whether a given object has a desired property, or is *far* from it, is a fundamental task both in the classical and quantum settings [Gol17, MW16], possessing myriad important applications, e.g., [Din07, DEL$^+$22, PK22]. In this context, understanding the resources (e.g., number of copies of a state or samples, efficiency of tester, etc) needed to test a given property is a central goal of property testing. A key property of interest in this paper is the *support size* of both distributions, a central property classically [Valiant and Valiant STOC'11], as well, as of quantum states [AKKT20].

In testing properties of classical probability distributions, one is given access to a distribution via its samples. In many cases, we are faced with high-dimensional distributions, which can be seen as assigning probabilities to $\{0,1\}^n$. A rich theory of classical property testing of distributions has emerged, and we now know the sample complexity of many properties of interest [R$^+$10, Val11, Rub12, Gol17, Can20, Can22]. There, one quickly learns that several properties require $2^{\Omega(n)}$ samples to be testable, e.g., distinguishing the support size between two families of distributions can require exponentially many samples even if they have vastly different support sizes and are promised to be flat distributions.

Theorem 1.1 (Failure of Classical Testing (Informal))**.** *Even given $2^{n/16}$ copies, no tester can distinguish between flat distributions of size $2^{n/8}$ from $2^{n/4}$ with probability better than $2^{-\Theta(n)}$.*

This kind of strong lower bound is pervasive in property testing of distributions [BFF$^+$01, BDKR02, RRSS07, VV11, VV17, Rub12, HR22], and it establishes severe limitations on our ability to test classical information. Roughly speaking, this is not surprising since by accessing a probability distribution via samples, we do not get a "global" hold on it, but rather, we just get random local pieces of it. In contrast, quantum mechanics allows us to manipulate objects that are global in the sense they are in superposition of possibly many different states of $\{0,1\}^n$. This phenomenon is known as *coherence*, and it is one of the fundamental pillars of quantum mechanics. We can then ask what improvements the setting of quantum information can provide, more specifically, whether this global nature of coherence can lead to substantial improvements in distinguishing vastly different support sizes.

How much can coherence help property testing?

We show that coherence alone cannot help. A quantum analog of a flat probability distribution is a subset state, namely, a quantum state of the form $1/\sqrt{|S|}\sum_{i\in S}|i\rangle$ for some $S \subseteq \{0,1\}^n$. In words, this state is a uniform superposition over some set S. These states are natural in their own right and they are commonly used in quantum complexity [VW16]. More precisely, we prove the following result analogous to the classical case.

Theorem 1.2 (Failure of Testing with Coherence (Informal))**.** *Even given $2^{n/16}$ copies, no tester can distinguish between subset states of size $2^{n/8}$ from $2^{n/4}$ with probability better than $2^{-\Theta(n)}$.*

We obtain the above result from a more general theorem about subset states. In fact, we show that subset states are actually indistinguishable from Haar random states, provided their support is not too small nor too big.

Theorem 1.3. *Let $\mathcal{H} = \mathbb{C}^d$ be a Hilbert space of dimension $d \in \mathbb{N}$, μ be the Haar measure on $\mathcal{H}$, and $S \subseteq [d]$ of size s. Then for any $k \in \mathbb{N}$,*

$$\left\| \int \psi^{\otimes k} d\mu(\psi) - \mathop{\mathbb{E}}_{S \subseteq [d], |S| = s} \phi_S^{\otimes k} \right\|_1 \leq O\left(\frac{k^2}{d} + \frac{k}{\sqrt{s}} + \frac{sk}{d} \right),$$

where $\phi_S = \left(\frac{1}{\sqrt{s}} \sum_{i \in S} |i\rangle \right) \left(\frac{1}{\sqrt{s}} \sum_{i \in S} \langle i| \right)$.

The above theorem leads to a new construction of pseudorandom states (PRS), which is an important primitive with broad applications in quantum cryptography [Kre21, KQST23], resolving an open problem from the seminal work of Ji, Liu and Song [JLS18]. It also leads to a new construction of pseudoentangled states [ABF$^+$24]. At the technical level, the proof Theorem 1.3 goes via spectral graph theory by analyzing some matrices in the so-called Johnson association scheme[1] [Del75].

Given that both classical and quantum property testing models fail spectacularly for our task, one can ask if there are other approaches to property testing.

How to go beyond the standard property testing models?

Very much like NP enhances P (and QMA enhances BQP) with adversarial *proofs*, one can enhance the standard property testing models with proofs (or certificates — "structured" proofs), namely, additional adversarial information intended to help testability. Here, we will consider the power and limitations of proofs and also interaction with provers in the context of property testing. One can imagine that a powerful untrustworthy entity prepares samples (or copies) together with certificates so that a less powerful entity can be convinced of a property, ideally using substantially fewer resources.

Classically, we show that even with exponentially many samples and interacting with exponentially many independent public-coin AM provers for exponentially many rounds, classical property testing still fails,

Theorem 1.4 (Failure of Classical Testing with Certificates (Informal))**.** *Even given $2^{\Omega(n)}$ samples of a classical flat distribution and interaction with $2^{\Omega(n)}$* AM *provers in $2^{\Omega(n)}$ rounds, no classical tester can estimate the support size up to factors $2^{\Omega(n)}$ with probability better than $2^{-\Theta(n)}$.*

At the heart of our proof of the above lower bound is a connection to fast mixing of high-dimensional expanders [AJK$^+$22]. This lower bound technique is quite general and holds even given any promise (say intended to make verification easier) on families of certifying distributions, which, in particular, captures the above public-coin AM lower bound with multitple independent provers with multiple rounds of interactions.

In fact, we show in the classical case, for distinguishing flat distributions of different support sizes, the proof provides no power—an optimal strategy for the honest provers is just to provide more samples by proofs.

Theorem 1.5 (Classical Certification Offer No Advantage (Informal))**.** *Given any public-coin* AM *protocol of communication cost of p bits, let the sample complexity distinguishing flat distribution of support size s and $2s$ with a constant advantage be t. Let t' be the sample complexity without proofs. Then, $p + t \geq \Omega(t')$.*

[1] These matrices also naturally arise in the study of complete high-dimensional expanders.

In sharp contrast to the classical case, the presence of polynomially many flat adversarial certificates (i.e., subset states) can dramatically reduce the number of copies for a property to be testable showing that coherence can also be extremely powerful.

Theorem 1.6 (Effective Quantum Certified Testing (Informal))**.** *With just polynomially many (i.e., $n^{O(1)}$) copies and subset state proofs (i.e., certificates of flat amplitudes), a tester can with high probability either approximate the support size of a subset state of arbitrary size, or detect that the certificates are malicious.*

Note that the above model corresponds to a QMA type tester with "structured" proofs (subset states in this case). It is also natural to ask if the same result can be achieved with a general (adversarial) proof instead of assuming additional structure on the proofs. Surprisingly, the answer is an emphatic no, and this holds for any quantum property whatsoever. More precisely, we can show the following severe information theoretic limitation on property testing with a general proof.

Theorem 1.7 (Informal)**.** *A general (adversarial) proof cannot (information theoretically) improve quantum property testing. More precisely, a general quantum proof can be replaced by at most polynomially many extra input states.*

Theorem 1.7 is obtained using the de-Merlinization ideas of Aaronson [Aar06, HLM17], so we do not claim technical novelty, but rather just make explicit their surprising implication to quantum property testing with a general proof. We point out that this result is (essentially) an information theoretic result since the above process of replacing a proof can incur an exponential increase in the running of a tester.

We also show how the study of property testing with structured proofs, sometimes even with very strong promises, can have interesting consequences to quantum-to-quantum state transformation lower bounds. One example of a quantum-to-quantum state transformation lower bound that can deduced from our work is the following.

Theorem 1.8 (Hardness of Absolute Amplitudes Transformation (Informal))**.** *Any transformation that takes k copies of an arbitrary n-qubit quantum state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle$ and produces a single n-qubit output state at least 0.001 close to $\sum_{x\in\{0,1\}^n} |\alpha_x| \, |x\rangle$ requires $k = 2^{\Omega(n)}$.*

Curiously, these quantum-to-quantum lower bounds can be deduced from the study of quantum-to-classical results.

This paper is organized as follows. In Section 1.1, we survey related work on quantum and classical property testing. In Section 1.2, we provide more details on the connection of our results with pseudorandomness and pseudoentanglement. In Section 2, we recall basic notation and terminology, and introduce some of the property testing models studied in this work. In Section 4, we give an overview on some of our technical results explaining, in particular, the role of the spectral analysis in the Johnson scheme and the use of fast mixing of high-dimensional expanders mentioned above. In Section 5, we establish the limitations of coherence, proving Theorem 1.2 and Theorem 1.3. We also provide additional results on indistinguishability of ensemble of quantum states that is not via Haar random states. In Section 6, we show the limitations of testing classical distributions leading to Theorem 1.4 and Theorem 1.5. In Section 7, we show that a general proof cannot improve quantum property testing leading to Theorem 1.7. In Section 8, we show that subset state proofs can substantially improve testability leading to Theorem 1.6. In Section 9, we conclude with a description of how quantum property testing relates to standard computational complexity, and to problems with quantum input.

1.1 Related Work on Property Testing

There is a vast body of work studying property testing both in the classical [Gol17] and quantum settings [MW16]. In the quantum setting, property testing of oracles has been more extensively investigated. There is by now a diverse and powerful set of query lower bound techniques which includes the polynomial method [BBC+01], adversary method [Amb00], generalized adversary method [HLS07], and others [AA18, AKKT20, ABK+21]. In the area of property testing of oracles, the analogous problem to testing support size is known as *approximate counting*. Approximately counting the weight of a classical oracle with a quantum QMA proof was considered by Aaronson et al. in [AKKT20]. They focoused on distinguising oracle weight w from $2w$ and established strong lower bounds using Laurent polynomials. They also considered the setting without proofs, but with access to subset states encoding the support of the classical oracle or access to a unitary that can produce it, also obtaining lower bounds. Subsequently, Belovs and Rosmanis [BR20] established lower bounds for approximate counting in the setting without proofs in the regime w versus $(1+\varepsilon)w$ for small $\varepsilon \in (0,1)$. In [DGRMT22], Dall'Agnol et al. investigate the power of adversarial quantum certificates in the study of property testing of unitaries. More recently, Weggemans [Weg24] showed additional lower bounds for testing some unitary properties with certificates and advice. The literature on property testing of quantum states rather than oracles (or unitaries) seems to be much sparser.

Property testing of classical probability distributions has been extensively studied [R+10, Val11, Rub12, Can20, Can22]. The case of testability of support size of distributions, or closely related notions such as entropy, uniformity, essential support, etc, are very natural, and they have been investigated in many works, including for flat distributions. Valiant and Valiant's $\Theta(N/\log N)$ bounds on testing support size of general distribution of domain size N is widely considered as a cornerstone of the area [VV11].

The notion of proofs (or certificates) is pervasive in theoretical computer science, and it was also studied in many forms in property testing. In [CG18], Chiesa and Gur investigate the power of adversarial certificates for property testing of probabilities distributions. They considered analgues of NP and MA, where certificates are bit strings. They also considered analogues of single prover interactive proofs AM and IP. They show that their corresponding NP, MA, and AM models can at most provide a quadratic advantage in general, whereas IP can give an exponential improvement. Following Chiesa and Gur's work, there are several works focusing on the upper bounds on property testing of general distribution including support size for various interactive proof models [HR22, HR23]. Here in our classical lower bounds, we consider multiple provers, whose certificates are probability distributions satisfying any desired (convex) promise intended to help testability (see Section 3 for the precise details). In particular, we consider interaction with multiple independent AM provers. Our technique based on fast mixing of high-dimensional expanders yields tight lower bounds.

1.2 Quantum Pseudorandomness from Our Results

We now explain the implications of Theorem 1.3 above for quantum pseudorandomness and pseudoentanglement. We recall some of the context about these concepts along the way.

Pseudorandom States. Pseudorandom quantum states (PRS) are a keyed family of quantum states that can be efficiently generated and are computationally indistinguishable from Haar random states, even when provided with polynomially many copies. PRSs have a wide range of applications including but not limited to statistically binding quantum bit commitments [MY22b] and private quantum coins [JLS18]. Notably, for certain applications like private quantum coins, PRSs represent the weakest primitive known to imply them. Moreover, PRSs imply other quantum pseudorandom objects, such as one-way state generators (OWSGs) [JLS18, MY22a] and EFI pairs (**e**fficiently samplable, statistically **f**ar but computationally **i**ndistinguishable pairs of quantum states) [BCQ23, MY22b]. Although all the existing PRS constructions rely on quantum-secure pseudorandom functions (PRFs) or pseudorandom permutations (PRPs), PRSs may be weaker than PRFs [KQST23].

Since the initial proposal of pseudorandom states [JLS18], various constructions have been investigated [BS19, AGQY22, BS20, ABF$^+$24, BBSS23]. Randomizing the phase was essential in the security proofs for all of these constructions. It is then natural to ask if it is possible to construct PRS without varying the phases, and indeed, Ji, Liu, and Song raised this question and conjectured that PRS can be constructed using *subset states.*

Consider a n-qubit system, represented by a 2^n dimensional Hilbert space. For any function $t(n) = \omega(\mathrm{poly}(n))$ and $t(n) \leq s \leq 2^n/t(n)$ and $k = \mathrm{poly}(n)$, the distance between k copies of a Haar random state and k copies of a random subset state of size s is negligible as per the above theorem. This range for the subset's size is tight, as otherwise efficient distinguishers exist between copies of a Haar random state and a random subset state.

An immediate corollary of Theorem 1.3 above is the following:

Corollary 1.9 (Pseudorandom States)**.** *Let* $\{\mathrm{PRP}_k : [2^n] \to [2^n]\}_{k\in\mathcal{K}}$ *be a quantum-secure family of pseudorandom permutations. Then the family of states* $\left\{\frac{1}{\sqrt{s}}\sum_{x\in[s]}|\mathrm{PRP}_k(x)\rangle\right\}_{k\in\mathcal{K}}$ *is a PRS on* n *qubits for* $t(n) \leq s \leq 2^n/t(n)$ *and any* $t(n) = \omega(\mathrm{poly}(n))$*.*

Tudor and Bouland [GTB23] indepdentently discovered a subset state PRS construction. Their analysis uses representation theory.

Pseudoentanglement. A closely related notion to the PRSs is that of *pseudoentangled* states studied recently by [ABF$^+$24]. Here we call a PRS $h(n)$-pseudoentangled if for any state $|\phi\rangle$ from the PRS, $|\phi\rangle$ in addition satisfies that its entanglement entropy across all cut is $O(h(n))$.[2] Note that for a Haar random state, the entanglement entropy is near maximal across all cuts. It's observed in [ABF$^+$24] that a subset state with respect to set S has entanglement entropy at most $O(\log|S|)$ across any cut for some function $h(n) : \mathbb{N} \to \mathbb{N}$, since the Schmidt rank of a subset state is at most $|S|$ across any cut. Therefore the subset states with small set size are good candidates for pseudoentangled states, which they left as an open problem.

[2] In [ABF$^+$24], another notion was considered. Roughly speaking, a pseudoentangled state ensemble consists of two efficiently constructible and computationally indistinguishable ensembles of states which display a gap in their entanglement entropy across all cuts.

Since a subset state with respect to a small subset has small entanglement entropy across all cut, we also have

Corollary 1.10 (Pseudoentangled States)**.** *Let* $\mathrm{PRP}_k : [2^n] \to [2^n]$ *be a quantum-secure family of pseudorandom permutations. For any* $h(n) = \omega(\log n)$ *and* $h(n) = n - \omega(\log n)$*, we have the following* $h(n)$*-pseudoentangled state from subset state of size* $s = 2^{h(n)}$*,*

$$\left\{ \frac{1}{\sqrt{s}} \sum_{x \in [s]} |\mathrm{PRP}_k(x)\rangle \right\}_{k \in \mathcal{K}} .$$

For a PRS, it is easy to see that if for some cut the entanglement entropy of a state $|\phi\rangle$ is $O(\log n)$, then $|\phi\rangle$ can be distinguished from Haar random states with polynomially many copies using swap test.

2 Preliminaries

General. We adopt the Dirac notation for vectors representing quantum states, e.g., $|\psi\rangle, |\phi\rangle$, etc. All the vectors of the form $|\psi\rangle$ will be unit vectors. Given any pure state $|\psi\rangle$, we adopt the convention that its density operator is denoted by the Greek letter without the "ket", e.g. $\psi = |\psi\rangle\langle\psi|$. The set of density operators in an arbitrary Hilbert space $\mathcal{H}$ is denoted $\mathfrak{D}(\mathcal{H})$, and the set of pure states is denoted by $\mathfrak{S}(\mathcal{H})$. For a mixed state denoted by capital letters, e.g., Ψ, Φ, we quite often treat it as a *set* of states together with some underlying distribution on the set. For mixed state Ψ, there can be many different ways to express it as a distribution on pure states. Normally, we fix some explicit set that will be clear from the context. So Ψ, Φ will have both the set interpretation and the density matrix interpretation. The Haar measure is referred to the uniform measure on the unit sphere of $\mathbb{C}^d$.

Given any matrix $M \in \mathbb{C}^{n \times n}$ denote by $\|M\|_1$ the trace norm, which is the sum of the singular values of M. We write $x \lesssim y$ to denote that there is a small constant $c \geq 1$, such that $x \leq cy$. For any set S, let $A(S, k) := \{(i_1, i_2, \ldots, i_k) \in S^k : i_j \neq i_{j'} \text{ for } j \neq j'\}$. We also adopt the notation S_n for the symmetric group. For two disjoint sets A, B, we use $A \sqcup B$ to denote the their union, emphasizing that A and B are disjoint.

Let $n^{\underline{k}} := n(n-1)\cdots(n-k+1)$. A simple calculation reveals that for $k = O(\sqrt{n})$,

$$\frac{n^{\underline{k}}}{(n-k+1)^{\underline{k}}} - 1 \leq \left(\frac{n+k-1}{n-k+1}\right)^k - 1 = O\left(\frac{k^2}{n}\right).$$

We will use this bound without referring to this calculation. Given any pure quantum state ρ over systems A, B, the entanglement entropy over the cut $A : B$ is the von Neumann entropy of the reduced density matrix of system A (or B), i.e., $-\operatorname{Tr}(\rho_A \log \rho_A)$.

Entropy and KL-divergence. Consider some discrete space Ω and a probability measure γ over Ω. If the random variable X is drawn from γ, we denote it by $X \sim \gamma$. We let $\ln x$ and $\log x$ stand for the natural logarithm of x and the logarithm of x to base 2, respectively. For any distribution γ over some discrete space Ω, the entropy function

$$H(\gamma) = \mathop{\mathbb{E}}_{x \in \Omega} \gamma(x) \log \frac{1}{\gamma(x)}.$$

Recall that the Kullback-Leibler divergence (KL-divergence) between two distributions μ_0, μ_1 over Ω is defined by the following formula

$$\mathrm{KL}(\mu_0 \,\|\, \mu_1) = \sum_{x \in \Omega} \mu_0(x) \log \frac{\mu_0(x)}{\mu_1(x)}.$$

If two random variables X_0, X_1 obey μ_0 and μ_1, respectively, we also use $\mathrm{KL}(X_0 \,\|\, X_1)$ to denote the KL-divergence between the two distributions. The KL-divergence satisfies the following chain rule:

$$\mathrm{KL}(X_0 Y_0 \,\|\, X_1 Y_1) = \mathrm{KL}(X_0 \,\|\, X_1) + \mathop{\mathbb{E}}_{x \sim X_0}\left[\mathrm{KL}\left(\frac{Y_0 \mid X_0 = x}{Y_1 \mid X_1 = x}\right)\right].$$ [3]

Flatness. A distribution μ is called *flat* if it is uniform over its support. There is one natural way to encode a classical distribution over a discrete domain X into a pure quantum state over the Hilbert space $\mathbb{C}^X$. That corresponds to the so-called *subset state.*

Definition 2.1 (Subset States/Flat States)**.** *We say that $|\psi\rangle \in \mathbb{C}^d$ is a subset state (or, equivalently, a flat state) if $|\psi\rangle$ is the uniform superposition over some subset $S \subseteq [d]$,*

$$|\psi\rangle = \frac{1}{\sqrt{|S|}} \sum_{i \in S} |i\rangle.$$

3 Property Testing Models

We now formally state some definitions and models of property testing we will use. Note that the definitions are made focusing on sample complexity. Some remarks on time complexity will be discussed in the final section. We start with the classical setting. Let Δ_d be the probability simplex in $\mathbb{R}^d$, i.e.,

$$\Delta_d := \left\{ (p_1, \dots, p_d) \in \mathbb{R}^d : \sum_{i=1}^{d} p_i = 1 \text{ and } p_1, \dots, p_d \geq 0 \right\}.$$

Analogously, for a finite set S, we denote by Δ_S the probability simplex in $\mathbb{R}^S$. Recall that a property of classical probability distributions is defined as follows.

Definition 3.1 (Property of Classical Distributions)**.** *A property is any family of probability distributions $\mathcal{P} = \sqcup_d \mathcal{P}_d$ where $\mathcal{P}_d \subseteq \Delta_d$.*

Definition 3.2 (Standard Classical Property Testing Model)**.** *For $d \in \mathbb{N}$, let $k = k(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropBPP}(k, a, b)$ *if there exists a verifier V such that for every $\nu \in \Delta_d$,*

(i) *if $\nu \in \mathcal{P}_d$, then V with k independent samples from ν accepts with probability at least a, and*

[3] Here, we use the fraction-like notation to also denote the KL-divergence for aesthetics, as we are comparing two conditional distributions. The numerator in the fraction-like notation corresponds to the first argument in the standard notation.

(ii) *if $\nu \in \Delta_d$ is ε-far in statistical distance from $\mathcal{P}_d$, then V with k independent samples from ν accepts with probability at most b.*

We proceed to discuss the classical property testing model enhanced with classical certificates. In the most standard notion of the Merlin-Arthur (MA) type proof, fix some property $\mathcal{P}$, one expect that for any $\nu \in \mathcal{P}_d$, there is an honest proof $\pi \in \{0,1\}^p$ that convince the verifier to accept with high probability after making k samples. On the other hand, for $\nu \in \Delta_d$ far from $\mathcal{P}$, no proof should fool the verifier to accept with high probability.

Definition 3.3 (Classical Property Testing with MA Proofs)**.** *For $d \in \mathbb{N}$, let $k = k(d), p = p(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropMA}(k,p,a,b)$ *with respect to certificates set $\mathcal{S} = \{0,1\}^p$ if there exists a verifier V such that for every $\nu \in \Delta_d$,*

(i) *if $\nu \in \mathcal{P}_d$, then there exist $\pi \in \mathcal{S}$ such that*

$$\Pr_{x_1,\ldots,x_k \sim \nu^{\otimes k}} [V(x_1,\ldots,x_k,\pi) \text{ accepts}] \geq a\,,$$

(ii) *if ν is ε-far from $\mathcal{P}_d$ in statistical distance, then for every $\pi \in \mathcal{S}$,*

$$\Pr_{x_1,\ldots,x_k \sim \nu^{\otimes k}} [V(x_1,\ldots,x_k,\pi) \text{ accepts}] \leq b\,.$$

A much stronger notion usually referred to as the public-coin Arthur-Merlin (AM) model, in its great generality, involves m provers, r rounds, and in each round, Arthur sends m independently uniformly random bit strings of length p to each of m Merlins who has no information about each other's communication, and Merlin responds with a 1 bit. After the r rounds of communication, Arthur should be able to decide whether the unknown distribution $\nu \in \mathcal{P}$ or far from it. Any such AM protocol Π gives rise to some distribution on the transcript $\pi \in \{0,1\}^{rm(p+1)}$ satisfying that in each round, the random bits send are uniformly random; and the communications with each Merlin are independent of each and only depends on communication history between the current Merlin and Arthur. Fix any valid AM protocol Π, we can let $\mu_\Pi \subseteq \Delta(\{0,1\}^{rm(p+1)})$ be the distribution on the communication transcript generated by Π.

Definition 3.4 (Classical Property Testing with AM Proofs)**.** *For $d \in \mathbb{N}$, let $k = k(d), m = m(r), p = p(d), r = r(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropAM}[r](k,m,r(mp+m),a,b)$,

(i) *if $\nu \in \mathcal{P}_d$, there exists a r-round, m-prover,* AM*-protocol* Π*, where in each round Arthur can send p uniformly random bits and get 1 bit answer, such that*

$$\Pr_{x_1,\ldots,x_k \sim \nu^{\otimes k}, \pi \sim \mu_\Pi} [V(x_1,\ldots,x_k,\pi) \text{ accepts}] \geq a\,,$$

(ii) *if ν is ε-far from $\mathcal{P}_d$ in statistical distance, for any r-round, m-prover,* AM*-protocol* Π*, where in each round Arthur can send p uniformly random bits and get 1 bit answer,*

$$\Pr_{x_1,\ldots,x_k \sim \nu^{\otimes k}, \pi \sim \mu_\Pi} [V(x_1,\ldots,x_k,\pi) \text{ accepts}] \leq b\,.$$

Analogously, we can define $\mathrm{PropIP}(k,p,a,b)$ which is like $\mathrm{PropAM}[\mathrm{poly}(n)](k,1,p,a,b)$, but now the verifier has private random coins. Consequently, in each round, the verifier's message can depend on the private coin and the communication transcript so far.

We now move to the quantum setting. Recall that $\mathfrak{S}(\mathbb{C}^d)$ denotes the set of pure states in $\mathbb{C}^d$. First, we review the notion of property of quantum states.

Definition 3.5 (Property of Quantum States)**.** *A property is any family of subsets $\mathcal{P} = \sqcup_d \mathcal{P}_d$ where $\mathcal{P}_d \subseteq \mathfrak{S}(\mathbb{C}^d)$.*

The standard quantum property testing model is defined as follows.

Definition 3.6 (Standard Quantum Property Testing Model)**.** *For $d \in \mathbb{N}$, let $k = k(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropBQP}(k, a, b)$ *if there exists a verifier V such that for every $|\psi\rangle \in \mathbb{C}^d$,*

(i) *if $|\psi\rangle \in \mathcal{P}_d$, then $V(|\psi\rangle^{\otimes k})$ accepts with probability at least a, and*
(ii) *if $|\psi\rangle$ is ε-far from $\mathcal{P}_d$ in trace distance, then $V(|\psi\rangle^{\otimes k})$ accepts with probability at most b.*

This model can be enhanced with certificates as follows.

Definition 3.7 (Quantum Property Testing with Certificates)**.** *For $d \in \mathbb{N}$, let $k = k(d), p = p(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropQMA}(k, m, mp, a, b)$ *if there exists a verifier V such that for every $|\psi\rangle \in \mathbb{C}^d$,*

(i) *if $|\psi\rangle \in \mathcal{P}_d$, then there exist m certificates $|\phi_1\rangle, \ldots, |\phi_m\rangle \in \mathbb{C}^{2^p}$ such that $V(|\psi\rangle^{\otimes k} \otimes |\phi_1\rangle \otimes \cdots \otimes |\phi_m\rangle)$ accepts with probability at least a, and*
(ii) *if $|\psi\rangle$ is ε-far from $\mathcal{P}_d$ (in trace distance), then then for every $|\phi_1\rangle, \ldots, |\phi_m\rangle \in \mathbb{C}^{2^p}$, $V(|\psi\rangle^{\otimes k} \otimes |\phi_1\rangle \otimes \cdots \otimes |\phi_m\rangle)$ accepts with probability at most b.*

We will also consider the promise version of these models in which we have a pair of disjoint properties $(\mathcal{P}_{\mathrm{YES}}, \mathcal{P}_{\mathrm{NO}})$. Then $\mathcal{P}_{\mathrm{YES}}$ will be the property of interest, and the item (ii) in the above definitions will only care about distributions or states from $\mathcal{P}_{\mathrm{NO}}$, a subset of distributions or states ε-far from $\mathcal{P}_{\mathrm{YES}}$ in this paper.

4 Technical Overview of Our Results

Our results intersect many fields: classical and quantum property testing; the power and limitations of classical and quantum proofs; the nature of classical versus quantum information via the role of coherence; cryptography via PRSs and entanglement via pseudoentanglement. At the heart of some of our technical results are connections to spectral graph theory via the Johnson scheme and also to fast mixing of high-dimensional expanders. We now give an overview of some of these connections, and precise technical details are left to the relevant sections.

4.1 Johnson Scheme and Limitations of Coherence

We will now discuss how subset states of the form $|\phi_S\rangle$, devoided of negative and imaginary phases by definition, can be indistinguishable from Haar random states over the sphere $\mathbb{C}^d$. This indistinguishability happens whenever the subset size $|S|$ is not too small nor too large. In turn, this implies that subset states of vast different support sizes are indistinguishable since they are both indistinguishable from Haar random states.

The key technical contribution is realizing a connection to the so-called Johnson scheme and conducting a spectral analysis using it to obtain the above indistinguishability. Recall that the matrices of the Johnson scheme $\mathcal{J}([d], k)$ have rows and columns indexed by sets from $\binom{[d]}{k}$. Moreover, given a matrix $\mathcal{D}$ on the scheme, each entry $\mathcal{D}(A,B)$ only depends on the size of the intersection $A \cap B$. For $t \in \{0,1,\ldots,k\}$, one defines a basis matrix $\mathcal{D}_t$, whose rows and columns are indexed by elements in $\binom{[d]}{k}$ as follows

$$\mathcal{D}_t(A,B) = \begin{cases} 1 & \text{if } |A \cap B| = t \\ 0 & \text{otherwise} \end{cases}$$

for every $A, B \in \binom{[d]}{k}$. It follows that any matrix $\mathcal{D}$ in the Johnson scheme can be decomposed as $\mathcal{D} = \sum_{t=0}^{d} \alpha_t \mathcal{D}_t$, where each α_t is a scalar. This association scheme has a variety of remarkable properties,[4] but we will be mostly concerned with the spectral properties of $\mathcal{D}_t$.

Now, we proceed to give an idea of how the Johnson scheme arises in the analysis. The starting point is the well-known fact that the expectation over Haar random states $\int \psi^{\otimes k} d\mu$ is equal to[5]

$$\binom{d+k-1}{k}^{-1} \frac{1}{k!} \sum_{\pi \in S_k} \sum_{\vec{i} \in [d]^k} |\vec{i}\rangle\langle \pi(\vec{i})| \approx \binom{d+k-1}{k}^{-1} \frac{1}{k!} \sum_{\pi \in S_k} \sum_{\vec{i} \in A([d],k)} |\vec{i}\rangle\langle \pi(\vec{i})| =: \tilde{\Psi},$$

where the last approximation assumes $k \ll \sqrt{d}$. Although the matrix on the RHS is not in the Johnson scheme, note that its only non-zero entries have a fixed value and occur on entries indexed by row $(i_1,\ldots,i_k)$ and column $(j_1,\ldots,j_k)$ if and only if $|\{i_1,\ldots,i_k\} \cap \{j_1,\ldots,j_k\}| = k$. This means that this matrix can be written as $\mathcal{D}_k \otimes J$, where J is a $k! \times k!$ all-ones matrix.

Next, we turn our attention to uniform average of subset states of a fixed size s. As before, it will be convenient to work with tuples of distinct indices as

$$\tilde{\Phi} = \mathop{\mathbb{E}}_{S:|S|=s} \left[\frac{1}{s^{\underline{k}}} \sum_{\vec{i},\vec{j} \in A(S,k),} |\vec{i}\rangle\langle \vec{j}| \right],$$

and the approximation error is small provided $k \ll \sqrt{s}$. Note that

$$\tilde{\Phi}((i_1,\ldots,i_k),(j_1,\ldots,j_k)) = \frac{1}{s^{\underline{k}}} \Pr_{|S|=s}[\vec{i},\vec{j} \in A(S,k)] = \frac{1}{s^{\underline{k}}} \frac{\binom{d-2k+t}{s-2k+t}}{\binom{d}{s}}, \tag{4.1}$$

where $t = |\{i_1,\ldots,i_k\} \cap \{j_1,\ldots,j_k\}|$. Since s and k are fixed, this means that the entries only depend on the size of the intersection of the set of elements in the tuples, and thus $\tilde{\Phi} = \sum_{t=0}^{k} \alpha_t \mathcal{D}_t \otimes J$, for scalars α_t. To compute the trace distance between $\tilde{\Psi}$ and $\tilde{\Phi}$ in order to conclude that these states are close, we rely on the spectral properties of the Johnson scheme. We also show the indistinguishability of some ensembles with dense support (see Section 5.4), but this time, it is not via indistinguishability from Haar.

4.2 Fast Mixing of High-dimensional Expanders and Classical Limitations

Suppose our goal is to distinguish flat distributions of support size s, the yes-case, from those with support size $w \gg s$, the no-case. Suppose these distributions are on $[N]$, where

[4] It forms a commutative algebra of matrices under addition and matrix multiplication.

[5] A normalized version of the projector onto the symmetric subspace.

$N = 2^n$. We can imagine that we have a complete simplicial complex $X = \cup_{i=1}^{w} X(i)$, with $X(i) = \binom{[N]}{i}$. Taking t independent samples from a flat distribution with support $S \in X(s)$ is approximately the same as taking a uniform subset of size t from S provided $t \ll \sqrt{s}$. In the yes-case, each flat distribution, represented by a set S, has an associated certifying distribution π_S. We can think that we choose a set $S \in X(s)$ uniformly at random with its corresponding certificate. This naturally gives rise to a pair of coupled random variables $(\mathcal{S} = S, \Pi = \pi_S)$. Now sampling from Π induces a conditional distribution on $\mathcal{S}|\Pi$; equivalently, we have a distribution μ on $X(s)$. This possibly very sparsely-supported distribution corresponds to how much we learned from the proof. Now, we take a uniformly random subset T of size t from $\mathcal{S}$. The connection to fast mixing of high-dimensional expanders now emerges. We recall that the complete simplicial complex X is a very strong HDX. A well known random walk on X is the Down walk $D_{i \to i-1}$, defined for every $i \in \{2, \ldots, w\}$, as walk operator from $X(i)$ to $X(i-1)$

$$D_{i \to i-1}(A, B) = \begin{cases} \frac{1}{i} & \text{if } A \subseteq B \\ 0 & \text{otherwise} \end{cases}$$

In this language, observe that T is distributed as $\mu D_{s \to s-1} D_{s-1 \to s-2} \ldots D_{t+1 \to t}$. The closeness of T to uniform on $X(t)$ is given by how fast the down random walk mixes starting with distribution μ on $X(s)$. Roughly speaking, since sampling T as above is statistically similar to sampling a uniform set from $X(t)$, this means that the certificate was not very informative, and this will allow us to deduce lower bounds on distinguishing the yes, and no cases. We illustrate this fast mixing from μ on $X(s)$ to close to uniform on $X(t)$ in Fig. 1.

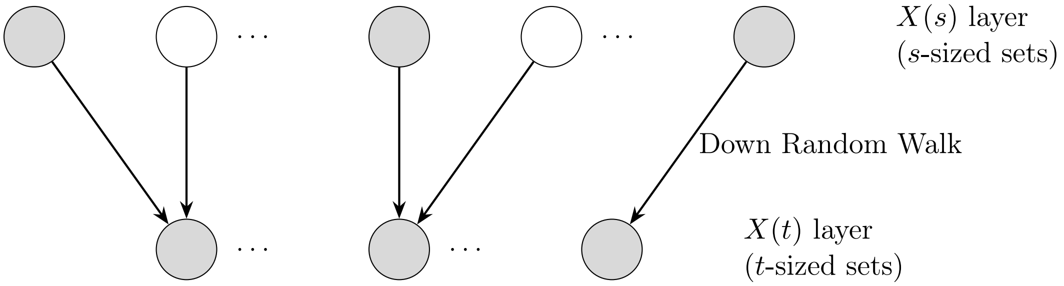


Figure 1: Mixing to uniform measure on $X(t)$ with Down random walk starting from $X(s)$. Gray vertices on top indicate support of initial measure on $X(s)$, whereas gray vertices on the bottom indicate the support on $X(t)$.

It is easy to see that if μ was just a delta distribution on a single set S, then mixing cannot happen. Therefore, we need to also ensure that sampling Π leaves us with enough entropy on μ so that mixing happens. In trying to help the verification protocol as much as possible, we can assume that the certifying distributions only come from a desired promised convex set, and this proof technique is oblivious to this choice. In particular, we can consider that the certifying distributions come from multiple AM provers and the lower bound still applies.

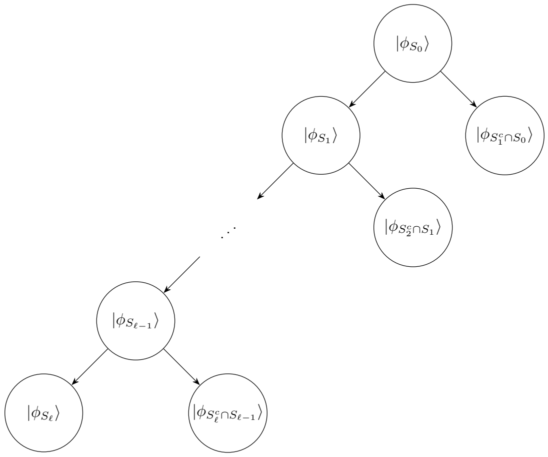
4.3 Coherence Strikes Back via Certificates

As discussed above, coherence alone is not enough to imply distinguishability between subset states of vastly different support size. In contrast, coherence in the form of additional (adversarial) certifying subset states will enable us to give a multiplicative approximation to the support size of an arbitrary subset state $|\psi_S\rangle$ regardless of the size of $S \subseteq \{0,1\}^n$.

An honest prover will consider an arbitrary sequence of nested sets such that $\{0,1\}^n = S_0 \supseteq S_1 \supseteq S_2 \supseteq \cdots \supseteq S_\ell = S$ and $|S_i| \,/\, |S_{i-1}| = 1/2$ for every $i \in [\ell]$ (we assumed that $|S|$ is a power of two for convenience). For each S_i, the prover will send multiple copies of $|\phi_{S_i}\rangle$ and $|\phi_{S_i^c \cap S_{i-1}}\rangle$. We show that with multiple purported copies of $|\phi_{S_{i-1}}\rangle$, $|\phi_{S_i}\rangle$, $|\phi_{S_i^c \cap S_{i-1}}\rangle$ it is possible to test if

$$\frac{|S_i|}{|S_{i-1}|} \;=\; \frac{1}{2} \pm \delta\,,$$

or reject if the prover is dishonest. We illustrate the expected nested sequence of subset states below.


$|\phi_{S_0}\rangle$
$|\phi_{S_1}\rangle$
$|\phi_{S_1^c \cap S_0}\rangle$
$|\phi_{S_2^c \cap S_1}\rangle$
$|\phi_{S_{\ell-1}}\rangle$
$|\phi_{S_\ell}\rangle$
$|\phi_{S_\ell^c \cap S_{\ell-1}}\rangle$

Coherence allows us to recursively apply this with enough control on the errors so that we obtain the telescoping conclusion

$$|S| \;=\; |S_0| \,\frac{|S_1|}{|S_0|}\frac{|S_2|}{|S_1|} \cdots \frac{|S_\ell|}{|S_{\ell-1}|} \;=\; \left(\frac{1}{2} \pm \delta\right)^{\ell} 2^n\,.$$

Since t is at most n, we obtain a non-trivial multiplicative approximation to $|S|$ with $\delta = 1/\mathrm{poly}(n)$.

5 Testing without Certificates: A Fiasco

In this section, we illustrate some natural examples of untestable quantum properties.

5.1 The Lower Bound Meta-Technique

To start, we review the generic lower bound technique on the power of a tester in distinguishing two collections of quantum states $\mathcal{A}$ and $\mathcal{B}$, representing states with and without certain abstract properties, respectively. If there is a tester that distinguishes between any state in $\mathcal{A}$ from any state in $\mathcal{B}$, then by linearity, it should distinguish any distributions over states from $\mathcal{A}$ and $\mathcal{B}$. In other words, it implies the distinguishability of ensembles. By the contrapositive, if the distinguishability of ensembles fails for any pair of ensembles, this means that there is no tester distinguishing $\mathcal{A}$ and $\mathcal{B}$. This meta-technique for indistinguishability is standard in the quantum as well as the classical[6] setting. The key innovations are the ensembles for which we show indistinguishability results.

Lemma 5.1 (Pointwise Distinguishability Implies Ensemble Distinguishability)**.** *Let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{C}^d$ and $1 \geq a > b \geq 0$. If there exists a measurement M such that*

$$\forall |\phi\rangle \in \mathcal{A}, \;\; \mathrm{Tr}(M\phi^{\otimes k}) \geq a, \;\; \textit{and,}$$
$$\forall |\phi\rangle \in \mathcal{B}, \;\; \mathrm{Tr}(M\phi^{\otimes k}) \leq b.$$

Then for any distributions $\mu_{\mathcal{A}}, \mu_{\mathcal{B}}$ on $\mathcal{A}$ and $\mathcal{B}$, respectively, we have

$$\mathrm{Tr}(M\rho_{\mathcal{A}}) \geq a \quad \text{and} \quad \mathrm{Tr}(M\rho_{\mathcal{B}}) \leq b\,,$$

where $\rho_{\mathcal{A}} = \mathbb{E}_{\mu_{\mathcal{A}}}\phi^{\otimes k}$ and $\rho_{\mathcal{B}} = \mathbb{E}_{\mu_{\mathcal{B}}}\phi^{\otimes k}$.

Proof. By linearity, we have

$$\mathrm{Tr}(M\rho_{\mathcal{A}}) = \mathbb{E}_{\mu_{\mathcal{A}}} \mathrm{Tr}(M\phi^{\otimes k}) \geq a,$$

where the last inequality follows from the assumption. An analogous computation establishes that $\mathrm{Tr}(M\rho_{\mathcal{B}}) \leq b$, concluding the proof. □

By the contrapositive of Lemma 5.1, one obtains the following lemma asserting that it suffices to find indistinguishable ensembles to rule out the existence of a property tester.

Lemma 5.2 (Ensemble Indistinguishability Implies Pointwise Indistinguishability)**.** *Let $\mathcal{A}, \mathcal{B} \subseteq \mathbb{C}^d$. If there exist distributions $\mu_{\mathcal{A}}, \mu_{\mathcal{B}}$ on $\mathcal{A}$ and $\mathcal{B}$, respectively, such that*

$$\left\|\rho_{\mathcal{A}} - \rho_{\mathcal{B}}\right\|_1 < \varepsilon,$$

where $\rho_{\mathcal{A}} = \mathbb{E}_{\mu_{\mathcal{A}}}\phi^{\otimes k}$ and $\rho_{\mathcal{B}} = \mathbb{E}_{\mu_{\mathcal{B}}}\phi^{\otimes k}$. Then there is no measurement M satisfying

$$\mathrm{Tr}(M\phi^{\otimes k}) \geq a, \forall |\phi\rangle \in \mathcal{A}, \quad \text{and} \quad \mathrm{Tr}(M\phi^{\otimes k}) \leq b, \forall |\phi\rangle \in \mathcal{B},$$

with $b - a \geq \varepsilon$.

5.2 Warm-ups: Product States v.s. Nonproduct States

Here, we show that testing *productness*, i.e., given a state $|\psi\rangle$ determine if it's a (multipartite) product state or $(1-\varepsilon)$-far from being a product state, is impossible. Based on the previous discussion, to rule out a property tester, one needs to come up with indistinguishable ensembles. For example, consider the following two ensembles:

$$\mathcal{E}_1^k = \{|\psi\rangle^{\otimes k} : |\psi\rangle \in \mathbb{C}^d, \langle \psi \,|\, \psi \rangle = 1\};$$
$$\mathcal{E}_0^k = \Big\{ \frac{1}{\sqrt{k!}} \sum_{\pi \in S_k} |\psi_{\pi(1)}\rangle \cdots |\psi_{\pi(k)}\rangle : |\psi_1\rangle, \ldots, |\psi_k\rangle$$
$$\text{are the first } k \text{ columns of a Haar random unitary } U \in \mathbb{C}^{d \times d} \Big\}.$$

[6] Suitably stated for probability distributions in the classical setting.

Note that states from the first ensemble are k-partite product, while states from the second ensemble have small overlap with any product state. We consider the Haar measure on states for $\mathcal{E}_1^k$ and the Haar measure on unitaries for $\mathcal{E}_0^k$. Note that $\mathbb{E}_{\psi\in\mathcal{E}_1^k}\psi$ and $\mathbb{E}_{\psi\in\mathcal{E}_0^k}\psi$ are both invariant under the action of $U^{\otimes k}$ for any unitary U since the corresponding Haar measures are invariant. By Schur's lemma from representation theory [S+77], we have

$$\mathop{\mathbb{E}}_{\psi\in\mathcal{E}_1^k}\psi = \mathop{\mathbb{E}}_{\psi\in\mathcal{E}_0^k}\psi.$$

By Lemma 5.2, no tester has any advantage testing product states and those far from being product using a single copy. We remark that it is known that insisting on perfect completeness means accepting everything because the product states span the entire space. This example rules out any other possible test giving up perfect completeness.

5.3 Subset States v.s. Haar Random States

Testing productness is impossible with one copy, but is testable with two copies of the state [HM13]. In this section, we demonstrate a property impossible to test even with polynomially many copies, that is subset states of fixed size. In particular, we show that testing the size s of the subset is impossible by considering the naive ensembles for different support size $s,\ell\in[N]$:

$$\mathcal{E}_1 = \left\{\phi_S = \frac{1}{\sqrt{|S|}}\sum_{i\in S}|i\rangle \;:\; |S| = s\right\},$$

$$\mathcal{E}_0 = \left\{\phi_S = \frac{1}{\sqrt{|S|}}\sum_{i\in S}|i\rangle \;:\; |S| = \ell\right\}.$$

The key technical result will be:

Theorem 5.3 (Theorem 1.3 restated)**.** *Let $\mathcal{H}=\mathbb{C}^d$ be a Hilbert space of dimension $d\in\mathbb{N}$, μ be the Haar measure on $\mathcal{H}$, and $S\subseteq[d]$ of size s. Then for any $k\in\mathbb{N}$,*

$$\left\|\int \psi^{\otimes k}d\mu(\psi) - \mathop{\mathbb{E}}_{S\subseteq[d],|S|=s}\phi_S^{\otimes k}\right\|_1 \le O\left(\frac{k^2}{d}+\frac{k}{\sqrt{s}}+\frac{sk}{d}\right),$$

where $\phi_S = \left(\frac{1}{\sqrt{s}}\sum_{i\in S}|i\rangle\right)\left(\frac{1}{\sqrt{s}}\sum_{i\in S}\langle i|\right)$.

In the above theorem, the subset state is compared with Haar random states, which by triangle inequality translates to a comparison between subset state of different subset sizes. The theorem itself implies new construction of quantum pseudorandom states as explained in the introduction. Note that the theorem is optimal in the following sense: When the subset size $s = O(\mathrm{poly}(n))$ is small, collision attacks illustrate that measuring some subset state in the computational basis of support size s for $O(\sqrt{s})$ times, a collision will be observed, that distinguishes subset state from Haar random state. When the support size is large, in particular, if $s=\Omega(d/\mathrm{poly}(n))$, then the overlap between the subset state and the uniform superposition of the computational basis will be significant, i.e., $1/\mathrm{poly}(n)$.

Then with polynomially many copies, the subset state will be distinguishable from the Haar random state.

As a corollary of Lemma 5.2 and Theorem 1.3, we have the following.

Theorem 5.4 (Failure of Standard Testing)**.** *Even given $\lceil 2^{n/16}\rceil$ copies, no tester can distinguish between subset states of size $\lceil 2^{n/8}\rceil$ from $\lceil 2^{n/4}\rceil$ with probability better than $O(2^{-n/16})$.*

So for the task of distinguishing subset state of very different support size, even with exponentially many copies, the advantage is still exponentially small.

Proof. Let $d = 2^n$, $k = \lceil 2^{n/16}\rceil$, $s = \lceil 2^{n/8}\rceil$, and $s' = \lceil 2^{n/4}\rceil$. Set

$$\mathcal{A} = \{\phi_S \mid S \subseteq [d], |S| = s\} \qquad \text{and} \qquad \mathcal{B} = \left\{\phi_S \mid S \subseteq [d], |S| = s'\right\}.$$

Let $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$ be uniform distributions on $\mathcal{A}$ and $\mathcal{B}$, respectively. By Theorem 1.3 and triangle inequality, we obtain

$$\begin{aligned}\|\rho_{\mathcal{A}} - \rho_{\mathcal{B}}\|_1 &\le \left\|\rho_{\mathcal{A}} - \int \psi^{\otimes k} d\mu(\psi)\right\|_1 + \left\|\int \psi^{\otimes k} d\mu(\psi) - \rho_{\mathcal{B}}\right\|_1 \\ &\le O\left(\frac{k^2}{d} + \frac{k}{\sqrt{s}} + \frac{sk}{d} + \frac{k}{\sqrt{s'}} + \frac{s'k}{d}\right) \\ &\le O\left(\frac{1}{2^{n/16}}\right),\end{aligned}$$

where the last inequality follows from our choices of d,k,s, and s'. Now, applying Lemma 5.2 to $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{B}}$, we conclude the proof. □

We now set off to prove Theorem 1.3. There will be three steps: 1. Give an approximate of the Haar random states; 2. Give an approximate of the random subset state; 3. Show that the two approximates are indistinguishable.

Approximate the Mixture of Haar Random States. First, let's look at the Haar random state. A well-known fact by representation theory gives an explicit formula for the mixture of Haar random states, $\Psi = \int \psi^{\otimes k} d\mu$, where μ is the Haar measure. For a detailed proof, see for example [Har13].

Fact 5.5.

$$\int \psi^{\otimes k} d\mu = \binom{d+k-1}{k}^{-1} \cdot \frac{1}{k!} \sum_{\pi \in S_k} \sum_{\vec{i} \in [d]^k} |\vec{i}\rangle\langle \pi(\vec{i})|.$$

Instead of working with Ψ directly, we look at the operator $\tilde{\Psi} = \Pi\Psi\Pi$, where Π is the projection onto the subspace of

$$\mathrm{span}\{|\vec{i}\rangle : \vec{i} \in A([d], k)\} \subseteq \mathcal{H}^{\otimes k}.$$

Immediately,

$$\tilde{\Psi} = \binom{d+k-1}{k}^{-1} \cdot \frac{1}{k!} \sum_{\pi \in S_k} \sum_{\vec{i} \in A([d],k)} |\vec{i}\rangle\langle \pi(\vec{i})|. \tag{5.1}$$

As long as k is small, we expect that $\Psi \approx \tilde{\Psi}$. This is simple and known. For completeness, we present a proof.

Proposition 5.6. $\|\Psi - \tilde{\Psi}\|_1 = O(k^2/d)$.

Proof. Consider the following decomposition of $\Psi := \tilde{\Psi} + \mathcal{R}$. Note that

$$\mathcal{R} = (I - \Pi)\Psi(I - \Pi).$$

It's clear that $\tilde{\Psi}$ and $\mathcal{R}$ are both positive semi-definite, and $\tilde{\Psi}\mathcal{R} = 0$. Therefore, the nonzero eigenspaces of Ψ correspond to those of $\tilde{\Psi}$ and $\mathcal{R}$, respectively. Consequently,

$$\begin{aligned} \left\|\Psi - \tilde{\Psi}\right\|_1 = 1 - \|\tilde{\Psi}\|_1 &= 1 - \frac{d^{\underline{k}}}{(d+k-1)^{\underline{k}}} \\ &= 1 - \frac{d}{d+k-1} \cdot \frac{d-1}{d+k-2} \cdots \frac{d-k+1}{d} \\ &\leq O\left(\frac{k^2}{d}\right). \end{aligned}$$

□

Approximate the Mixture of Random Subset State. Next, we turn to random subset states. Let $\Phi = \mathbb{E}_{|S|=s}\, \phi_S^{\otimes k}$, and consider $\Pi\Phi\Pi$, but normalized.[7] In particular,

$$\tilde{\Phi} = \mathop{\mathbb{E}}_{S:|S|=s}\left[\frac{1}{s^{\underline{k}}} \sum_{\vec{i},\vec{j}\in A(S,k),} |\vec{i}\rangle\langle\vec{j}|\right].$$

Analogous to Proposition 5.6, we have

Proposition 5.7. $\|\Phi - \tilde{\Phi}\|_1 = O(k/\sqrt{s})$.

Proof. Let γ be the uniform distribution over subset $S \subseteq [d]$ of size s,

$$\begin{aligned} &\left\|\int_S \left(\frac{1}{s^k}\sum_{\vec{i},\vec{j}\in S^k} |\vec{i}\rangle\langle\vec{j}| - \frac{1}{s^{\underline{k}}}\sum_{\vec{i},\vec{j}\in A(S,k)} |\vec{i}\rangle\langle\vec{j}|\right) d\gamma\right\|_1 \\ &\qquad \leq \int_S \left\|\frac{1}{s^k}\sum_{\vec{i},\vec{j}\in S^k} |\vec{i}\rangle\langle\vec{j}| - \frac{1}{s^{\underline{k}}}\sum_{\vec{i},\vec{j}\in A(S,k)} |\vec{i}\rangle\langle\vec{j}|\right\|_1 d\gamma \leq O\left(\frac{k}{\sqrt{s}}\right). \end{aligned}$$

□

All that is left to do is to show that $\|\tilde{\Phi} - \tilde{\Psi}\|_1$ is small. Fix any $\vec{i},\vec{j} \in A([d],k)$, and let $\ell = \ell(\vec{i},\vec{j})$ be the total number of distinct elements in the union of the elements of the vectors $\vec{i},\vec{j}$. Then the $(\vec{i},\vec{j})$'th entry of $\tilde{\Phi}$ is

$$\tilde{\Phi}(\vec{i},\vec{j}) = \frac{1}{s^{\underline{k}}} \Pr_{|S|=s}[\vec{i},\vec{j} \in A(S,k)] = \frac{1}{s^{\underline{k}}} \frac{\binom{d-\ell}{s-\ell}}{\binom{d}{s}} = \frac{s^{\underline{\ell}}}{s^{\underline{k}} \cdot d^{\underline{\ell}}}. \tag{5.2}$$

[7] Although in the case of Haar random state we didn't normalize, this doesn't really matter. Our choice is for simplicity of proof.

Comparing the Approximates.

Proposition 5.8. *For any $k \ll s \leq d$, it holds that*

$$\|\tilde{\Phi} - \tilde{\Psi}\|_1 = O\left(\frac{sk}{d}\right).$$

Proof. Let

$$\mathcal{D} = \tilde{\Phi} - \frac{(d+k-1)^{\underline{k}}}{d^{\underline{k}}}\tilde{\Psi}.$$

This factor is chosen so that $\mathcal{D}(\vec{i}, \vec{j}) = 0$ for any $\vec{i}$ and $\vec{j}$ such that $\vec{j} = \pi(\vec{i})$ for some permutation π. By triangle inequality,

$$\|\tilde{\Phi} - \tilde{\Psi}\|_1 \leq \|\mathcal{D}\|_1 + \left\|\tilde{\Psi} - \frac{(d+k-1)^{\underline{k}}}{d^{\underline{k}}}\tilde{\Psi}\right\|_1,$$

where the second term is bounded by $O(k^2/d)$.

We turn to $\mathcal{D}$. Let $\vec{j} \sim \vec{i}$ to denote that $\vec{j}$ is a permutation of $\vec{i}$. Note that for any $\vec{i} \sim \vec{j}$, $\mathcal{D}(\vec{i}, \cdot) = \mathcal{D}(\vec{j}, \cdot)$, and similarly $\mathcal{D}(\cdot, \vec{i}) = \mathcal{D}(\cdot, \vec{j})$. Therefore $\mathcal{D} = \tilde{\mathcal{D}} \otimes J$ where $J \in \mathbb{C}^{k! \times k!}$ is the all 1 matrix and $\tilde{\mathcal{D}} \in \mathbb{C}^{\binom{[d]}{k} \times \binom{[d]}{k}}$, s.t. for any $A, B \in \binom{[d]}{k}$,

$$\tilde{\mathcal{D}}(A, B) = \mathcal{D}(\vec{i}, \vec{j}), \qquad \vec{i}, \vec{j} \text{ contain } A \text{ and } B, \text{ respectively.}$$

Next, decompose $\tilde{\mathcal{D}} := \sum_{t=0}^{k-1} \alpha_t \mathcal{D}_t$, where in view of (5.2),

$$\alpha_t = \frac{(s-k)\cdots(s-2k+t+1)}{d\cdots(d-2k+t+1)},$$
$$\mathcal{D}_t(A, B) = \begin{cases} 1 & |A \cap B| = t, \\ 0 & \text{otherwise.} \end{cases}$$

$\mathcal{D}_t$ is the adjacency matrix for the well-studied generalized Johnson graphs [Del73]. In particular, we will need the following fact (explained in Appendix).

Fact 5.9. *For any $0 \leq t \leq k-1$, and for $k = O(\sqrt{d})$*

$$\|\mathcal{D}_t\|_1 \lesssim \binom{d-k}{k-t}\binom{d}{t}2^{k-t}.$$

Assisted by the above fact, we can bound $\|\mathcal{D}\|_1$ for $sk = O(d)$ as below,

$$\begin{aligned}
\|\mathcal{D}\|_1 = k!\|\tilde{\mathcal{D}}\|_1 \leq k!\sum_{t=0}^{k-1}\alpha_t\|\mathcal{D}_t\|_1 &\lesssim k!\sum_{t=0}^{k-1}\frac{s^{k-t}}{d^{2k-t}} \cdot \frac{d^k}{t!(k-t)!} \cdot 2^{k-t} \\
&= \sum_{t=0}^{k-1}\left(\frac{2s}{d}\right)^{k-t}\binom{k}{t} = \left(1 + \frac{2s}{d}\right)^k - 1 \lesssim O\left(\frac{2sk}{d}\right).
\end{aligned}$$

□

Theorem 1.3 follows by triangle inequality on Propositions 5.6-5.8.

5.4 Indistinguishability of Support Size in the Dense Regime

We have discussed that subset state with small support of fixed size are information-theoretically indistinguishable from Haar random states. This fact rules out property testing distinguishing general states with support size $s_0 = \omega(\mathrm{poly}(n))$ and $s_1 = 2^n/\omega(\mathrm{poly}(n))$. Then one may hope that property testing for large support size (constant density) may be possible.[8] In this section, we adapt a similar proof to show that this is also impossible.

In particular, we consider the following two ensembles for parameters $s, t, p \in (0, 1)$:

$$\mathcal{E}_1 = \left\{ \phi_S = \frac{1}{\sqrt{|S|}} \sum_{i \in S} |i\rangle \ : \ |S| \approx pd \right\},$$

$$\mathcal{E}_0 = \left\{ \phi_{S,T} = \frac{1}{\sqrt{2|S|}} \sum_{i \in S} |i\rangle - \frac{1}{\sqrt{2|T|}} \sum_{j \in T} |j\rangle \ : \ |S| \approx sd, |T| \approx td, S \cap T = \varnothing \right\}.$$

To be precise, the underlying distributions of the two ensembles are:

(i) For $\mathcal{E}_1$, sample state ϕ_S by letting $i \in S$ w.p. p for all $i \in [d]$ independently at random;

(ii) For $\mathcal{E}_0$, sample state $\phi_{S,T}$ by the following process: For each $i \in [d]$ independently, sample a uniformly random $r \in [0, 1]$, then let $i \in S$ if $r < s$; let $i \in T$ if $s \le r < s + t$.

Choose $p, s, t \in (0, 1)$ to be some constant such that

$$\sqrt{\frac{s}{2}} - \sqrt{\frac{t}{2}} = \sqrt{p}. \tag{5.3}$$

The choice is made so that states from the two ensembles have about the same overlap with $|\mu\rangle = \sum_{i \in [d]} |i\rangle / \sqrt{d}$. Otherwise, comparing with $|\mu\rangle$ will be a valid attack distinguishing the two ensembles. On the other hand, this overlap condition is the only thing matters: In $\mathcal{E}_0$, there are positive part and negative part, it is totally fine to have both parts positive, the analysis works equally well.

For concreteness, one can set $s = 8p$, and $t = 2p$ in (5.3). The support size of a random state from $\mathcal{E}_1$ will be $pd \pm \varepsilon d$ almost surely for arbitrarily small constant $\varepsilon > 0$; while the support size of a random state from $\mathcal{E}_0$ will be $(s + t)d \pm \varepsilon d$ almost surely. Note that $s + t = 10p$, i.e., states in $\mathcal{E}_0$ has 10 times larger support size than $\mathcal{E}_1$. A slight abuse of notation, we also use $\mathcal{E}_0, \mathcal{E}_1$ to denote the mixed state for states of the average state $\mathcal{E}_0, \mathcal{E}_1$, respectively. Our goal is to show that

$$\|\mathcal{E}_0 - \mathcal{E}_1\| = \mathrm{negl}(n).$$

Approximates of $\mathcal{E}_1$. We consider approximates of the average of state from $\mathcal{E}_1$,

$$\mathcal{E}_1 \xrightarrow{\Pi} \mathcal{E}_1' \xrightarrow{\text{flatten}} \mathcal{E}_1''.$$

Recall Π be the projection onto the subspace $\mathrm{span}\{|\vec{i}\rangle : \vec{i} \in A([d], k)\}$, then $\mathcal{E}_1' = \Pi \mathcal{E}_1 \Pi$. In the approximate $\mathcal{E}_1''$, we pretend that the "amplitudes" do not depend on $|S|$, in other words

[8] In the sparse regime, by viewing the subset state as a distribution on polynomially many things, the problem can be understood via classical argument.

we “flatten” the distribution that we sample the state. In particular,

$$\mathcal{E}_1 = \mathop{\mathbb{E}}_{S}\left[|S|^{-k}\left(\sum_{i\in S}|i\rangle\right)^{\otimes k}\left(\sum_{i\in S}\langle i|\right)^{\otimes k}\right];$$

$$\mathcal{E}_1' = \mathop{\mathbb{E}}_{S}\left[|S|^{-k}\sum_{\vec{i},\vec{j}\in A(S,k)}|\vec{i}\rangle\langle\vec{j}|\right];$$

$$\mathcal{E}_1'' = \mathop{\mathbb{E}}_{S}\left[(pd)^{-k}\sum_{\vec{i},\vec{j}\in A(S,k)}|\vec{i}\rangle\langle\vec{j}|\right].$$

We compare $\mathcal{E}_1$ and $\mathcal{E}_1'$ as follows

$$\begin{aligned}\|\mathcal{E}_1 - \mathcal{E}_1'\| &\le \mathop{\mathbb{E}}_{S}|S|^{-k}\left\|\sum_{\vec{i},\vec{j}\in S^k}|\vec{i}\rangle\langle\vec{j}| - \sum_{\vec{i},\vec{j}\in A(S,k)}|\vec{i}\rangle\langle\vec{j}|\right\| \\ &\le O\left(\mathop{\mathbb{E}}_{S}\sqrt{\frac{k^2}{|S|}}\right) = O\left(\frac{k}{\sqrt{d}}\right).\end{aligned}$$

Next compare $\mathcal{E}_1'$ and $\mathcal{E}_1''$, we claim

$$\|\mathcal{E}_1'' - \mathcal{E}_1'\| \le O\left(\frac{k}{d^{2/5}}\right). \tag{5.4}$$

Consider the interval $L = pd \pm d^{3/5}$. By Chernoff Bound, the probability that $|S| \notin L$ is $\exp(-\Omega(d^{1/5}))$. Then (5.4) is a direct conclusion from the following two bounds.

1. For $S \in L$,

$$\begin{aligned}\left\|(|S|^{-k} - (pd)^{-k})\sum_{\vec{i},\vec{j}\in S^k}|\vec{i}\rangle\langle\vec{j}|\right\| &= |S|^{\underline{k}}\cdot\frac{|S|^k - (pd)^k}{(|S|pd)^k} \\ &\le \left(1 + \frac{k^2}{|S|}\right)\left(\left(1 + \frac{1}{pd^{2/5}}\right)^k - 1\right) \\ &\le O\left(\frac{k}{d^{2/5}}\right).\end{aligned}$$

2. For $S \notin L$,

$$\sum_{S:|S|\notin L}\Pr[S]\left\|(pd)^{-k}\sum_{\vec{i},\vec{j}\in S^k}|\vec{i}\rangle\langle\vec{j}|\right\|,\ \sum_{S:|S|\notin L}\Pr[S]\left\||S|^{-k}\sum_{\vec{i},\vec{j}\in S^k}|\vec{i}\rangle\langle\vec{j}|\right\| \le \exp(-\Omega(d^{1/5})).$$

If follows that

$$\|\mathcal{E}_1 - \mathcal{E}_1''\| \le O\left(\frac{k}{d^{2/5}}\right).$$

For any $\vec{i},\vec{j} \in A([d],k)$, compute the entry of $\mathcal{E}''(\vec{i},\vec{j})$ explicitly. Note the entry depends only on $\ell = |\vec{i}\cup\vec{j}|$,

$$\langle\vec{i}|\mathcal{E}_1'|\vec{j}\rangle = p^{\ell-k}\cdot d^{-k}. \tag{5.5}$$

Approximates of $\mathcal{E}_0$. We consider approximates of the average of state from $\mathcal{E}_0$ completely analogously (and a lot more tedious) to $\mathcal{E}_1$,

$$\mathcal{E}_0 \xrightarrow{\Pi} \mathcal{E}_0' \xrightarrow{\text{flatten}} \mathcal{E}_0''.$$

In particular,

$$\begin{aligned}
\mathcal{E}_0 &= \mathop{\mathbb{E}}_{S,T}\left[\left(\frac{1}{\sqrt{2|S|}}\sum_{i\in S}|i\rangle - \frac{1}{\sqrt{2|T|}}\sum_{j\in T}|j\rangle\right)^{\otimes k}\left(\frac{1}{\sqrt{2|S|}}\sum_{i\in S}\langle i| - \frac{1}{\sqrt{2|T|}}\sum_{j\in T}\langle j|\right)^{\otimes k}\right];\\
\mathcal{E}_0' &= \Pi\mathcal{E}_0\Pi;\\
\mathcal{E}_0'' &= \Pi\left(\mathop{\mathbb{E}}_{S,T}\left[\left(\frac{1}{\sqrt{2sd}}\sum_{i\in S}|i\rangle - \frac{1}{\sqrt{2td}}\sum_{j\in T}|j\rangle\right)^{\otimes k}\left(\frac{1}{\sqrt{2sd}}\sum_{i\in S}\langle i| - \frac{1}{\sqrt{2td}}\sum_{j\in T}\langle j|\right)^{\otimes k}\right]\right)\Pi.
\end{aligned}$$

The analysis would also be completely analogous to that of $\mathcal{E}_1$. We omit the calculations here,

$$\|\mathcal{E}_0 - \mathcal{E}_0''\| \le O\left(\frac{k}{d^{2/5}}\right).$$

Now we compute the entry for $\mathcal{E}_0''$ explicitly. Fix any $\vec{i}, \vec{j} \in A([d], k)$, let $\ell = \ell(\vec{i}, \vec{j})$ be the total number of distinct elements in vectors $\vec{i}$ union $\vec{j}$. Let $a := 2(\ell - k), b := 2k - \ell$. So a is the number of elements that appeared only in $\vec{i}$ or $\vec{j}$, while b is the number of elements that appeared in both $\vec{i}$ and $\vec{j}$. Then

$$\begin{aligned}
\langle\vec{i}|\mathcal{E}_0''|\vec{j}\rangle &= \sum_{\ell_1=0}^{a}\sum_{\ell_2=0}^{b}\binom{a}{\ell_1}\binom{b}{\ell_2}\left(\frac{1}{\sqrt{2sd}}\right)^{\ell_1+2\ell_2} s^{\ell_1+\ell_2}\left(\frac{-1}{\sqrt{2td}}\right)^{a-\ell_1+2(b-\ell_2)} t^{a-\ell_1+b-\ell_2}\\
&= \sum_{\ell_1=0}^{a}\sum_{\ell_2=0}^{b}\binom{a}{\ell_1}\binom{b}{\ell_2}\left(\frac{s}{\sqrt{2sd}}\right)^{\ell_1}\left(\frac{1}{\sqrt{2d}}\right)^{2\ell_2}\left(\frac{-t}{\sqrt{2td}}\right)^{a-\ell_1}\left(\frac{1}{\sqrt{2d}}\right)^{2(b-\ell_2)}\\
&= \sum_{\ell_1=0}^{a}\binom{a}{\ell_1}\left(\frac{s}{\sqrt{2sd}}\right)^{\ell_1}\left(\frac{-t}{\sqrt{2td}}\right)^{a-\ell_1}\sum_{\ell_2=0}^{b}\binom{b}{\ell_2}\left(\frac{1}{2d}\right)^{b}\\
&= \left(\frac{s}{\sqrt{2sd}} - \frac{t}{\sqrt{2td}}\right)^{a}\left(\frac{1}{d}\right)^{b}\\
&= \left(\sqrt{\frac{p}{d}}\right)^{a}\left(\frac{1}{d}\right)^{b}\\
&= \frac{p^{\ell-k}}{d^k}. \qquad (5.6)
\end{aligned}$$

Indistinguishability of The Two Ensembles Note, $\mathcal{E}_0'' = \mathcal{E}_1''$ by (5.5) and (5.6). By triangle inequality,

$$\|\mathcal{E}_0 - \mathcal{E}_1\| \le O\left(\frac{k}{d^{2/5}}\right) = \text{negl}(n).$$

5.5 Quantum and Classical Property Testing

The most natural classical counterpart of the above quantum property testing would be the property testing for classical probability distributions. In the property testing for classical distribution, one is given a unknown probability distribution ν where the only way to access ν is to draw independent samples. The goal is to test whether the unknown distribution ν satisfies certain property, e.g., whether ν has large support size or not. To make the connection between quantum property testing and classical property testing for distributions precise, we formalize this connection by encoding classical distributions as states and show that lower bounds on quantum property testing imply lower bounds for property testing of classical distributions.

Let $\mathcal{A} \subseteq \Delta_N$ be a collection of probability distributions in $\mathbb{R}^N$. We say that $|\psi\rangle$ has classical shadow in $\mathcal{A}$ if $|\psi\rangle = \sum_{i=1}^N \alpha_i |i\rangle$ satisfying $(|\alpha_1|^2, \ldots, |\alpha_N|^2) \in \mathcal{A}$. This definition generalizes to mixed states, i.e., ψ has classical shadow in $\mathcal{A}$ if ψ can be expressed as some mixture of pure states where each pure state has a classical shadow in $\mathcal{A}$. Now a further generalization to reflect more than one "samples" from classical distribution, a general state ρ has a classical k-shadow in $\mathcal{A}$ if for some distribution λ on the ℓ_2-unit sphere of $\mathbb{C}^N$ and $k \in \mathbb{N}$, the state ρ can be expressed as

$$\rho = \mathbb{E}_{|\psi\rangle \sim \lambda} \left(|\psi\rangle\langle\psi| \right)^{\otimes k},$$

with every $|\psi\rangle \sim \lambda$ has a shadow in $\mathcal{A}$. So quantum states with k-shadow of $\mathcal{A}$ corresponds to the natural quantum counterparts for a mixture $\mathcal{D}$ of distributions in $\mathcal{A}$ from where k samples will be drawn. Precisely, let Λ be the channel that measures each copy $|\psi\rangle\langle\psi|$ in the standard computational basis. Then the effect of Λ is to make k samples from a random distribution $\nu \in \mathcal{D}$.

Claim 5.10. *If ρ has a classical k-shadow in $\mathcal{A} \subseteq \Delta_N$, then*

$$\Lambda(\rho) = \mathbb{E}_{\nu \sim \mathcal{D}} \left(\sum_{i=1}^N \nu_i |i\rangle\langle i| \right)^{\otimes k}, \tag{5.7}$$

where $\mathcal{D}$ is some distribution on $\mathcal{A}$.

Using Theorem 1.3 and Claim 5.10, we deduce the following lower bound.

Corollary 5.11 (Failure of Standard Classical Testing)**.** *Even given $\lceil 2^{n/16} \rceil$ samples of a flat distribution, no tester can distinguish between support size $\lceil 2^{n/8} \rceil$ from $\lceil 2^{n/4} \rceil$ with probability better than $O(2^{-n/16})$.*

Proof. Consider two ensemble of subset states $\rho_{\mathcal{A}}$ and $\rho_{\mathcal{B}}$ with subset sizes $\lceil 2^{n/8} \rceil$ and $\lceil 2^{n/4} \rceil$, respectively. From Theorem 1.3, we deduce that $\|\rho_{\mathcal{A}} - \rho_{\mathcal{B}}\|_1 \leq O(2^{-n/16})$. Note that Λ does not increase the trace distance, so $\|\Lambda(\rho_{\mathcal{A}}) - \Lambda(\rho_{\mathcal{B}})\|_1 \leq O(2^{-n/16})$ concluding the proof. □

6 Testing with Classical Certificates: Another Fiasco

In this section, we discuss property testing in the presence of classical certificates. We deliberately choose the word certificates to contrast the notion of proofs in the definition of

PropMA. It's hard to give a formal definition, our intention is that certificates would mean a more general notion—we will see examples.

Our investigation in this section focuses on the concrete problem: establishing lower bounds for certifying the support size. We consider the problem $\mathrm{GapSupp}_{s,\ell}$, that is to distinguish the following two ensembles:

Definition 6.1 (The Gap Support Size Problem)**.** *The* $\mathrm{GapSupp}_{s,\ell} = (\mathcal{P}_{\mathrm{yes}}, \mathcal{P}_{\mathrm{no}})$, *where*

(YES) $\mathcal{E}_1 = \{$*uniform distribution on support* $S : S \subseteq [N], |S| = s\}$
(NO) $\mathcal{E}_0 = \{$*uniform distribution on support* $S : S \subseteq [N], |S| = \ell\}$.

Since $\mathrm{GapSupp}_{s,\ell}$ can be thought of a special case of quantum state property due to Section 5.5, we think N as some exponentially large quantity, and let $n = \log N$.

6.1 Lower Bounds on Classical Testing with Proofs via HDX Fast Mixing

Our main result in the section is the following.

Theorem 6.2 (Classical Indistinguishability for Subset Distribution with Certificates)**.** *For some parameter* $s = \omega(\mathrm{poly}(n))$*, given any proof of length* p *and allowing* t *samples, then the verifier can distinguish* $\mathrm{GapSupp}_{s,2s}$ *with an advantage at most*

$$O\left(\sqrt{\frac{tp}{s}} + \frac{t^2}{s}\right).$$

Remark 6.3 (Optimality)**.** *The above lower bound is tight in general. To see this, note that one obvious strategy that the honest prover can do is to send set* T *consisting of* p *elements from the support of size* s *at the cost about* $p\log(N/p)$ *communication complexity. In the yes case, the probability of seeing a collision with* T *sampling* s/p *elements is* $(1 - p/s)^{s/p}$*; while in the no case, the probability seeing a collision with* T *in the samples is* $(1-p/2s)^{s/p}$*. The two probabilities can differ by* $\Omega(1)$*.*

Proof. Suppose we are assisted with a proof of length p. For any string $\Pi \in \{0,1\}^p$, let $\mathcal{F}_\Pi$ denote the set of $S \in \mathcal{E}_1$ such that the faithful prover will provide the proof Π. Now there will be two situations, one with a faithful prover, one with the adversarial prover:

(Yes) Consider the YES distribution indicated by its support S chosen randomly from $\mathcal{E}_1$. Let Π be the faithful proof associated with S. The verifier will sample t elements from the distribution. Overall, the verifier observes $X_1, X_2, \ldots, X_t$ together with the proof Π.

(NO) Suppose a uniformly random distribution indicated by its support S' is chosen from $\mathcal{E}_0$. The adversary will send a proof Π' to the verifier with probability $|\mathcal{F}_{\Pi'}|/\binom{N}{s}$, independent to S'. The verifier samples t elements from the uniform distribution on S' together with the adversary proof Π'. So the verifier sees $Y_1, Y_2, \ldots, Y_t$ together with an adversarial proof Π'.

Let ν_1 and ν_0 denote the distribution on samples together with the proofs that the verifier sees in the YES and NO case, respectively.

Now, note that for $t \ll s$, the probability that $X_1, X_2, \ldots, X_t$ consist some collision is $O(t^2/s)$. Therefore in YES case, we can alternatively think of the distribution of $X_1, X_2, \ldots, X_t, \Pi$, as follows

Modified Faithful Process (To generate $\tilde{\nu}_1$ on $X_1, X_2, \ldots, X_t, \Pi$):

(i) Sample Π with probability $|\mathcal{F}_\Pi|/\binom{N}{s}$;

(ii) Sample $S \in \mathcal{F}_\Pi$ at random;

(iii) Sample a subset $T \subseteq S$ of size t uniformly at random;

(iv) Let $X_1, X_2, \ldots, X_t$ be some uniformly random permutation τ on T.

$$\nu \xrightarrow{\text{collisionless}} \tilde{\nu}.$$

The new distribution, denoted $\tilde{\nu}_1$, differs from the old ν_1 in statistical distance by $O(t^2/s)$. Analogously, consider the distribution on $\tilde{\nu}_0$ which differs from ν_0 in statistical distance at most $O(t^2/s)$ described below,

Modified Adversarial Process (To generate $\tilde{\nu}_0$ on $Y_1, Y_2, \ldots, Y_t, \Pi'$):

(i) Sample Π' based on $|\mathcal{F}_{\Pi'}|/\binom{N}{s}$;

(ii) Sample $R \in \mathcal{E}_0$ uniformly at random;

(iii) Sample a subset $S' \in R$ uniformly at random of size s;

(iv) Sample a subset $T' \subseteq S'$ of size t uniformly at random;

(v) Let $Y_1, Y_2, \ldots, Y_t$ be some uniformly random permutation τ' on T.

In the above description for $\tilde{\nu}_0$, step (iii) is redundant as sampling t-subset T' from an s-subset S' that itself is a random subset of R is the same as sampling a t-subset T' from R directly. Furthermore, the overall distribution of T' is uniform over t-subset of $[N]$ (and S' will be a uniform s-subset of $[N]$). This redundancy is introduced for the purpose of analysis.

Therefore, $\tilde{\nu}_1$ corresponds to essentially what the verifier reads in the YES case, and $\tilde{\nu}_0$ corresponds to what the verifier reads in the NO case. Based on the discussion so far, to prove our claimed bound in the theorem

$$\|\nu_0 - \nu_1\| \leq \sqrt{\frac{tp}{2s}} + O(t^2/s),$$

it suffices to show

$$\|\tilde{\nu}_0 - \tilde{\nu}_1\|_1 \leq \sqrt{\frac{tp}{2s}}. \tag{6.1}$$

Now we justify the inequality (6.1).

$$\begin{aligned}
2\|\tilde{\nu}_0 - \tilde{\nu}_1\|^2 &\overset{(1)}{\leq} \mathrm{KL}(\nu_0 \,\|\, \nu_1) = \mathrm{KL}(\Pi T \tau \,\|\, \Pi' T' \tau') \\
&\overset{(2)}{=} \mathrm{KL}(\Pi\tau \,\|\, \Pi'\tau') + \mathop{\mathbb{E}}_{\pi,\sigma} \mathrm{KL}\left(\begin{array}{c} T \mid \Pi = \pi, \tau = \sigma \\ \hline\hline T' \mid \Pi' = \pi, \tau' = \sigma \end{array}\right) \\
&\overset{(3)}{=} \mathop{\mathbb{E}}_{\pi} \mathrm{KL}\left(\begin{array}{c} T \mid \Pi = \pi \\ \hline\hline T' \mid \Pi' = \pi \end{array}\right) \\
&\overset{(4)}{=} \mathop{\mathbb{E}}_{\pi} \mathrm{KL}\left(\begin{array}{c} T \mid \Pi = \pi \\ \hline\hline T' \end{array}\right) \\
&\overset{(5)}{\leq} \mathop{\mathbb{E}}_{\pi} \frac{t}{s} \mathrm{KL}\left(\begin{array}{c} S \mid \pi \\ \hline\hline S' \end{array}\right) \\
&\overset{(6)}{\leq} \frac{tp}{s}, \qquad (6.2)
\end{aligned}$$

where (1) uses Pinsker's inequality, and note there is a natural bijection between Π, T, τ and $\Pi, X_1, X_2, \ldots, X_t$ (analogously for Π', T', τ' and $\Pi', Y_1, Y_2, \ldots, Y_t$); (2) is by Chain rule for KL-divergence; (3) holds because $\Pi\tau$ and $\Pi'\tau'$ have the same distribution and the random permutation τ (τ') is independent from T, Π (T', Π'); (4) holds because in the adversary case the proof is independent with T; (5) invokes a divergence contraction result Lemma 6.4 that we explain later; and (6) holds because S' is uniform over $\binom{[N]}{s}$ as S, and by definitions of mutual information and KL-divergence,

$$\mathop{\mathbb{E}}_{\pi} \mathrm{KL}((S \mid \pi) \,\|\, S) = I(S; \Pi) \leq H(\Pi) \leq p. \qquad \square$$

The missing technical component for the above proof is the following divergence contraction lemma, which is studied in the theory of higher-dimensional expanders. In particular, it is an application of the more general theorems proved in [CGM19, AJK+22].

Lemma 6.4 (Divergence contraction)**.** *Let μ_0 be the uniform distribution over $\binom{[N]}{s}$, and μ_1 be some distribution over $\binom{[N]}{s}$. Consider the following random process for $i \in \{0, 1\}$:*

(i) *Sample S from μ_i,*

(ii) *Sample subset T of size t from S uniformly at random.*

The above random process introduces a distribution λ_i. Then

$$\mathrm{KL}(\lambda_1 \,\|\, \lambda_0) \leq \frac{t}{s} \cdot \mathrm{KL}(\mu_1 \,\|\, \mu_0). \qquad (6.3)$$

For completeness, we provide a self-contained proof using a language consistent with our discussion so far, where (6.3) is replaced with a slightly weaker bound:

$$\mathrm{KL}(\lambda_1 \,\|\, \lambda_0) \leq \frac{t}{s - t + 1} \cdot \mathrm{KL}(\mu_1 \,\|\, \mu_0).$$

Note that for our application, $t \leq O(\sqrt{s})$, thus, t/s and $t/(s - t + 1)$ are essentially the same. For the tighter bound, see [AJK+22, Thoerem 5].

Proof. Consider the random variables $X_1, X_2, \ldots, X_s$ which are obtained by drawing a random subset S from μ_1, and then permute the elements by a random permutation τ. Similarly, the random variables $Y_1, Y_2, \ldots, Y_s$ will be obtained by first drawing a random subset S' from μ_0, then randomly order the elements in the subset by τ'. Note that

$$\begin{aligned}\mathrm{KL}(X_1X_2\ldots X_s \,\|\, Y_1Y_2\ldots Y_s) &= \mathrm{KL}(\tau S \,\|\, \tau' S')\\ &= \mathrm{KL}(\tau \,\|\, \tau') + \mathop{\mathbb{E}}_{\sigma} \mathrm{KL}\left(\frac{S \mid \tau = \sigma}{S' \mid \tau' = \sigma}\right) = \mathrm{KL}(S \,\|\, S') = \mathrm{KL}(\mu_1 \,\|\, \mu_0),\end{aligned}$$

where the second step follows the chain rule, and third step holds as the permutations τ, τ' are independent of S, S'. Analogously,

$$\mathrm{KL}(X_1X_2\ldots X_t \,\|\, Y_1Y_2\ldots Y_t) = \mathrm{KL}(\lambda_1 \,\|\, \lambda_0).$$

By chain rule, for any $\ell \le s$,

$$\mathrm{KL}(X_1X_2\ldots X_\ell \,\|\, Y_1Y_2\ldots Y_\ell) = \sum_{i=1}^{\ell} \mathop{\mathbb{E}}_{x \in A([N],s)} \mathrm{KL}\left(\frac{X_i \mid X_{<i} = x_{<i}}{Y_i \mid Y_{<i} = x_{<i}}\right).$$

Now we need the following claim.

Claim 6.5 (cf. Theorem 4 [AJK$^+$22])**.** *For any $x \in A([N], s)$ and $1 \le i \le s$,*

$$\mathrm{KL}\left(\frac{X_i \mid X_{<i} = x_{<i}}{Y_i \mid Y_{<i} = x_{<i}}\right) \le \frac{1}{s-i+1} \cdot \mathrm{KL}\left(\frac{X_iX_{i+1}\ldots X_s \mid X_{<i} = x_{<i}}{Y_iY_{i+1}\ldots Y_s \mid Y_{<i} = x_{<i}}\right). \tag{6.4}$$

The proof of the claim is deferred to the appendix. With the claim, we can finish the proof.

$$\begin{aligned}\mathrm{KL}(\lambda_1 \,\|\, \lambda_0) &= \sum_{i=1}^{t} \mathop{\mathbb{E}}_{x \in A([N],s)} \mathrm{KL}\left(\frac{X_i \mid X_{<i} = x_{<i}}{Y_i \mid Y_{<i} = x_{<i}}\right).\\ &\le \sum_{i=1}^{t} \frac{1}{s-i+1} \cdot \mathop{\mathbb{E}}_{x \in A([N],s)} \mathrm{KL}\left(\frac{X_iX_{i+1}\ldots X_s \mid X_{<i} = x_{<i}}{Y_iY_{i+1}\ldots Y_s \mid Y_{<i} = x_{<i}}\right)\\ &\le \sum_{i=1}^{t} \frac{1}{s-i+1} \cdot \mathrm{KL}(X_1X_2\ldots X_s \,\|\, Y_1Y_2\ldots Y_s)\\ &\le \frac{t}{s-t+1} \cdot \mathrm{KL}(\mu_1 \,\|\, \mu_0).\end{aligned}$$

□

In Theorem 6.2, the lower bound is stated for $\mathrm{GapSupp}_{s,2s}$, the YES case corresponds to the distribution of the smaller support size, and the NO case corresponds to the distribution of the larger support size. The support size of the NO case is somewhat arbitrary, and one can consider $\mathrm{GapSupp}_{s,\ell}$ for $\ell > 2s$, or simply $\mathrm{GapSupp}_{s,N}$ where the NO case is simply the uniform distribution over $[N]$ (no ensemble at all). This a-priory makes the distinguishing task easier, but the same analysis holds and results in the same asymptotic bound, i.e., allowing t samples it's impossible to distinguish YES case from uniform distribution with an advantage better than $O(\sqrt{tp/s} + t^2/s)$.

One can also consider $\mathrm{GapSupp}_{2s,s}$, i.e., the NO case having the smaller support size. This case is also captured by the same analysis. However, when the NO case has a smaller support size, it admits a much stronger lower bound that trivializes the problem. We discuss this case in Section 6.4.

6.2 A Generalization: Testing with Structured Classical Certificates

Note that in our proof to Theorem 6.2, regardless of whether the proof Π associated with each subset S is fixed or randomized, the analysis is exactly the same as long as (i) in the modified faithful process, the joint distribution of the proof and the distribution indicated by S are set correctly; (ii) in the modified adversarial process, the marginal distribution of Π' is set correctly. Therefore, our analysis is robust. We discuss the implication of the robustness formally.

Stated in the most general way, our lower bound technique works for certificates coming from any promised (intended to help testability) convex subset $\mathcal{C}$ of the probability simplex $\Delta_{\mathcal{S}}$ in $\mathbb{R}^{\mathcal{S}}$, where $\mathcal{S}$ is an arbitrary finite set representing all the possible certificates. In view of the proof to Theorem 6.2, Π can be any random variable depending only on the distribution $\mu_1 \in \mathcal{E}_1$ to test. In other words, an element from $\Delta_{\mathcal{S}}$ may represent some certificate on a YES input μ_1. A certificate can be the proof from the prover in the case of PropMA model stated in Theorem 6.2, and can be the entire communication transcript in the case of the (public coin) PropAM model.

Starting from the trivial example, the delta-distributions on $\mathcal{S} = \{0,1\}^p$, i.e., distributions supported on a single element in $\mathcal{S}$. Such distributions correspond to the case where for a fixed input, there is a fixed proof of length m. Therefore, we can let Δ_p^{MA} denote the set of delta-distributions on $\mathcal{S}$,

$$\Delta_p^{\mathrm{MA}} = \{\text{singleton distribution} \in \Delta_{\mathcal{S}} : \mathcal{S} = \{0,1\}^p\}\,.$$

Allowing the convex hull of Δ^{MA}, we obtain all distributions $\Delta_{\mathcal{S}}$—indeed a trivial example. This model captures MA proofs: Because for any MA protocol, in the yes case there is a delta-distribution corresponding to the honest MA proof that will be accepted with high probability. In the no case, no proof will be accepted with high probability. Therefore for any mixed strategy from $\mathrm{conv}\left(\Delta^{\mathrm{MA}}\right)$, the verifier will reject with high probability. Consequently, we can give an alternative definition of the MA type property testing model.

Definition 6.6 (Classical Property Testing with MA Type Certificates)**.** *Let $k = k(d), p = p(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropMA}(k,p,a,b)$ *with respect to certificates $\mathcal{C} = \mathrm{conv}\left(\Delta_p^{\mathrm{MA}}\right)$ if there exists a verifier V such that for every $\mu \in \Delta_d$,*

(i) *if $\nu \in \mathcal{P}_d$, then there exist $\mu \in \mathcal{C}$ such that*

$$\Pr_{x_1,\ldots,x_k \sim \mu^{\otimes k}, y \sim \mu}[V(x_1,\ldots,x_k,y) \text{ accepts}] \geq a\,,$$

(ii) *if ν is ε-far from $\mathcal{P}_d$ (in statistical distance), then for every certificate $\mu \in \mathcal{C}$ such that*

$$\Pr_{x_1,\ldots,x_k \sim \mu^{\otimes k}, y \sim \mu}[V(x_1,\ldots,x_k,y) \text{ accepts}] \leq b\,.$$

Replacing $\mathrm{conv}\left(\Delta^{\mathrm{MA}}\right)$ in Definition 6.6 by any other convex set $\mathcal{C} \subseteq \Delta_p$, one obtains $\mathcal{C}$ type certificates. Then Theorem 6.2 can be stated in a more generalized way.

Theorem 6.7 (Indistinguishability for Subset Distribution with Structured Certificates)**.** *For some parameter $s = \omega(\mathrm{poly}(n))$, given any certificates of type $\mathcal{C} \subseteq \Delta_p$ and allowing t samples,* $\mathrm{GapSupp}_{s,2s}$ *with an advantage at most*

$$O\left(\sqrt{\frac{tp}{s}} + \frac{t^2}{s}\right).$$

We now instantiate $\mathcal{C}$ to certificates arising from a valid AM protocol—the communication transcripts. To illustrate, suppose $\mathcal{Z} = \{0,1\}^p$ where $p = rn + r$, and consider r-round AM protocols where the verifier sends n uniformly random bits to the prover and receives 1 answer bit at each round. A distribution μ on $\mathcal{Z}$ naturally defines random variables $R_1A_1 \dots R_rA_r$, where each R_i takes values in $\{0,1\}^n$ and $A_i \in \{0,1\}$, such that $\Pr[(R_1 = r_1, A_1 = a_1, \dots, R_r = r_r, A_r = a_r)] = \mu(r_1, a_1, \dots, r_r, a_r)$. The collection of distributions encoding valid AM protocols of this form is given by

$$\Delta_p^{\mathrm{AM}} = \left\{ \mu \in \Delta_{\mathcal{Z}} \ : \begin{array}{r} \forall i, r_1, \dots, r_i,\ (R_i \,|R_1 = r_1 \dots R_{i-1} = r_{i-1} \text{ is uniform}) \\ \text{and } (A_i \mid R_1 = r_1 \dots R_i = r_i \text{ is fixed}) \end{array} \right\}.$$

Analogous to the discussion in the last paragraph, taking $\mathrm{conv}\left(\Delta_p^{\mathrm{AM}}\right)$ captures the power of AM provers.

For another more structured example, suppose $\mathcal{S}$ is equal to the Cartesian product $\mathcal{Z} \times \cdots \times \mathcal{Z}$ with m copies, we can take $\mathcal{C} = \mathrm{conv}\left(\{\mu^{\otimes k} \mid \mu \in \Delta_{\mathcal{Z}}\}\right)$. In words, $\mathcal{C}$ is the convex hull of of i.i.d. distributions. Using this notation, we can for instance take $\mathrm{conv}\left(\{\mu^{\otimes m} : \mu \in \Delta^{\mathrm{AM}}\}\right)$, or $\mathrm{conv}\left(\{\mu_1^{\otimes m_1} \otimes \cdots \otimes \mu_\ell^{\otimes m_\ell} : \mu_1, \dots, \mu_\ell \in \Delta^{\mathrm{AM}}\}\right)$. This can be thought of as capturing AM with multiple independent provers. Analogously, $\mathrm{PropAM}(m)$ for multiple provers.

It then follows that

Corollary 6.8. *For* PropAM *tester with even unboundedly many independent provers and unboundedly many rounds, to solve* $\mathrm{GapSupp}_{s,\ell}$ *with a constant advantage, where $\ell > s$, the sample complexity t and the proof complexity p, must satisfy*

$$t(p+t) = \Omega(\sqrt{s}).$$

6.3 A Private-coin AM Protocol

In the previous subsection, we proved a tight lower bound for GapSupp with classical certificates. The lower bounds holds even when allowing public-coin AM type certificates. In this section, we point out that the public-coin restriction is important as there is a very efficient private-coin AM tester, i.e., two-round IP tester. This tester is adapted from [HR22].

Theorem 6.9 (cf. [HR22])**.** *There is a private-coin* AM *tester, e.g. Algorithm 6.10, for* $\mathrm{GapSupp}_{\frac{N}{3},\frac{2N}{3}}$ *using $O(1)$ samples and $O(1)$ communication.*

Algorithm 6.10: A Private-coin AM Tester for $\mathrm{GapSupp}_{N/3,2N/3}$

Input: Unknown distribution μ

Arthur:

(i) Make k samples from μ for some large enough constant k, denote the set of elements sampled from μ by S;

(ii) Make k samples uniformly at random from $[N]$, denote the set R;

(iii) Send the set $M := S \cup R$ to Merlin in a random order.

Merlin: Merlin return a subset $M' \in M$ such that $M' = M \cap \mathrm{supp}\,\mu$

Accept if $|M'| \leq 1.5k$ and $S \subseteq M'$.

Proof. For large enough N, almost surely, the set M contains $2k$ elements. If the prover is honest, $M' = M \cap \operatorname{supp} \mu$. If $\operatorname{supp} \mu = N/3$, by chernoff bound, $|M'| \leq |S| + |R|/2$ with probability at least $1 - \exp(-\Omega(k))$. This establishes the completeness case.

In the soundness case, $\operatorname{supp} \mu \geq 2N/3$. Then with probability $1 - \exp(-\Omega(k))$, $|M \cup \operatorname{supp} \mu| \geq |S| + 7|R|/12 = 1.5k + k/12$. To fool Arthur, Merlin needs to send $|M'| \leq 1.5k$, thus with probability $1 - \exp(-\Omega(k))$,

$$\left|(M \cup \operatorname{supp} \mu) \setminus M'\right| \geq \frac{k}{12}.$$

However, the verifier has no information about which element in $M \cup \operatorname{supp} \mu$ is in S, that is M' is determined by $M \cup \operatorname{supp} \mu$ and independent of S. Note that S is just a uniformly random subset of $M \cup \operatorname{supp} \mu$ of size k. Thus, with probability at most $\exp(-\Omega(k))$, $S \subseteq M'$. It concludes that in the soundness case, Arthur accepts with probability at most $\exp(-\Omega(k))$. □

6.4 One-sidedness of the GapSupp

So far, we considered the upper and lower bounds for $\mathrm{GappSupp}_{s,\ell}$, where $s \leq \ell$. In words, we wanted to test that the flat distribution has a small support. We now justify this choice. In particular, suppose we want to test whether a given distribution has large support, e.g., consider $\mathrm{GapSupp}_{\ell,s}$ for $\ell > s$, then the proof does not improve the testability at all no matter how long it is.

Theorem 6.11. *Consider two ensembles of distributions for some parameters $\ell \gg s$. Given $t = o(\sqrt{s})$ samples from the distribution, there is no tester that can solve* $\mathrm{GapSupp}_{\ell,s}$ *with proof of arbitrary length.*

Proof. Given any proof π, take a random distribution $\mu_L \in \mathcal{F}_\pi$ where L is the support of μ_L. Let ν_0 be the distribution on t samples that a tester sees when first sample $\mu_L \in \mathcal{F}_\pi$, then sample t elements from μ_L.

Consider an adversarial strategy. Given proof π, the adversary chooses $\mu_L \in \mathcal{F}_\pi$; then he chooses $S \subseteq L$ with $|S| = s$. Provide the distribution $\mu_S \in \mathcal{E}_1$ to the tester. Let ν_1 be the distribution of what tester sees when sample t elements from such experiment.

It's easy to see that as long as $t \ll \sqrt{s}$, ν_0 is statistically close to ν_1. □

Note that this argument can be adapted even if we consider interactive proof rather than the PropMA model. Suppose is some interactive proof type tester Π that accepts the uniform distribution with high probability. Now for any distribution μ_S, where S is of small support. As long as the sample complexity $s \ll \sqrt{N}$ is in the non-collision regime. There is a trivial adversary strategy: Inductively, say at round t, τ_t is the transcript communicated between the verifier and the prover so far. Maybe the verifier will make a few additional sample $S_t \sim \mu$ and roll some dice R_t, and send message $m_t = m_t(\tau_t, R_1, R_2, \ldots, R_t, S_1, S_2, \ldots, S_t)$. The adversary prover receiving m_t would simply respond pretending the underlying distribution is uniform.

Corollary 6.12. *To test uniformity, the sample complexity is $\Omega(\sqrt{N})$ even in the* PropIP *model.*

7 Testing with a General Quantum Proof: Yet Another Fiasco

In Section 6, we discussed at length that for the GapSupp problem, assisted with a standard MA type proof, or even some very general structured certificates which include AM type proof, would not improve the testability unless the certificates are exponentially long.

One might ask how general such phenomenon is, if it holds for any property, and for any certificates even structured ones. Our answer is two-fold: There are both good news and bad news. We follow a convention to state the bad news first. For MA type proof, in fact, for QMA type proof, the proof would not improve the testability significantly for any property of interest. We prove that quantum proof does not increase testability, and it subsumes the classical proof. This fact follows from the de-Merlinization ideas of Aaronson [Aar06] combined with the follow-up work of Harrow et al. [HLM17].

Theorem 7.1. *Suppose there is a protocol to test property $\mathcal{P}$ using with completeness and soundness errors $1/3$, using k copies of input state $|\psi\rangle$ and a w-qubit proof. Then there is an protocol to test property $\mathcal{P}$ with the same completeness and soundness, using $O(kw\log w)$ copies of ρ and no proof. Moreover, if the original protocol has running time t, then the new protocol has running time $O(tw^w)$.*

Proof. Denote the given protocol for $\mathcal{P}$ by M. We begin by amplifying M to have exponentially-small soundness error. From [Aar06, Lemma 15] we have that there is another protocol for $\mathcal{P}$, denoted by M', with completeness error $1/3$ and soundness error $5^{-\Omega(w\log w)}$. M' uses $O(kw\log^2 w)$ copies of $|\psi\rangle$ and a $O(w\log w)$-qubit proof state $|\omega\rangle$. M' also has runtime $\widetilde{O}(tw\log w)$.

We now apply Algorithm 1 from Harrow–Lin–Montanaro [HLM17]. Let Ψ be the amplified input state on $O(nkw\log^2 w)$ qubits. If ψ is a positive instance, then there exists a proof state ω such that M' accepts $\Psi\otimes\omega$ with probability at least $2/3$. If ψ is a negative instance, then for all proof states, M' succeeds with probably at most $5^{-\Omega(w\log w)}$. Applying [HLM17, Corollary 13] with these parameters, we find Algorithm 1, using only one copy of Ψ, accepts positive instances with probability at least $4/63$ and negative instances with probability at most $o(1)$. Correctness can be suitably amplified with $O(1)$ repetitions, and Algorithm 1 runs in time $\widetilde{O}(tw\log w)O(2^{w\log w}) = O(tw^w)$. □

Because w can be of order $\mathrm{poly}(n)$, the approach presented in 7.1 may have exponential runtime. This inefficiency is actually *necessary* unless BQP = QMA. This is because QMA embeds in propQMA via the natural embedding $x \mapsto |x\rangle$ for $x\in\{0,1\}^n$.

8 Testing with Structured Quantum Certificates: A Triumph

In the previous section, we announced the bad news that general QMA proof would not improve testability for any properties significantly.

We can now announce the good news: With naturally structured quantum certificates, one can increase the testability for testing subset state of different sizes dramatically. This is in sharp contrast to the classical problem GapSupp, as in Section 6.2, we proved that structured certificates in an abstract manner as convex subset of $\Delta_{\mathcal{S}}$ gives almost no gain in terms of testability when the certificates is short.

We focus on one natural restriction: restricting the certificates to be subset states. Namely, both honest and adversary prover can only send subset state. We define the corresponding property testing model $\mathrm{PropQMA}_{\mathsf{flat}}(k, m, pm, a, b)$ analoguously to Definition 3.7.

Definition 8.1 (Quantum Property Testing with Certificates)**.** *For $d \in \mathbb{N}$, let $k = k(d), p = p(d) : \mathbb{N} \to \mathbb{N}$, $1 \geq a > b \geq 0$. A property $\mathcal{P} = \sqcup \mathcal{P}_d$ belongs to* $\mathrm{PropQMA}(k, m, mp, a, b)$ *if there exists a verifier V such that for every $|\psi\rangle \in \mathbb{C}^d$,*

(i) *if $|\psi\rangle \in \mathcal{P}_d$, then there exist m subset states $|\phi_1\rangle, \ldots, |\phi_m\rangle \in \mathbb{C}^{2^p}$ such that $V(|\psi\rangle^{\otimes k} \otimes |\phi_1\rangle \otimes \cdots \otimes |\phi_m\rangle)$ accepts with probability at least a, and*
(ii) *if $|\psi\rangle$ is ε-far from $\mathcal{P}_d$ (in trace distance), then then for any subset states $|\phi_1\rangle, \ldots, |\phi_m\rangle \in \mathbb{C}^{2^p}$, $V(|\psi\rangle^{\otimes k} \otimes |\phi_1\rangle \otimes \cdots \otimes |\phi_m\rangle)$ accepts with probability at most b.*

This section aims to prove that (adversarial) flat quantum certificates allow us to obtain a multiplicative approximation to the support size of a flat state. More precisely, we show the following.

Theorem 8.2 (Effective Quantum Certified Testing)**.** *With just polynomially many (i.e.,* $\mathrm{poly}(n)$*) copies and certificates of flat amplitudes, a polynomial time quantum tester can,*

(i) *either certify the support size of an arbitrary subset state (the target state) is s within a $(1 \pm \varepsilon)$ multiplicative factor for any constant $\varepsilon > 0$,*[9]
(ii) *or detect that the certificates are malicious.*

Note that it suffices to consider the case where there is only one copy of the target state. We show that if the target state has the correct support size s, which is part of the proof, the certificates are accepted with a probability exponentially close to 1; while if the prover tries to fool the verifier that the target state has support size s which is ε-far from the correct one, then the verifier will reject with a probability constant away from 1. Theorem 8.2 implies that with polynomial size certificates (at least restricted), the ensembles studied in Section 5 are distinguishable, using even just a single copy of the state.

For simplicity, assume that $2^n/s$ is a power of 2. Our testing strategy is simple. However, a lot of care needs to be taken to handle the adversarial situation. Therefore, we start by presenting the overall idea.

8.1 Proof Outline

Let $\rho = \phi_T$ be the target subset state with support T, for which we want to certify its support size. The prover will send classical ℓ supposedly equal to $n - \log s$. Furthermore, the prover will be asked to provide for $i = 1, 2, \ldots, \ell$ states ϕ_i supposedly equal to some subset state ϕ_{S_i}, such that

$$\begin{aligned} &T = S_\ell \subseteq S_{\ell-1} \subseteq \cdots \subseteq S_1 \subseteq S_0 = [2^n], \\ &|S_i| = 2|S_{i+1}|, && i = 0, 1, 2, \ldots, \ell - 1. \end{aligned}$$

The task reduces to testing whether indeed the support of the given states halves each time. This motivates a key technical lemma establishing that the support size of two subset states, $|\phi_H\rangle$ and $|\phi_S\rangle$, satisfies $S = |H|/2$. The intent is that $S \subset H$ and $|S| / |H| \approx 1/2$.

[9] Or even any $\varepsilon = \Omega(1/\mathrm{poly}(n))$.

Lemma 8.3 (Support Halving Lemma (informal))**.** *Suppose $|\phi_H\rangle, |\phi_S\rangle, |\phi_{S'}\rangle$ are subset states satisfying*

(i) $|\langle \phi_S \mid \phi_{S'} \rangle|^2 = \frac{1}{\mathrm{poly}(n)}$,

(ii) $\left| |\langle \phi_H \mid \phi_S \rangle|^2 - \frac{1}{2} \right| = \frac{1}{\mathrm{poly}(n)}$,

(iii) $\left| |\langle \phi_H \mid \phi_{S'} \rangle|^2 - \frac{1}{2} \right| = \frac{1}{\mathrm{poly}(n)}$.

Then, we have

$$\frac{|S|}{|H|} = \frac{1}{2} \pm \frac{1}{\mathrm{poly}(n)} .$$

It is evident how the above lemma will be used to design an algorithm to approximate the support size of a subset state $|\phi_T\rangle$. We will apply the support halving lemma iteratively. Start with $i = 0$, $S_0 = \{0,1\}^n$; for any $i \geq 0$, S_{i+1} can be any subset satisfying $T \subset S_{i+1} \subset S_i$ and

$$\frac{|S_{i+1}|}{|S_i|} = \frac{1}{2} \pm \frac{1}{\mathrm{poly}(n)}, \tag{8.1}$$

which will be guaranteed using a test that implements Lemma 8.9. Then, we obtain the telescoping multiplication

$$|S_\ell| \approx |S_0| \frac{|S_1|}{|S_0|} \frac{|S_2|}{|S_1|} \cdots \frac{|S_\ell|}{|S_{\ell-1}|} = \frac{1}{2^\ell} \left(1 \pm \frac{1}{\mathrm{poly}(n)}\right)^\ell |S_0| .$$

As long as the $1/\mathrm{poly}(n)$ is small enough, $(1 + 1/\mathrm{poly}(n))^\ell = 1 \pm \varepsilon$. Therefore $T = S_\ell$ has size about $2^{n-\ell}$.

To obtain good estimates on

$$\langle \phi_{S_i}, \phi_{S_{i+1}} \rangle, \quad \langle \phi_{S_i}, \phi_{S'_{i+1}} \rangle, \quad \langle \phi_{S_{i+1}}, \phi_{S'_{i+1}} \rangle$$

for $i = 0, 1, 2, \ldots, \ell - 1$, we will ask to prover to provide many copies of each of these subset states. In particular, the prover should supply collections of copies of states

$$\Phi_i, \Psi_i, \qquad i = 1, 2, \ldots, \ell,$$

where Φ_i corresponds to m copies of states ϕ_{S_i}, and Ψ_i corresponds to m copies of states $\phi_{S'_i}$. Supposedly, $S'_i = S_{i-1} \setminus S_i$. Then Chernoff bound tells us that with probability at most $\exp(-\Omega(m/\mathrm{poly}(n)))$, the estimate differs from the actual correlation by at most $1/\mathrm{poly}(n)$. Hence choose $m = \mathrm{poly}(n)$ would suffice.

8.2 δ-tilted States and Symmetry Test

The first challenge to face is that once we are dealing with proofs that are supposed to supply copies of the same states, we need to deal with the adversarial situation where the proofs are not the same states. Here we bring the *δ-tilted states* and *symmetry test* from [JW23, Section 4].

Definition 8.4 (δ-tilted states)**.** *A collection of states $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_k\rangle$ defined on a same space is an δ-tilted state if there is a subset $R \subseteq [k]$ such that $|R| \geq (1-\delta)k$ and for any $i, j \in R$,*

$$\mathrm{D}(|\psi_i\rangle, |\psi_j\rangle) \leq \sqrt{\delta}.$$

Furthermore, we call $|\psi_i\rangle$ a representative state *for any $i \in R$, and the subset $\{|\psi_i\rangle : i \in R\}$ the* representative set*.*

The symmetry test are used to test if the collection of states Φ provided by the prover are essentially the same, i.e., being δ-tilted, or not.

```
Algorithm 8.5: Symmetry Test
Input: $\Phi = \{\phi_1, \phi_2, \ldots, \phi_m\}$, a collection of pure states for some even number $m$.
  (i) Sample a random matching $\pi$ within $1, 2, \ldots, m$.
  (ii) SwapTest on the pairs based on the matching $\pi$.
Accept if all SwapTests accept.
```

Theorem 8.6 (Symmetry Test [JW23])**.** *Suppose Ψ is not an δ-tilted state. Then the symmetry test passes with probability at most $\exp(-\Theta(\delta^2 m))$. On the contrary, for 0-tilted state Ψ, the symmetry test accepts with probability 1.*

For a collection of states Φ, one can view it as a mixed state, or simply a set. We will take both perspectives. Next, we collect some facts about δ-tilted collection Φ. The first one states that we can combine two δ-tilted states into one in the natural way.

Proposition 8.7 (Tensorization of tilted states [JW23])**.** *If Ψ is an δ-tilted state and Φ is a γ-tilted state, and $|\Psi| = |\Phi| = m$. Then $\Psi \otimes \Phi$ is an $(\delta + \gamma)$-tilted state, where*

$$\Psi \otimes \Phi = \{\psi_i \otimes \phi_i : \Psi = \{\psi_1, \psi_2, \ldots, \psi_m\}, \Phi = \{\phi_1, \phi_2, \ldots, \phi_m\}\}.$$

The second fact states that the bahavior of δ-tilted states is like that of any representative state. This fact has two folds: If Ψ is viewed as mixed states, Ψ is close to its representative state in trace distance; if Ψ is viewed as sets, then concentration holds.

Proposition 8.8 ([JW23])**.** *For any quantum algorithm $\mathcal{A}$, let $\mathcal{A}(|\psi\rangle)$ denote the probability that $\mathcal{A}$ accepts $|\psi\rangle$. Let Ψ be an δ-tilted state, and $|\psi\rangle$ any representative state of Ψ. Then*

$$|\mathcal{A}(|\psi\rangle) - \mathcal{A}(\Psi)| \leq 3\sqrt{\delta}. \tag{8.2}$$

Furthermore, when apply $\mathcal{A}$ to Ψ, let α be the fraction of accepted executions of $\mathcal{A}$. Then

$$\Pr[|\alpha - \mathcal{A}(\Psi)| \geq \sqrt{\delta}] \leq \exp(-\delta|\Psi|/2), \tag{8.3}$$

and therefore,

$$\Pr[|\alpha - \mathcal{A}(|\psi\rangle)| \geq 4\sqrt{\delta}] \leq \exp(-\delta|\Psi|/2). \tag{8.4}$$

8.3 Subset Test

We now proceed to prove the support halving lemma, and present the subset test which will help us test (8.1).

Lemma 8.9 (Support Halving Lemma)**.** *Let $\mu \in (0,1)$ be a constant and $\delta \in (0, C\mu^4)$, where $C > 0$ is universal constant. Suppose $|\phi_H\rangle, |\phi_S\rangle, |\phi_{S'}\rangle$ are subset states satisfying*

(i) $|\langle \phi_S \,|\, \phi_{S'}\rangle|^2 \leq \delta$,

(ii) $\left| |\langle \phi_H \,|\, \phi_S\rangle|^2 - \mu \right| \leq \delta$,

(iii) $\left| |\langle \phi_H \,|\, \phi_{S'}\rangle|^2 - (1-\mu) \right| \leq \delta$.

Then, we have

$$\frac{|S|}{|H|} = (\mu \pm O(\delta^{1/4}))\,.$$

Proof. Using the first assumption, we have

$$\delta \geq |\langle \phi_S \,|\, \phi_{S'}\rangle|^2 = \frac{|S \cap S'|^2}{|S|\,|S'|}\,. \tag{8.5}$$

The second assumption states that

$$\delta \geq \left| |\langle \phi_H \,|\, \phi_S\rangle|^2 - \mu \right| = \left| \frac{|S \cap H|^2}{|S|\,|H|} - \mu \right| ,$$

in particular, it implies

$$\mu - \delta \leq |\langle \phi_H \,|\, \phi_S\rangle|^2 = \frac{|S \cap H|^2}{|S|\,|H|} \leq \min\left\{ \frac{|H|}{|S|}, \frac{|S|}{|H|}, \frac{|S \cap H|}{|H|} \right\} , \tag{8.6}$$

since $|S \cap H| \leq \min\{|S|, |H|\}$. Similarly, from the third assumption

$$\delta \geq \left| |\langle \phi_H \,|\, \phi_{S'}\rangle|^2 - \mu \right| = \left| \frac{|S' \cap H|^2}{|S'|\,|H|} - (1-\mu) \right|$$

we obtain

$$(1-\mu) - \delta \leq |\langle \phi_H \,|\, \phi_{S'}\rangle|^2 \leq \min\left\{ \frac{|S'|}{|H|}, \frac{|H|}{|S'|}, \frac{|S' \cap H|}{|H|} \right\} . \tag{8.7}$$

Using bounds from Eq. (8.6) and Eq. (8.7) in Eq. (8.5), we get

$$\frac{\delta}{(\mu-\delta)((1-\mu)-\delta)} \geq \frac{|S \cap S'|^2}{|H|^2} \geq \frac{|S \cap S' \cap H|^2}{|H|^2}\,. \tag{8.8}$$

Let $\gamma = |S \cap H| \,/\, |S|$ and $\gamma' = |S' \cap H| \,/\, |S'|$. From the second and third assumptions, we have

$$\begin{aligned}
1 \pm 2\delta &= \frac{|S \cap H|^2}{|S|\,|H|} + \frac{|S' \cap H|^2}{|S'|\,|H|} \\
&= \gamma \frac{|S \cap H|}{|H|} + \gamma' \frac{|S' \cap H|}{|H|} \\
&= \gamma \frac{|(S \setminus S') \cap H|}{|H|} + \gamma' \frac{|(S' \setminus S) \cap H|}{|H|} + (\gamma + \gamma') \frac{|S \cap S' \cap H|}{|H|} \\
&= \gamma\alpha + \gamma'\beta + (\gamma + \gamma') \frac{|S \cap S' \cap H|}{|H|} \\
&= \gamma\alpha + \gamma'\beta + O(\sqrt{\delta})\,, \qquad (8.9)
\end{aligned}$$

where the $O(\sqrt{\delta})$ bound follows from Eq. (8.8), and we set $\alpha = |(S \setminus S') \cap H| \,/\, |H|$ and $\beta = |(S' \setminus S) \cap H| \,/\, |H|$. In view of (8.8) and (8.6), we have

$$\alpha \geq \mu - O(\sqrt{\delta})$$

and similarly, using (8.8) and (8.7),

$$\beta \geq (1 - \mu) - O(\sqrt{\delta}).$$

Since $1 \geq \alpha + \beta$ and $\alpha, \beta \geq 0$, we decude that $\alpha = \mu \pm O(\sqrt{\delta})$ and $\beta = (1 - \mu) \pm O(\sqrt{\delta})$. Using (8.9) for the first equality, and since $\mu = \Omega(\delta^{1/4})$, and $\gamma, \gamma' \leq 1$,

$$1 \pm O(\sqrt{\delta}) = \gamma\alpha + \gamma'\beta = \gamma\mu + \gamma'(1 - \mu) \pm O(\sqrt{\delta}) \leq 1 - \mu(1 - \gamma) \pm O(\sqrt{\delta})\,,$$

we conclude that $\gamma \geq 1 - O(\delta^{1/4})$. From this, we get $|S \cap H| = (1 - O(\delta^{1/4}))\,|S|$. Using the second assumption, we get

$$\delta \geq \left| \frac{|S \cap H|^2}{|S|\,|H|} - \mu \right| = \left| (1 - O(\delta^{1/4}))^2 \frac{|S|}{|H|} - \mu \right| = \left| (1 - O(\delta^{1/4})) \frac{|S|}{|H|} - \mu \right|\,,$$

or

$$O(\delta) \geq \left| \frac{|S|}{|H|} - (1 \pm O(\delta^{1/4}))\mu \right|\,,$$

as desired. □

Below is a formal description of the subset test that incorporates our setting and implements Lemma 8.9 as a test for (8.1). The parameter γ will be specified later.

```
Algorithm 8.10: SubSet Test
Input: Collections of states $\Phi_1, \Phi_2, \Psi_2$, and a target density some constant $\mu \in (0,1)$.
Supposedly, the three collections corresponds to some subset state $\phi_H, \phi_S, \phi_{S'}$, respectively.
Take the following steps:
  (i) Partition the each of collections into two parts of equal size:
          $\Phi_1 = \Phi_1' \sqcup \Phi_1'', \quad \Phi_2 = \Phi_2' \sqcup \Phi_2'', \quad \Psi_2 = \Psi_2' \sqcup \Psi_2''.$
  (ii) Estimate $|\langle \phi_H \mid \phi_S \rangle|^2$: Applying $m/2$ swap tests on $\{\Phi_1'\} \otimes \{\Phi_2'\}$. Let the fraction
       of accepted pairs be $\alpha$.
  (iii) Estimate $|\langle \phi_H \mid \phi_{S'} \rangle|^2$: Applying $m/2$ swap tests on $\{\Phi_1''\} \otimes \{\Psi_2'\}$. Let the fraction
       of accepted pairs be $\beta$.
  (iv) Estimate $|\langle \phi_S \mid \phi_{S'} \rangle|^2$: Applying $m/2$ swap tests on $\{\Phi_2''\} \otimes \{\Psi_2''\}$. Let the fraction
       of accepted pairs be $\zeta$.
Accept if all the inequalities hold: $|(2\alpha - 1) - \mu| \leq \gamma$; $|(2\beta - 1) - (1 - \mu)| \leq \gamma$; $|2\zeta - 1| \leq \gamma$.
```

8.4 Subset Support Certification Algorithm and Analysis

Now we present a formal description of the algorithm that used to certify support size of a given target state. Set the parameters:

ε, the estimate error tolerance parameter in Theorem 8.2, $\leq 1/2$.
δ, the symmetry test error tolerance parameter as used in Lemma 8.9, $= \varepsilon^{16}/(320^2 n^8)$.
γ, the subset test error tolerance parameter, $= \varepsilon^8/(80n^4)$.
m, the size of each collections Φ_i, Ψ_i, $= O(n^{16}/\varepsilon^{32})$.

```
Algorithm 8.11: Subset State Support Test
Input:$\rho, \Phi_0, \Phi_1, \Psi_1, \Phi_2, \Psi_2 \ldots \Phi_\ell, \Psi_\ell$
Apply one of the following tests:
  (i) Symmetry Test on $(\Phi_i, \Psi_i)$, for all $i$;
  (ii) (Even) Subset Test on $(\Phi_{2i}, \Phi_{2i+1}, \Psi_{2i+1})$, for all $0 < i < \ell/2$;
  (iii) (Odd) Subset Test on $(\Phi_{2i+1}, \Phi_{2i+2}, \Psi_{2i+2})$, for all $0 \leq i < \ell/2$;
  (iv) Swap Test on $\rho$ and a random state $\phi \in \Phi_\ell$.
Accept if the chosen test accepts.
```

Proof of Theorem 8.2. Now we show the above Subset State Support Test satisfies Theorem 8.2. The completeness is straightforward. The (i) symmetry test and (iv) swap test will pass with probability 1. In (ii) and (iii), each Subset Test will passes will probability $1 - \exp(-\Omega(\gamma^2 m))$. Overall, the test pass with probability $1 - \ell \exp(-\Omega(\gamma^2 m)) = 1 - \exp(-\Omega(n^8))$, certifying the subset state has size $N/2^\ell$.

In the adversarial case, we consider the possible ways that the adversary may cheat and show eventually the testing will catch these cheats.

Case 1: Attack caught by (i) Symmetry Test. If any collection Φ_i, Ψ_i of the states is not δ-tilted, then by Theorem 8.6, with probability at most $\exp(-\Omega(\delta^2 m)) = \exp(-\Omega(1))$, it passes the symmetry test. From now on, assume that all the collections are δ-tilted.

Case 2: Attack caught by (iv) Swap Test. Consider Φ_ℓ, take any representative state

$\phi \in \Phi_\ell$. Suppose that ϕ corresponds to some subset state ϕ_S with $|S| \notin (1 \pm \varepsilon^2)|T|$, then

$$|\langle \phi_S \mid \rho\rangle|^2 \le 1 - \varepsilon^2 \;\Rightarrow\; \Pr[\text{Swap Test accepts } (\phi_S, \rho)] \le 1 - \frac{\varepsilon^2}{2}.$$

Suppose ϕ is a representative state. Then by definition with probability at least $1-\delta$, a random state ϕ' in Φ_ℓ satisfies $\mathrm{D}(\phi', \phi_S) \le \sqrt{\delta}$, or in other words $|\langle \phi_S, \phi'\rangle|^2 \ge 1-\delta$. Therefore, ϕ' corresponds to a subset state of size $\notin (1\pm\varepsilon^2)(1\pm\delta)|T|$. Hence the probability that any other representative state passes the swap test is at most $\frac{1+(1-\varepsilon^2)(1-\delta)}{2}$. We can conclude, with probability at least

$$(1-\delta)\cdot\left(1 - \frac{1+(1-\varepsilon^2)(1-\delta)}{2}\right) \ge \frac{\varepsilon^2}{2} - \delta,$$

the swap test rejects. Consequently, from now on we further assume that all the representative states in Φ_ℓ corresponds to a subset state of support size $s \in (1\pm\varepsilon^2)|T|$, as otherwise the accepting probability will be a constant away from 1.

Case 3: Attack caught by (ii)-(iii) Subset Test. Pick some arbitrary representative state ϕ_i of Φ_i for $i = 0, 1, \ldots, \ell$, let s_i be the support size for each ϕ_i, then

$$s_0 \cdot \frac{s_1}{s_0} \cdot \frac{s_2}{s_1} \cdots \frac{s_\ell}{s_{\ell-1}} = s_\ell \in (1\pm\varepsilon^2)|T|. \tag{8.10}$$

To cheat, the adversary can tell a wrong estimate of $|T|$, meaning $2^{-\ell}|N| \notin (1\pm\varepsilon)|T|$.

Claim 8.12. *If the adversary tells a wrong estimate of $|T|$. Then for some $i \ge 1$, either,*

$$\frac{s_i}{s_{i-1}} \ge \frac{1}{2} + \frac{\ln(1+\varepsilon^2)}{\ell},$$

or,

$$\frac{s_i}{s_{i-1}} \le \frac{1}{2} - \frac{\varepsilon^2}{2\ell}.$$

Proof. For the purpose of contradiction, suppose the claim is false. That is for all i, the fraction between s_i and s_{i-1} is very close to $1/2$. Consider two possible situation, first, if $2^{-\ell}N > (1+\varepsilon)|T|$, then

$$s_\ell \ge \left(\frac{1}{2} - \frac{\varepsilon^2}{2\ell}\right)^\ell s_0 \ge (1-\varepsilon^2)(1+\varepsilon)2^{-\ell}s_0 > (1-\varepsilon^2)(1+\varepsilon)|T| > (1+\varepsilon^2)|T|.$$

This contradicts (8.10). Second, if $2^{-\ell}N < (1-\varepsilon)|T|$, then

$$s_\ell \le \left(\frac{1}{2} + \frac{\ln(1+\varepsilon^2)}{2\ell}\right)^\ell s_0 \le 2^{-\ell}s_0 \exp(\ln(1+\varepsilon^2)) = (1+\varepsilon^2)2^{-\ell}s_0 < (1-\varepsilon^2)|T|,$$

again, contradicting (8.10). □

W.l.o.g., say $i = 1$ is an index satisfying the above claim. We show that the (odd) subset test rejects with high probability. In the subset test, each collection is partitioned into two parts of equal size,

$$\Phi_1 = \Phi_1' \sqcup \Phi_1'', \quad \Phi_2 = \Phi_2' \sqcup \Phi_2'', \quad \Psi_2 = \Psi_2' \sqcup \Psi_2''.$$

By definition, each part will be a 2δ-tilted state. Furthermore, by Proposition 8.7 they form three collections of 4δ-tilted states,

$$\Gamma_0 := \Phi_1' \otimes \Phi_2', \quad \Gamma_1 := \Phi_1'' \otimes \Psi_2', \quad \Gamma_2 := \Phi_2'' \otimes \Psi_2''.$$

In subset test, $\Gamma_0, \Gamma_1, \Gamma_2$ will be fed to swap test and estimate s_2/s_1. By Claim 8.12,

$$\frac{s_2}{s_1} \notin \frac{1}{2} \pm \frac{\varepsilon^2}{2\ell}.$$

Let $\kappa = \Theta(\varepsilon^2/(2\ell))$, then one of the following must be true by Lemma 8.9,

(i) $|\langle \phi_2, \psi_2 \rangle|^2 \geq \kappa^4$,
(ii) $||\langle \phi_1, \phi_2 \rangle|^2 - \frac{1}{2}| \geq \kappa^4$,
(iii) $||\langle \phi_1, \psi_2 \rangle|^2 - \frac{1}{2}| \geq \kappa^4$.

Without loss of generality say (i) hold. Then

$$\left| \Pr[\text{SwapTest}(\phi_1, \phi_2) \text{ accept}] - \frac{1}{2} \right| \geq \kappa^4/2. \tag{8.11}$$

By Proposition 8.8, the estimate ζ from Γ_0 will be

$$\begin{aligned} &\Pr_{\zeta}\left[|\zeta - \Pr[\text{SwapTest}(\phi_1, \phi_2) \text{ accept}|] \geq 4\sqrt{4\delta} \right] = \exp(-\Omega(\delta m)) \\ \overset{(8.11)}{\Longrightarrow} \quad &\Pr_{\zeta}\left[|2\zeta - 1| \leq \kappa^4 - 16\sqrt{\delta} \right] = \exp(-\Omega(\delta m)). \end{aligned} \tag{8.12}$$

Choose suitable parameters that satisfy,

$$\gamma \leq \kappa^4 - 16\sqrt{\delta}, \ \kappa^4 \geq 17\sqrt{\delta}.$$

Hence, the probability that the subset test accepts is $\exp(-\Omega(\delta m)) = \exp(-\Omega(n^8))$. □

8.5 Discussion

Let us review some of our indistinguishability results in Section 5 and Section 6, we want to emphasize some remarkable perspectives of our upper bound result.

First in Section 5, we pointed out that there is no tester with any advantage for testing productness with a single copy of a given state $|\psi\rangle$. What if proofs are allowed? It is immediate that the honest prover can give another copy of $|\psi\rangle$ as the proof, then by product test [HM13], the verifier can distinguish product states and those far from being product. The unsatisfying feature is that the role of the proof is very limited, it serves as just another copy of the state. Quantitatively, if we count the total number of resources used in the algorithm, i.e., the proof complexity plus the given state $|\psi\rangle$, then proofs gain us nothing! Because with two copies of the state, one can carry out product test anyways. Therefore the more interesting question would be to demonstrate properties for which without proof the copy complexity is super-polynomial but with polynomial-size proofs the copy complexity becomes polynomial. This is what our example illustrates.

Second, think about the lower bound that we presented in Section 6 for distinguishing the flat distribution of support size s and $2s$. That would be the classical counterpart of the quantum property for which we demonstrate the power of proofs. However, in this classical setting, the power of proofs is completely gone! To see this note that $\Theta(\sqrt{s})$ samples are necessary and sufficient for distinguishing flat distribution of support size either s or $2s$. With certificates, on the other hand, our lower bound in Theorem 6.2 shows that only when the certificate length is $\Omega(\sqrt{s})$, the certificate can reduce the sample complexity. However, the total resources needed, i.e., certificate length plus sample complexity is $\Omega(\sqrt{s})$, match that without certificates. In fact, recall that in Remark 6.3, the optimal proof strategy is to send some extra samples.

Finally, for the "productness" example in Section 5, the best prover strategy would make proof an additional copy, not really providing any extra power; for the subset state, with flat certificates we can estimate the support size using exponentially smaller amount of resources. What about the other ensembles with different support size. It is not hard to see that a proof helps using analogous strategy: Ask the prover to provide certificates that will be subset state of size $\approx pd$, which as we have seen is testable with with flat certificate. In fact, in the dense regime, the flat certificate can be further relaxed to nonnegative amplitudes certificates [JW23] using their sparsity test.

8.6 Lower Bounds for Quantum-to-Quantum State Transformation

We will now discuss how the study of quantum property testing protocols even under very strong assumptions on the structure of the proofs can have interesting consequences for quantum-to-quantum state transformation. To this end, we will suppose that we have obtained the following results for some quantum property $\mathcal{P}$.

(i) We managed to design a tester with quantum proofs for property $\mathcal{P}$ using "few" copies of the input state, but assuming that the proofs satisfy the strong promise of being a chosen function of the state being tested.

(ii) We also managed to show that property $\mathcal{P}$ requires "many" copies to be tested (using only copies of the state to be tested).

Now we can consider a quantum-to-quantum transformation that takes a certain number of copies of a state and produces a state that is (close to) the chosen function (mentioned above) of the input states. Note that combining this hypothetical transformation with the tester for $\mathcal{P}$ with the promised structured proofs (from the first item above) yields a tester (using only copies the input state) for property $\mathcal{P}$. We illustrate this scenario in Fig. 2. By appealing to the second item above, we would deduce that "many" copies are needed to implement this quantum-to-quantum transformation. Curiously, these considerations also illustrate that the study of quantum-to-classical results can have implications for quantum-to-quantum results.

For instance, using the above template, we can deduce the following quantum-to-quantum transformation lower bounds. The first about transforming the amplitudes of a quantum state into their absolute values.

Theorem 8.13 (Hardness of Absolute Amplitudes Transformation (Informal)). *Any transformation that takes k copies of an arbitrary n-qubit quantum state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x|x\rangle$ and produces a signle n-qubit output state at least 0.001 close to $\sum_{x\in\{0,1\}^n} |\alpha_x|\,|x\rangle$ requires $k = 2^{\Omega(n)}$.*

The second transformation lower bound is for mapping the amplitudes to their complex conjugates.

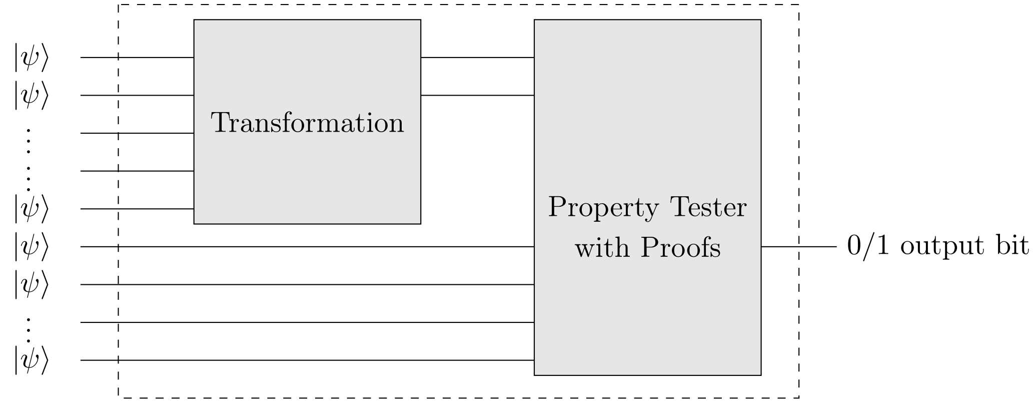


Figure 2: A pictorial representation of combining a quantum-to-quantum state transformation that generates suitable proofs with a property tester promised to receive copies of the input quantum state, as well as, these suitable (structured) proofs. This combination yields a property tester (depicted as the dashed enclosing box) using only copies of the input state and no proofs.

Theorem 8.14 (Hardness of Amplitude Conjugation Transformation (Informal))**.** *Any transformation that takes k copies of an arbitrary n-qubit quantum state $|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x|x\rangle$ and produces a signle n-qubit output state at least* 0.001 *close to $\sum_{x\in\{0,1\}^n} \alpha_x^*|x\rangle$ requires $k = 2^{\Omega(n)}$.*

9 Property Testing Complexity Classes and Hierarchy

Now we take the opportunity to define some obvious property testing complexity classes regarding the information theoretic sample/copy complexity. Consider some property $\mathcal{P} = \sqcup \mathcal{P}_N$, where $\mathcal{P}_N$ can be a subset of Δ_N or $\mathfrak{S}(\mathbb{C}^N)$.

Definition 9.1 (Property Testing Complexity Class)**.** *Let $n = \log(N)$, fix some arbitrary constant $1 > c > s > 0$,*

$$\begin{aligned}
\mathrm{Prop}\mathcal{C} &:= \mathrm{Prop}\mathcal{C}\left(\mathrm{poly}(n), c, s\right), && C \in \{\mathrm{BPP}, \mathrm{BQP}\};\\
\mathrm{PropMA} &:= \mathrm{PropMA}\left(\mathrm{poly}(n), \mathrm{poly}(n), c, s\right);\\
\mathrm{PropMA}_{\mathrm{exp}} &:= \mathrm{PropMA}_{\mathrm{exp}}\left(\mathrm{poly}(n), 2^{\mathrm{poly}(n)}, c, s\right);\\
\mathrm{PropEXP} &:= \mathrm{PropEXP}\left(2^{\mathrm{poly}(n)}, c, s\right);\\
\mathrm{PropQMA}(k) &:= \mathrm{PropQMA}(\mathrm{poly}(n), k, \mathrm{poly}(n) c, s);\\
\mathrm{PropQMA} &:= \mathrm{PropQMA}(1);\\
\mathrm{PropAM}[r] &:= \mathrm{PropAM}[r](\mathrm{poly}(n), 1, \mathrm{poly}(n), c, s);\\
\mathrm{PropIP} &:= \mathrm{PropIP}(\mathrm{poly}(n), \mathrm{poly}(n), c, s);\\
&\ \ \vdots
\end{aligned}$$

Note we did not make any distinction between testing classical distribution or quantum states. First, in the classical model PropBPP, it does not make sense to test quantum property. So one can just view classical distribution as degenerated quantum states. Second, it is normally very clear from the context, the problem of interest is to test classical distribution or to test quantum states, thus there is no need to make a different set of names. This also justifies the choice that the complexity is with respect to $n = \log(N)$ where n is the size of the discrete probability space in the case of property testing for classical distributions, and dimension in the case of quantum states.

Finally, the notations PropBPP, PropQMA, etc. are used for both the property testing models as well as the property testing complexity classes. This is a somewhat common abuse of notation. To give an alert ahead for the unfamiliar readers, consider the following two statement, for some property $\mathcal{P}$,

(i) $\mathcal{P} \in \mathrm{PropQMA}$,

(ii) $\mathrm{PropQMA}(\mathcal{P}) = \exp(\Omega(n))$.

In the first case, PropQMA is a complexity class. $\mathcal{P} \in \mathrm{PropQMA}$ is an upper bound result, i.e., $\mathcal{P}$ can be tested using $\mathrm{poly}(n)$ copies assisted with a QMA type prover. On the other hand, in the second case, PropQMA is the property testing model, and $\mathrm{PropQMA}(\mathcal{P}) = \exp(\Omega(n))$ is a lower bound result, meaning that in the PropQMA model, one needs $\exp(\Omega(n))$ copies to test $\mathcal{P}$.

Along our study of the concrete properties in the later sections, we will elaborate more on the relationship between different property testing complexity class. Here, we collect a few simple observations.

Proposition 9.2. *For both classical distribution and quantum state properties*

$$\mathrm{PropEXP} = \mathrm{ALL}. \tag{9.1}$$

Proof. The statement holds for both quantum and classical properties, because exponentially many properties are sufficient for learning to quantum states and classical distribution [OW16, HHJ$^+$16, dlVKM07]. □

This proposition justifies our definition of the property testing class $\mathrm{PropMA}_{\mathrm{exp}}$, where the number of samples/copies is polynomial instead of exponential. It turns out that $\mathrm{PropMA}_{\mathrm{exp}}$ is also a trivial upper bounds for any quantum state properties.

Proposition 9.3. *For quantum state properties*

$$\mathrm{PropMA}_{\mathrm{exp}} = \mathrm{ALL}. \tag{9.2}$$

Proof. The statement holds because the the prover can send a classical description of the state $|\psi\rangle$ to test. Then the verifier can prepare a state $|\phi\rangle$ based on the classical description and use swap test to check if the classical description is the correct. In particular, set $\delta = 1/\exp(\mathrm{poly}(d))$. Estimate the overlap $|\langle\phi \mid \psi\rangle|^2$.

For $|\psi\rangle \in \mathcal{P}$, the honest prover sends a correct description of $|\psi\rangle$ up to the precision $\delta = 1/\exp(\mathrm{poly}(d))$, thus $|\langle\psi \mid \phi\rangle|^2 \geq 1-\delta$. Therefore, $|\phi\rangle$ is δ close to $\mathcal{P}$. Given $k = \mathrm{polylog}(d)$ many copy of $|\psi\rangle$, the probability that all k swap test passes is

$$(1 - O(\delta))^k = 1 - 1/\exp(\mathrm{poly}(d)).$$

On the other hand, for $|\psi\rangle$ ε-far from $\mathcal{P}$, if the verifier lies by giving some $|\phi\rangle$ that is δ close to $\mathcal{P}$, then $|\langle\psi \mid \phi\rangle|^2 \leq 1-\varepsilon+o(\varepsilon)$. Therefore all the swap test passes with probability at most

$$(1 - \varepsilon + o(\varepsilon))^k = \exp(-\varepsilon k),$$

which is tiny for any constant ε. □

A natural question is whether the above proposition is true for classical distribution properties. As we see in Theorem 6.2, it is not. Therefore, this is another example of how quantum coherence enlarges the testability.

In terms of the proof system, the seminar work of Goldwasser and Sipser proved a surprising result $\mathrm{IP} = \mathrm{AM}[\mathrm{poly}(n)]$ [GS86], i.e., public-coin interactive proof system is as powerful as the private-coin proof system. However, in view of Theorem 6.9 and Corollary 6.8, private-coin is significantly more powerful in property testing for both classical distribution and quantum states,

$$\mathrm{PropAM}[\mathrm{poly}(n)] \subsetneq \mathrm{PropIP}. \tag{9.3}$$

Finally, Theorem 7.1 implies that

$$\mathrm{PropQMA} = \mathrm{PropBQP}. \tag{9.4}$$

$$\mathrm{PropMA} = \mathrm{PropBPP}. \tag{9.5}$$

9.1 Information Theoretic versus Computation Constrained Models

The study of property testing can be broadly divided into two main categories: information theoretic and computation constrained testing. In the former category, no computation assumption is made about the tester other than its existence (even allowing it to be non-uniform). In the latter category, we impose that a tester has to be generated uniformly by a Turing machine, and it has to obey the computation resource constraints of the corresponding complexity class (e.g., in this model a tester for PropQMA is required to be a BQP verifier). In particular, these models can capture the following behavior regarding property testing and computation complexity.

(i) Information theoretic: captures the inherent limitations imposed by quantum/classical information theory regardless of any computation limitations on a tester.

(ii) computation constrained with

- quantum input states: captures decision problems with quantum inputs under resource constraints.
- classical input states: captures standard complexity classes.

Remark 9.4. *Any language or promise problem in a complexity (or computability) class with classical inputs gives rise to two disjoint collection of bit strings L_{yes} and L_{no} consisting in yes and no instances, respectively. We remark that the information theoretic version of the class* PropBPP *and* PropBQP *trivially capture them.*

Remark 9.5. *By considering classical input states in the complexity constrained models of property testing, we can ask whether a classical bit string (given as input state to be tested) is a yes or no instance of a language or promise problem. Therefore, these models capture standard complexity classes. Under the assumption* $\mathrm{BQP} \neq \mathrm{QMA}$, *we have* $\mathrm{PropBQP} \neq \mathrm{PropQMA}$ *for their computation constrained models.*

Remark 9.6. *In contrast, for the information theoretic models, we have the collapse* $\mathrm{PropBQP} = \mathrm{PropQMA}$. *This means that a general quantum proof cannot substantially improve information theoretic testability of quantum properties (they can at best reduce polynomially the number of copies of the input state, or improve the efficiency of the tester).*

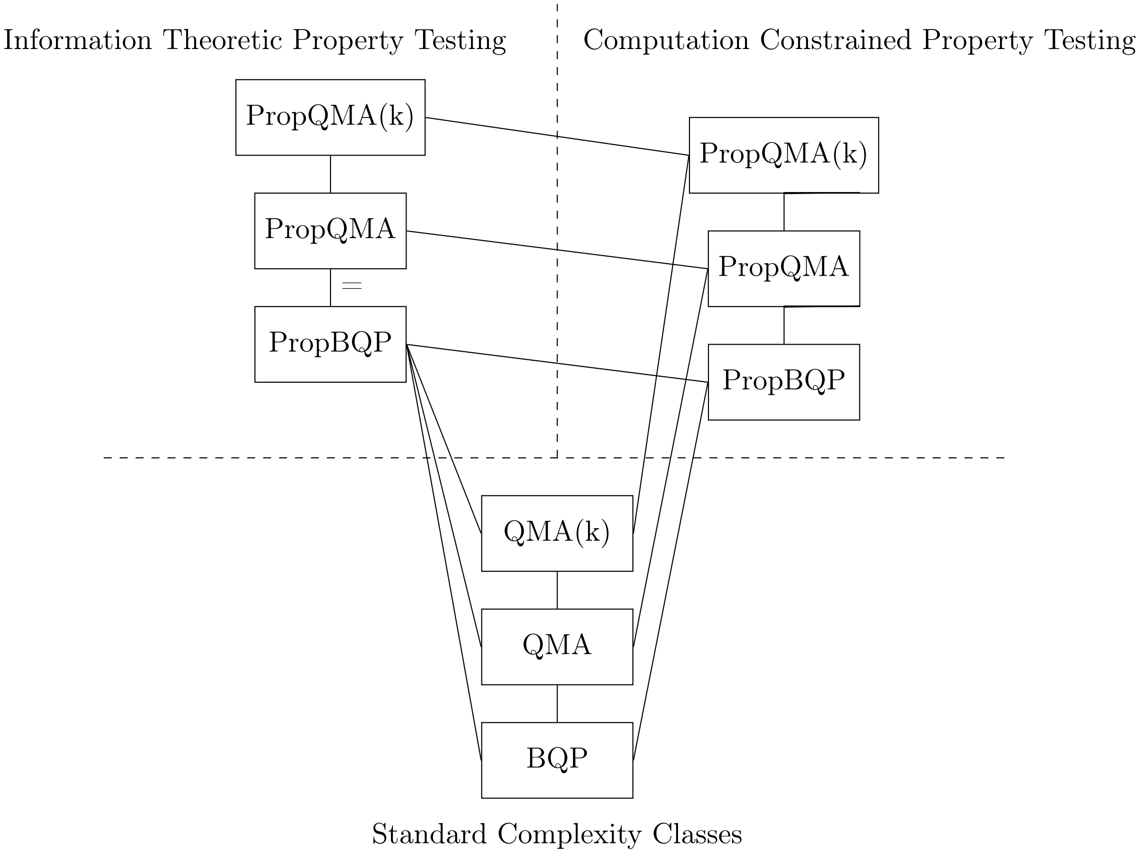

Information Theoretic Property Testing
Computation Constrained Property Testing
PropQMA(k)
PropQMA
=
PropBQP
PropQMA(k)
PropQMA
PropBQP
QMA(k)
QMA
BQP
Standard Complexity Classes

Figure 3: We depict the relationships among property testing both in the information theoretic models (on the upper left), the computation constrained models (on the upper right), and the standard complexity classes (on the bottom) for the case of BQP, QMA, and QMA(k). Line segments from bottom to top indicate containments (i.e., the model on top can test at least all the properties its connecting bottom model can).

We summarize the above remarks in Fig. 3.

9.2 Summary of Our Information Theoretic in terms of Property Testing Classes

In Fig. 4, we provide a visual summary of some of our information theoretic results for support size Theorems 1.2, 1.4 and 1.6 in terms of the property testing classes.

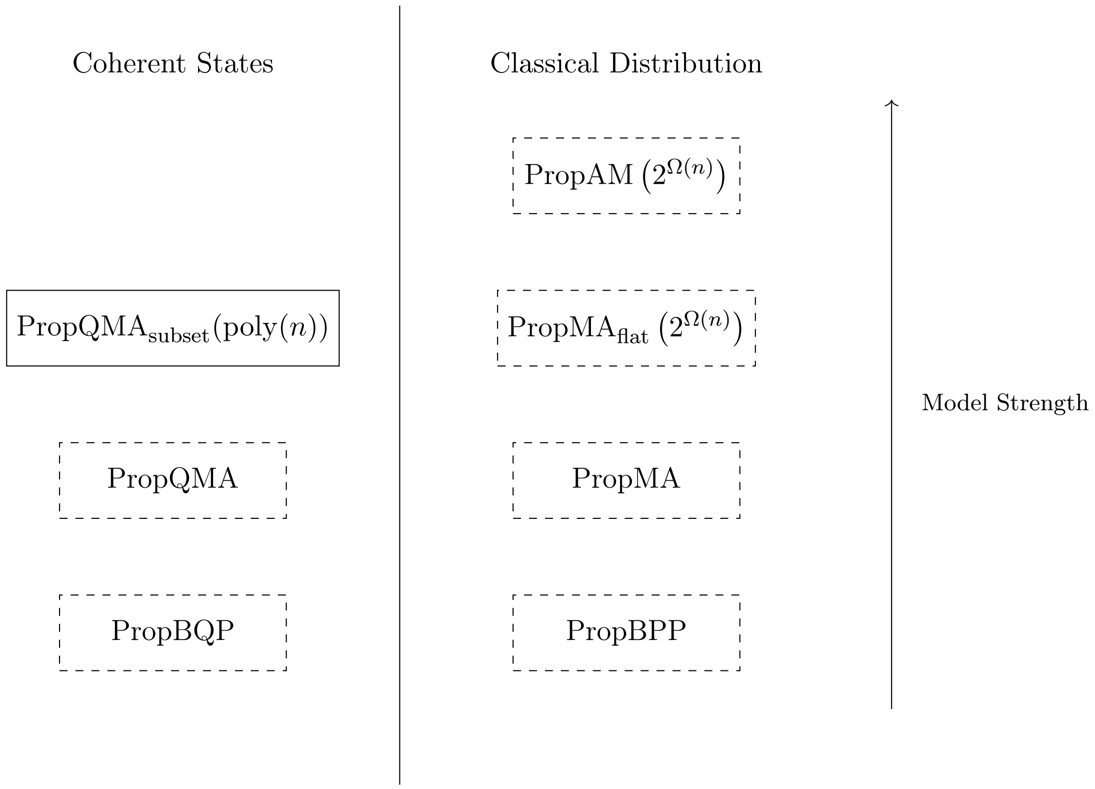

Coherent States
Classical Distribution
$\mathsf{PropAM}\left(2^{\Omega(n)}\right)$
$\mathsf{PropQMA}_{\mathsf{subset}}(\mathrm{poly}(n))$
$\mathsf{PropMA}_{\mathsf{flat}}\left(2^{\Omega(n)}\right)$
Model Strength
PropQMA
PropMA
PropBQP
PropBPP

Figure 4: A pictorial representation of the limitations in distinguishing support size of flat coherent quantum states (depicted on the left column) and flat classical distributions (on the right column). Dashed boxes indicate that the model fails in this task whereas a solid box indicates that the model succeeds.

References

[AA18] Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM Journal on Computing*, 2018. 4

[Aar06] Scott Aaronson. QMA/qpoly $\subseteq$ PSPACE/poly: de-merlinizing quantum protocols. In *Proceedings of the 21st IEEE Conference on Computational Complexity (CCC)*, 2006. 3, 29

[ABF$^+$24] Scott Aaronson, Adam Bouland, Bill Fefferman, Soumik Ghosh, Umesh Vazirani, Chenyi Zhang, and Zixin Zhou. Quantum pseudoentanglement. In *Proceedings of the 15th Innovations in Theoretical Computer Science Conference (ITCS)*, 2024. 2, 5

[ABK$^+$21] Scott Aaronson, Shalev Ben-David, Robin Kothari, Shravas Rao, and Avishay Tal. Degree vs. approximate degree and quantum implications of huang's sensitivity theorem. In Samir Khuller and Virginia Vassilevska Williams, editors, *Proceedings of the 53rd ACM Symposium on Theory of Computing (STOC)*, pages 1330–1342. ACM, 2021. 4

[AGQY22] Prabhanjan Ananth, Aditya Gulati, Luowen Qian, and Henry Yuen. Pseudorandom (function-like) quantum state generators: New definitions and applications. In *Proceedings of the Theory of Cryptography Conference (TCC)*, pages 237–265. Springer, 2022. 5

[AJK$^+$22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th ACM Symposium on Theory of Computing (STOC)*, 2022. 2, 24, 25

[AKKT20] Scott Aaronson, Robin Kothari, William Kretschmer, and Justin Thaler. Quantum lower bounds for approximate counting via laurent polynomials. In *Proceedings of the 35th IEEE Conference on Computational Complexity (CCC)*, 2020. 1, 4

[Amb00] Andris Ambainis. Quantum lower bounds by quantum arguments. In *Proceedings of the 32nd ACM Symposium on Theory of Computing (STOC)*, 2000. 4

[BBC$^+$01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 2001. 4

[BBSS23] Amit Behera, Zvika Brakerski, Or Sattath, and Omri Shmueli. Pseudorandomness with proof of destruction and applications. *Cryptology ePrint Archive*, 2023. 5

[BCQ23] Zvika Brakerski, Ran Canetti, and Luowen Qian. On the computational hardness needed for quantum cryptography. In *Proceedings of the 14th Innovations in Theoretical Computer Science Conference (ITCS)*, 2023. 5

[BDKR02] Tuundefinedkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating entropy. In *Proceedings of the 34th ACM Symposium on Theory of Computing (STOC)*, 2002. 1

[BFF$^+$01] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White. Testing random variables for independence and identity. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, 2001. 1

[BR20] Aleksandrs Belovs and Ansis Rosmanis. Tight quantum lower bound for approximate counting with quantum states, 2020. 4

[BS19] Zvika Brakerski and Omri Shmueli. (pseudo) random quantum states with binary phase. In *Proceedings of the Theory of Cryptography Conference (TCC)*, pages 229–250. Springer, 2019. 5

[BS20] Zvika Brakerski and Omri Shmueli. Scalable pseudorandom quantum states. In *Proceedings of the 40th Annual International Cryptology Conference (CRYPTO)*, pages 417–440. Springer, 2020. 5

[Can20] Clément L. Canonne. *A Survey on Distribution Testing: Your Data is Big. But is it Blue?* Number 9 in Graduate Surveys. Theory of Computing Library, 2020. 1, 4

[Can22] Clément L. Canonne. *Topics and Techniques in Distribution Testing.* Publisher: Now Foundations and Trends, 2022. 1, 4

[CG18] Alessandro Chiesa and Tom Gur. Proofs of Proximity for Distribution Testing. In *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS)*, 2018. 4

[CGM19] M. Cryan, H. Guo, and G. Mousa. Modified log-sobolev inequalities for strongly log-concave distributions. In *Proceedings of the 60th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2019. 24

[Del73] Philippe Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Res. Rep. Suppl.*, 10:vi+–97, 1973. 17, 49

[Del75] Philippe Delsarte. The association schemes of coding theory. In *Combinatorics: Proceedings of the NATO Advanced Study Institute held at Nijenrode Castle, Breukelen, The Netherlands 8–20 July 1974.* Springer, 1975. 2

[DEL$^+$22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In *Proceedings of the 54th ACM Symposium on Theory of Computing (STOC)*, 2022. 1

[DGRMT22] Marcel Dall'Agnol, Tom Gur, Subhayan Roy Moulik, and Justin Thaler. Quantum Proofs of Proximity. *Quantum*, October 2022. 4

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12–es, jun 2007. 1

[dlVKM07] Wenceslas Fernandez de la Vega and Claire Kenyon-Mathieu. Linear programming relaxations of maxcut. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 53–61, 2007. 40

[Gol17] Oded Goldreich. *Introduction to Property Testing.* Cambridge University Press, 2017. 1, 4

[GS86] S. Goldwasser and M. Sipser. Private coins versus public coins in interactive proof systems. In *Proceedings of the 18th ACM Symposium on Theory of Computing (STOC)*, pages 59–68, 1986. 41

[GTB23] Tudor Giurgica-Tiron and Adam Bouland. Pseudorandomness from subset states. *arXiv preprint arXiv:2312.09206*, 2023. 5

[Har13] Aram W Harrow. The church of the symmetric subspace. *arXiv preprint arXiv:1308.6595*, 2013. 15

[HHJ+16] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC)*, pages 913–925, 2016. 40

[HLM17] Aram W. Harrow, Cedric Yen-Yu Lin, and Ashley Montanaro. Sequential measurements, disturbance and property testing. In Philip N. Klein, editor, *Proceedings of the 28th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2017. 3, 29

[HLS07] Peter Hoyer, Troy Lee, and Robert Spalek. Negative weights make adversaries stronger. In *Proceedings of the 39th ACM Symposium on Theory of Computing (STOC)*, 2007. 4

[HM13] Aram W. Harrow and Ashley Montanaro. Testing product states, quantum merlin-arthur games and tensor optimization. *J. ACM*, 60(1), feb 2013. 14, 37

[HR22] Tal Herman and Guy N. Rothblum. Verifying the unseen: interactive proofs for label-invariant distribution properties. In Stefano Leonardi and Anupam Gupta, editors, *Proceedings of the 54th ACM Symposium on Theory of Computing (STOC)*, pages 1208–1219. ACM, 2022. 1, 4, 27

[HR23] Tal Herman and Guy N. Rothblum. Doubley-efficient interactive proofs for distribution properties. pages 743–751. IEEE, 2023. 4

[JLS18] Zhengfeng Ji, Yi-Kai Liu, and Fang Song. Pseudorandom quantum states. In *Proceedings of the 38th Annual International Cryptology Conference (CRYPTO)*, pages 126–152. Springer, 2018. 2, 5

[JW23] Fernando Granha Jeronimo and Pei Wu. The Power of Unentangled Quantum Proofs with Non-negative Amplitudes. In *Proceedings of the 55th ACM Symposium on Theory of Computing (STOC)*, 2023. 31, 32, 38

[KQST23] William Kretschmer, Luowen Qian, Makrand Sinha, and Avishay Tal. Quantum cryptography in algorithmica. In *Proceedings of the 55th ACM Symposium on Theory of Computing (STOC)*, 2023. 2, 5

[Kre21] William Kretschmer. Quantum Pseudorandomness and Classical Complexity. *Leibniz Int. Proc. Inf.*, 2021. 2

[Mac30] Colin MacLaurin. Iv. a second letter from mr. colin mclaurin. *Philosophical Transactions of the Royal Society of London*, 36(408), 1730. 48

[MW16] Ashley Montanaro and Ronald de Wolf. *A Survey of Quantum Property Testing*. Number 7 in Graduate Surveys. Theory of Computing Library, 2016. 1, 4

[MY22a] Tomoyuki Morimae and Takashi Yamakawa. One-wayness in quantum cryptography. *arXiv preprint arXiv:2210.03394*, 2022. 5

[MY22b] Tomoyuki Morimae and Takashi Yamakawa. Quantum commitments and signatures without one-way functions. In *Proceedings of the 42nd Annual International Cryptology Conference (CRYPTO)*, pages 269–295. Springer, 2022. 5

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the 48th ACM Symposium on Theory of Computing (STOC)*, pages 899–912, 2016. 40

[PK22] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical ldpc codes. In *Proceedings of the 54th ACM Symposium on Theory of Computing (STOC)*, 2022. 1

[R$^+$10] Dana Ron et al. *Algorithmic and analysis techniques in property testing*. Now Publishers, Inc., 2010. 1, 4

[RRSS07] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007. 1

[Rub12] Ronitt Rubinfeld. Taming big probability distributions. *XRDS*, 2012. 1, 4

[S$^+$77] Jean-Pierre Serre et al. *Linear representations of finite groups*, volume 42. Springer, 1977. 14

[SV14] Mohit Singh and Nisheeth K Vishnoi. Entropy, optimization and counting. In *Proceedings of the 46th ACM Symposium on Theory of Computing (STOC)*, pages 50–59, 2014. 48

[Val11] Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 2011. 1, 4

[VV11] Gregory Valiant and Paul Valiant. Estimating the unseen: an n/log(n)-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, 2011. 1, 4

[VV17] Gregory Valiant and Paul Valiant. Estimating the unseen: Improved estimators for entropy and other properties. *J. ACM*, 2017. 1

[VW16] Thomas Vidick and John Watrous. Quantum proofs. *Foundations and Trends® in Theoretical Computer Science*, 2016. 1

[Weg24] Jordi Weggemans. Lower bounds for unitary property testing with proofs and advice, 2024. 4

A Divergence Contraction

In this section, we finish the proof of Lemma 6.4. It remains to establish Claim 6.5. Before we prove Claim 6.5, we need to introduce some definitions and recall several facts about elementary symmetric polynomials and KL-divergence.

Auxiliary Definitions and Facts. Recall that the *downwalk operator* from $\binom{[N]}{s}$ to $\binom{[N]}{t}$ is defined as follows

$$D_{s\to t}(S,T) := \begin{cases} \frac{1}{\binom{s}{t}}, & T \subseteq S; \\ 0, & \text{otherwise}, \end{cases}$$

for every $S \in \binom{[N]}{s}$ to $T \in \binom{[N]}{t}$. Thus, viewing the distributions λ_i, μ_i in Lemma 6.4 as row vectors, then $\lambda_i = \mu_i D_{s\to t}$

Definition A.1 (Generating polynomial)**.** *Given a distribution μ on $\binom{[N]}{s}$, its generating polynomial $P_\mu \in \mathbb{R}[X_1, X_2, \ldots, X_N]$ is*

$$P_\mu(X) := \sum_{S\subseteq[N]:|S|=s} \mu(S) \prod_{i\in S} X_i.$$

The well-known MacLaurin's inequality on elementary symmetric polynomials [Mac30] reads: For nonnegative $X_1, X_2, \ldots, X_N$, and integers $s \geq t > 0$,

$$\left(\mathop{\mathbb{E}}_{S\in\binom{[N]}{s}} \prod_{i\in S} X_i \right)^{1/s} \leq \left(\mathop{\mathbb{E}}_{T\in\binom{[N]}{t}} \prod_{i\in T} X_i \right)^{1/t}. \tag{A.1}$$

An immediate corollary is that for the uniform distribution μ, its generating polynomial P_μ is log-concave on nonnegative inputs.

The next lemma about KL-divergence minimization follows from duality theory of convex optimization.

Lemma A.2 (See Appendix B of [SV14])**.** *Given a distribution μ on $\binom{[N]}{s}$, and a distribution $q : [N] \to \mathbb{R}$, then*

$$\inf_{\nu:\binom{[N]}{s}\to\mathbb{R}} \{\mathrm{KL}(\nu \,\|\, \mu) : q = \nu D_{s\to 1}\} = -\log \left(\inf_{x_1,x_2,\ldots,x_N>0} \frac{P_\mu(x)}{(x_1^{q(1)} x_2^{q(2)} \cdots x_N^{q(N)})^s} \right). \tag{A.2}$$

Proof of Claim 6.5. Now we are ready to prove Claim 6.5. Without loss of generality, say $x_i = i$. Let μ_1', μ_0' be the induced distribution of μ, μ_1 on $\binom{\{i,i+1,\ldots,N\}}{s-i+1}$ conditioning on

$[i] \subseteq S, S'$. Then μ'_0 is uniform on $\binom{\{i,i+1,\dots,N\}}{s-i+1}$. Let $q := \mu'_1 D_{s-i+1\to 1}$, then

$$\begin{aligned}
&\mathrm{KL}\left(\frac{X_i X_{i+1}\dots X_s \mid X_{<i} = x_{<i}}{Y_i Y_{i+1}\dots Y_s \mid Y_{<i} = x_{<i}}\right)\\
&\qquad = \mathrm{KL}(\mu'_1 \,\|\, \mu'_0)\\
&\qquad \geq \inf_{\mu''_1}\{\mathrm{KL}(\mu''_1 \,\|\, \mu'_0) : \mu''_1 D_{s-i+1\to 1} = q\}\\
&\qquad = -\log\left(\inf_{z_i,z_{i+1},\dots,z_N>0} \frac{P_{\mu'_0}(z)}{(z_i^{q(i)} z_{i+1}^{q(i+1)} \cdots z_N^{q(N)})^{s-i+1}}\right)\\
&\qquad \geq -\log\left(\inf_{z_i,z_{i+1},\dots,z_N>0} \left(\frac{\mathbb{E}_{j\in\{i,i+1,\dots,N\}}\, z_j}{z_i^{q(i)} z_{i+1}^{q(i+1)} \cdots z_N^{q(N)})}\right)^{s-i+1}\right).
\end{aligned}$$

where the second step is due to Lemma A.2; the third step is due to MacLaurin's inequality. Set $z_i = (N-s+1)q_i$, then

$$\begin{aligned}
\mathrm{KL}(\mu'_1 \,\|\, \mu'_0) &\geq -(s-i+1)\sum_{j=i}^{N} q(j)\log\frac{q_j}{1/(N-s+1)}\\
&= (s-i+1)\mathrm{KL}(\mu'_1 D_{s-i+1\to 1} \,\|\, \mu'_0 D_{s-i+1\to 1})\\
&= (s-i+1)\mathrm{KL}\left(\frac{X_i \mid X_{<i} = x_{<i}}{Y_i \mid Y_{<i} = x_{<i}}\right).
\end{aligned}$$

B Spectra of $\mathcal{D}_t$ from Johnson Scheme

The spectra of $\mathcal{D}_t$ is known [Del73]. In particular, fix any $0 \leq t \leq k-1$, there are $k+1$ distinct eigenvalues $\lambda_0, \lambda_1, \dots, \lambda_k$, such that

$$\begin{aligned}
\lambda_0 &= \binom{k}{t}\binom{d-k}{k-t},\\
\lambda_j &= \sum_{\ell=\max\{0,j-t\}}^{\min\{j,k-t\}} (-1)^\ell \binom{j}{\ell}\binom{k-j}{k-t-\ell}\binom{d-k-j}{k-t-\ell}, \qquad j = 1,2,\dots,k,
\end{aligned}$$

with multiplicity

$$\begin{aligned}
m_0 &= 1,\\
m_j &= \binom{d}{j} - \binom{d}{j-1}, \qquad j = 1,2,\dots,k.
\end{aligned}$$

For us, we simplify λ_j for $k = O(\sqrt{d})$,

$$\begin{aligned}
|\lambda_j| &\lesssim \frac{\binom{k-j}{t-j}}{\binom{k}{t}}\lambda_0, \qquad j = 1,2,\dots,t,\\
|\lambda_j| &\lesssim \frac{\binom{j}{t}(k-t)!}{\binom{k}{t}(k-j)!}\cdot\frac{1}{d^{j-t}}\lambda_0, \qquad j = t+1,\dots,k.
\end{aligned}$$

We bound $\|\mathcal{D}_t\|_1$ as follows

$$\begin{aligned}
\|\mathcal{D}_t\|_1 &= \lambda_0 + \sum_{j=1}^{k} m_j |\lambda_j| \\
&\lesssim \lambda_0 \left(1 + \sum_{j=1}^{t} \frac{d^{\underline{j}}}{j!} \frac{\binom{k-j}{t-j}}{\binom{k}{t}} + \sum_{j=t+1}^{k} \frac{d^{\underline{t}}}{j!} \frac{\binom{j}{t} (k-t)^{\underline{j-t}}}{\binom{k}{t}} \right) \\
&\lesssim \lambda_0 \left(1 + \frac{d^{\underline{t}}}{k^{\underline{t}}} + \sum_{j=t+1}^{k} \frac{d^{\underline{t}}}{k^{\underline{t}}} \binom{k-t}{j-t} \right) \\
&\leq \lambda_0 \left(1 + \frac{d^{\underline{t}}}{k^{\underline{t}}} 2^{k-t} \right) \\
&= \binom{k}{t} \binom{d-k}{k-t} \left(1 + \frac{d^{\underline{t}}}{k^{\underline{t}}} 2^{k-t} \right) \\
&\lesssim \binom{k}{t} \binom{d-k}{k-t} \frac{d^{\underline{t}}}{k^{\underline{t}}} 2^{k-t} = \binom{d}{t} \binom{d-k}{k-t} 2^{k-t}.
\end{aligned}$$

This proves Fact 5.9.