Proceedings of Machine Learning Research vol 196:1–33, 2024 37th Annual Conference on Learning Theory

The role of randomness in quantum state certification with unentangled measurements

Yuhan Liu YL2976@CORNELL.EDU
Cornell University

Jayadev Acharya ACHARYA@CORNELL.EDU
Cornell University

Editors: Shipra Agrawal and Aaron Roth

Abstract

Given n copies of an unknown quantum state $\rho \in \mathbb{C}^{d\times d}$, quantum state certification is the task of determining whether $\rho = \rho_0$ or $\|\rho - \rho_0\|_1 > \varepsilon$, where ρ_0 is a known reference state. We study quantum state certification using unentangled quantum measurements, namely measurements which operate only on one copy of ρ at a time. When there is a common source of randomness available and the unentangled measurements are chosen based on this randomness, prior work has shown that $\Theta(d^{3/2}/\varepsilon^2)$ copies are necessary and sufficient. This holds even when the measurements are allowed to be chosen adaptively. We consider deterministic measurement schemes (as opposed to randomized) and demonstrate that $\Theta(d^2/\varepsilon^2)$ copies are necessary and sufficient for state certification. This shows a separation between algorithms with and without randomness.

We develop a lower bound framework for both fixed and randomized measurements that relates the hardness of testing to the well-established Lüders rule. More precisely, we obtain lower bounds for randomized and fixed schemes as a function of the eigenvalues of the Lüders channel which characterizes one possible post-measurement state transformation.

1. Introduction

We study the problem of quantum state certification (O'Donnell and Wright, 2015; Wright, 2016; Badescu et al., 2019), where we are given n copies of an *unknown* quantum state with density $\rho \in \mathbb{C}^{d\times d}$, and complete description of a known state ρ_0. The goal is to use quantum measurements to test whether $\rho = \rho_0$ or $\|\rho - \rho_0\|_1 > \varepsilon$, where $\|\cdot\|_1$ is the trace norm. A special case of this problem is *mixedness testing*, which is the case when $\rho_0 = \rho_{\text{mm}} := \mathbb{I}_d/d$ is the maximally mixed state. Quantum certification is motivated by practical applications where one wants to verify whether the output state of a quantum algorithm is indeed the state we desire.

A related problem is of *closeness testing*, where we are given copies of two *unknown* states ρ and ρ_0 and the goal is again to test whether $\rho = \rho_0$ or $\|\rho - \rho_0\|_1 > \varepsilon$. The motivation to study this problem is to test whether two quantum algorithms produce the same state.

We are interested in determining how many copies of the unknown state(s) are needed to perform the task of testing. This task of understanding the copy complexity quantum state certification was studied initiated in O'Donnell and Wright (2015) and later in Badescu et al. (2019). They showed that when we are allowed to perform arbitrary entangled quantum measurements over the n copies, then $n = \Theta(d/\varepsilon^2)$ copies are necessary and sufficient for testing. However, entangled measurements are currently infeasible to implement in practice, even for moderate values of n and d. It is desirable to use unentangled measurements, where a quantum measurement is done on one copy of ρ (and ρ_0 if it is also unknown) at a time. Such *unentangled measurements* (also referred to as *incoherent* and *independent* in previous literature) can be categorized into three different protocols:

1. **Fixed/Deterministic measurements.** The set of measurements(POVMs) to be performed are fixed *ahead of time*. Once the copies of the quantum states are available, we use these fixed measurements for the task of testing. A key advantage of such protocols is that the same set of measurements can be used for multiple repetitions of the testing problem. Moreover, there is no latency since the measurements are not designed after the states are made available, which is a drawback of the following protocols.

2. **Randomized non-adaptive measurements.** In this setting, there is common randomness available, and the set of measurements at the different copies are all chosen simultaneously as a function of this common randomness. Every time we want to test for a state, we need to instantiate the common randomness and select the set of measurements using a new instance of the common randomness. This is done after the copies of the state are made available.[1]

3. **Randomized adaptive measurements.** In this setting common randomness is available across the measurements. Furthermore, the measurements are applied sequentially to each copy of ρ, and the measurement on the next copy of ρ can depend on the outcome of previous measurements[2]. A primary drawback of this scheme is the latency and complications associated with designing measurements one after another.

1.1. Prior Works

O'Donnell and Wright (2015) initiated the study of copy complexity of the task of quantum state certification. They considered entangled measurements and showed that $n = \Theta(d/\varepsilon^2)$ copies are necessary and sufficient for testing. This is also the copy complexity of closeness testing (Badescu et al., 2019).

Given the practical relevance of unentangled measurements, it has been considered in several prior works. For the task of quantum mixedness testing, Bubeck et al. (2020) showed that when randomized non-adaptive unentangled measurements are allowed, then $n = \Theta(d^{3/2}/\varepsilon^2)$ copies are necessary and sufficient. Chen et al. (2022b) extended the results to the cases when ρ_0 need not be the maximally mixed state, and also when it is unknown (closeness testing).

Chen et al. (2022a) futher showed that adaptivity does not help and the number of copies necessary is still $n = \Omega(d^{3/2}/\varepsilon^2)$. Yu (2023) achieved $n = \tilde{\Theta}(d^2/\varepsilon^2)$ using randomly sampled Pauli measurements, which are more restrictive yet easier to implement. The drawback of all the algorithms in these works is the necessity of randomization in the measurements.

Quantum tomography. In *quantum tomography* (O'Donnell and Wright, 2017; Flammia and O'Donnell, 2023), the goal is to estimate the unknown state ρ to within ε in trace distance. O'Donnell and Wright (2016, 2017); Haah et al. (2017) established the optimal copy complexity for this task as $\Theta(d^2/\varepsilon^2)$ with entangled measurements. For unentangled measurements, various works Kueng et al. (2017); Haah et al. (2017) have shown that $\Theta(d^3/\varepsilon^2)$ are necessary and sufficient to estimate a full-rank ρ, even when adaptivity is allowed (Chen et al., 2023). Guţă et al. (2020) showed that the bound is achievable up to log factors using *fixed* structured POVMs, e.g. SIC-POVM (Zauner, 1999; Renes et al., 2004), maximal MUB (Klappenecker and Rotteler, 2005).

1. If the set of measurements is finite, we can prepare all measurements beforehand and sample with classical randomness. However, this could still be difficult if the set is very large.
2. In principle, one can use the first measurement outcome as a source of randomness for all other measurements, so adaptive measurements are essentially randomized.

1.2. New results

We consider state certification with fixed measurements, where the POVMs are fixed ahead of time, and can be used for multiple repetitions of the problem.

The naive solution is to apply the fixed unentangled measurements for quantum tomography in Guţă et al. (2020) giving an upper bound of $n = \tilde{O}\big(d^3/\varepsilon^2\big)$. However, since tomography is strictly harder than testing, we expect to do much better than d^3. Indeed, Yu (2021) designed a simple algorithm with fixed measurements that achieves $O(d^2/\varepsilon^2)$ copy complexity. The lower bound, however, was left as an outstanding open problem. Without randomness, it is unknown whether $n = O(d^{3/2}/\varepsilon^2)$ copies are sufficient to perform quantum state certification, or if we need more copies due to the lack of randomness.

We establish the copy complexity of quantum state certification with fixed unentangled measurements. Our main result, stated below, shows that indeed there is a cost in copy complexity that we have to pay for having schemes that are *fixed and reusable*. Please see Section 2.2 for the formal problem definition.

Theorem 1 *For fixed unentangled POVMs, $n = \Theta(d^2/\varepsilon^2)$ copies are necessary and sufficient to test whether $\rho = \rho_0$ or $\|\rho - \rho_0\|_1 > \varepsilon$ with probability at least 2/3.*

Table 1 places our work in the context of existing results for other types of measurements. There is a strict $\Theta(\sqrt{d})$-factor separation between fixed and randomized non-adaptive schemes. We note that the randomness source can be entirely independent of the quantum states, so it is in some sense surprising that a piece of irrelevant random information leads to substantial improvement in copy complexity. This demonstrates that randomness is a valuable and important resource in unentangled quantum state certification.

Measurement type		Upper bound	Lower bound
Entangled		$\frac{d}{\varepsilon^2}$	$\frac{d}{\varepsilon^2}$
Unentangled	Adaptive	$\frac{d^{3/2}}{\varepsilon^2}$	$\frac{d^{3/2}}{\varepsilon^2}$
	Randomized	$\frac{d^{3/2}}{\varepsilon^2}$	$\frac{d^{3/2}}{\varepsilon^2}$
	Fixed	$\frac{d^2}{\varepsilon^2}$	$\frac{d^2}{\varepsilon^2}$ **(This work)**

Table 1: Existing and new worst-case copy complexity results for quantum state certification.

We develop an information-theoretic framework for non-adaptive schemes that leads to both the lower bound of $\Omega(d^2/\varepsilon^2)$ for fixed measurements and the bound of $\Omega(d^{3/2}\varepsilon^2)$ of randomized ones. Details are elaborated in Section 3.

1.3. Related works

Learning information about quantum states. Our work falls into the line of quantum state certification O'Donnell and Wright (2015); Bubeck et al. (2020); Chen et al. (2022a). In addition to worst-case bounds that depend on d, Chen et al. (2022b,a) considered general quantum state certification where the copy complexity decreases when ρ_0 is approximately low rank. Other closeness measures such as fidelity and Bures χ^2-divergence were considered in Badescu et al. (2019).

Many works have studied other related problems such as closeness testing Badescu et al. (2019); Yu (2021, 2023) (test whether *two unknown* states ρ and σ are equal or ε far), hypothesis testing Ogawa and Nagaoka (2000); Brandão et al. (2020); Regula et al. (2023) (distinguish between two known states), and hypothesis selection Badescu and O'Donnell (2021); Fawzi et al. (2023) (determine ρ from a finite set of hypothesis sets). Shadow tomography (Aaronson, 2020; Huang et al., 2020; Brandão et al., 2019) considers the problem of learning the statistics of the state ρ over a finite set of observables, which is simpler than tomography. Algorithms for shadow tomography can be applied to quantum hypothesis selection (Badescu and O'Donnell, 2021; Fawzi et al., 2023).

In addition to the four types of measurements discussed before, *Pauli* measurements have also attracted significant interest (Flammia and Liu, 2011; Liu, 2011; Cai et al., 2016; Yu, 2023) due to ease in implementation despite being less powerful. Moreover, Fawzi et al. (2023) considered *sequential* strategies which allow the number of measurements to depend on previous outcomes (e.g. one can choose to stop measuring the remaining copies if the outcomes so far yield a good estimate), which is parallel to the adaptivity of measurements.

Classical distribution testing. Quantum state certification can be viewed as the quantum equivalent of testing identity of discrete distributions from samples. Here the task is to decide from samples whether a distribution is equal to a given known distribution. The problem has been well studied starting with the works of Batu et al. (2001); Paninski (2008) which establish the sample complexity of this task when all the samples are available. This is similar to using entangled measurements in the quantum case. Recently there has been significant work on distributed testing of distributions, where instead of having all samples at the same place, they are distributed across users, and we obtain only limited information about each sample, e.g., a communication-constrained (Barnes et al., 2020; Acharya et al., 2020a), or privacy-preserving information (Duchi et al., 2013; Acharya et al., 2021; Han et al., 2015). Thinking of each sample analogous to one copy, this distributed testing is in spirit similar to unentangled measurements, where we perform measurements on one copy at a time. Acharya et al. (2020b, 2022) derived a unified information-theoretic framework in terms of the channel constraints. In particular, Acharya et al. (2020b,a, 2021) showed that there is a separation for distributed testing under communication and privacy constraints between the cases when common randomness was available versus not. Furthermore, Acharya et al. (2022) show that adaptivity does not help in these problems beyond common randomness. Our results are qualitatively similar to these classical distributed testing results. We show in this work that these ideas can be generalized to quantum state certification and a similar separation also holds. We refer the readers to Canonne (2022) for a comprehensive survey of the above topics.

Outline. The rest of the paper is organized as follows. In Section 3 we overview our main technical contributions. In Section 2, we give the precise problem definition and provide some mathematical terminology and definitions. In Section 4 we introduce our unified lower bound framework for non-adaptive measurements. In Appendix D we describe the algorithm that achieves the copy complexity upper bound for fixed measurements.

2. Preliminaries

2.1. Basics of quantum computing

Quantum states The space of d-dimensional complex vectors $\mathbb{C}^d$ forms a Hilbert space. We use the Dirac notation $|\psi\rangle \in \mathbb{C}^d$ to denote a vector, and $\langle\psi|$ is its conjugate transpose, which is a row vector. The Hilbert-Schmidt inner product between $|\psi\rangle$ and $|\phi\rangle$ is $\langle\psi|\phi\rangle$. In a d-dimensional quantum system, the state ρ is a $d \times d$ positive semi-definite Hermitian matrix with $\mathrm{Tr}[\rho] = 1$. If the rank of ρ is 1, then ρ is a *pure* state and $\rho = |\psi\rangle\langle\psi|$ for some unit-norm $|\psi\rangle \in \mathbb{C}^d$. Otherwise, the state is a *mixed state*. A special case is $\rho_{\text{mm}} := \mathbb{I}_d/d$, which is the maximally mixed state.

Measurements All measurements can be formulated as *positive operator-valued measure* (POVM). Let $\mathcal{X}$ be a finite set of outcomes. Then a POVM $\mathcal{M} = \{M_x\}_{x\in\mathcal{X}}$, where M_x is p.s.d. and $\sum_{x\in\mathcal{X}} M_x = \mathbb{I}_d$. Let X be the outcome when applying measurement $\mathcal{M}$ to ρ, then the probability of observing x is given by the *Born's rule*,

$$\Pr[\, X = x\,] = \mathrm{Tr}[\rho M_x].$$

We note that the outcome set $\mathcal{X}$ need not be finite, in which case POVMs and Born's rule can be generalized. However, finite POVMs are without loss of generality. In principle, all physically feasible measurements are finite. Moreover, our argument extends easily to infinite POVMs.

2.2. Problem setup

Given n independent copies of an unknown quantum state $\rho \in \mathbb{C}^{d\times d}$, the goal is to design

- n POVMs $\mathcal{M}^n = (\mathcal{M}_1, \ldots, \mathcal{M}_n)$ that are applied to the n copies of the state that produce the measurement outcomes $\mathbf{x} = (x_1, \ldots, x_n)$,
- a tester T such that when $\rho = \rho_0$ it outputs `YES` with probability at least 2/3 and it outputs `NO` with probability at least $2/3$ if $\|\rho - \rho_0\|_1 > \varepsilon$,

$$\Pr_{\rho=\rho_0}(T(\mathbf{x}) = \texttt{YES}) \geq \frac{2}{3}, \text{ and } \inf_{\rho:\|\rho-\rho_0\|_1>\varepsilon} \Pr(T(\mathbf{x}) = \texttt{NO}) \geq \frac{2}{3}.$$

When $\rho_0 = \rho_{\text{mm}} := \mathbb{I}_d/d$, the problem is called *mixedness testing*. The smallest value of n for which we can design such a tester for all ρ_0 is the *copy complexity* of quantum state certification.

We apply measurements for each copy individually. More precisely, for the i-th copy, we apply a POVM $\mathcal{M}_i = \{M_x^i\}_{x=1}^k$ where M_x^i is p.s.d. and $\sum_x M_x^i = \mathbb{I}_d$. Let x_i be the measurement outcome after applying $\mathcal{M}_i$ on the i-th copy. When the quantum state is ρ, x_i follows a discrete distribution $\mathbf{p}_\rho^i = [\mathbf{p}_\rho^i(1), \ldots, \mathbf{p}_\rho^i(k)]$ given by *Born's rule*, $\mathbf{p}_\rho^i(x) = \mathrm{Tr}[M_x^i\rho], \quad x = 1, \ldots, k$.

According to (Chen et al., 2021, Lemma 4.8), general finite POVMs can be simulated using rank-1 POVMs if we only consider the measurement outcomes and disregard the post-measurement quantum state. Thus it is without loss of generality to only consider rank-1 POVMS, i.e.,

$$M_x^i = |\psi_x^i\rangle\langle\psi_x^i|, \quad \sum_{x=1}^k |\psi_x^i\rangle\langle\psi_x^i| = \mathbb{I}_d \tag{1}$$

Note that $|\psi_x^i\rangle$ may not be normalized.

We can mathematically formulate the three unentangled measurement schemes as follows,

In *fixed* measurement schemes, each $\mathcal{M}_i$ is fixed before receiving the quantum state ρ. The n outcomes $x_1, \ldots, x_n$ follow a product distribution

$$\mathbf{P}_\rho := \bigotimes_{i=1}^n \mathbf{p}_\rho^i. \tag{2}$$

In *randomized non-adaptive schemes*, there is a common random seed $R \sim \mathcal{R}$ independent of ρ, and the measurements are then chosen based on R, $\mathcal{M}_i = \mathcal{M}_i(R)$. The outcomes are independent conditioned on the random seed R, and thus we can write $\mathbf{P}_\rho(R) = \bigotimes_{i=1}^n \mathbf{p}_\rho^i(R)$.

For *randomized adaptive schemes*, in addition to the common randomness, the ith measurement depends on the previous outcomes, namely, $\mathcal{M}_i = \mathcal{M}_i(x_1, \ldots, x_{i-1}, R)$. The n outcomes are in general not independent.

2.3. Closeness measures of distributions

Let $\mathbf{p}$ and $\mathbf{q}$ be two distributions over a finite discrete domain $\mathcal{X}$. The *total variation distance* is defined as, $d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) := \sup_{S \subseteq \mathcal{X}}(\mathbf{p}(S) - \mathbf{q}(S)) = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mathbf{p}(x) - \mathbf{q}(x)|$. The Kullback-Leibler (KL) divergence $\mathrm{KL}(\mathbf{p} \,||\, \mathbf{q})$ and chi-square divergence $\mathrm{d}_{\chi^2}(\mathbf{p} \,||\, \mathbf{q})$ are defined as $\mathrm{KL}(\mathbf{p} \,||\, \mathbf{q}) := \sum_{x\in\mathcal{X}} \mathbf{p}(x) \log \frac{\mathbf{p}(x)}{\mathbf{q}(x)}$, $\mathrm{d}_{\chi^2}(\mathbf{p} \,||\, \mathbf{q}) := \sum_{x\in\mathcal{X}} \frac{(\mathbf{p}(x)-\mathbf{q}(x))^2}{\mathbf{q}(x)}$. The three quantities can be related using Pinsker's inequality and concavity of logarithms respectively,

$$2d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})^2 \le \mathrm{KL}(\mathbf{p} \,||\, \mathbf{q}) \le \mathrm{d}_{\chi^2}(\mathbf{p} \,||\, \mathbf{q}).$$

We may also define ℓ_p distances between distributions, $\|\mathbf{p} - \mathbf{q}\|_p := \left(\sum_{x\in\mathcal{X}} |\mathbf{p}(x) - \mathbf{q}(x)|^p\right)^{1/p}$.

2.4. Linear operators and superoperators

Hilbert space over complex matrices The space of complex matrices $\mathbb{C}^{d\times d}$ is a Hilbert space when equipped with the matrix inner product defined as $\langle A, B\rangle := \mathrm{Tr}[A^\dagger B]$, where $A, B \in \mathbb{C}^{d\times d}$. The subspace of all Hermitian matrices, denoted as $\mathbb{H}_d$, is a *real* Hilbert space (i.e. the associated field is $\mathbb{R}$) with the same matrix inner product. Any positive semi-definite Hermitian matrix M has a unique p.s.d. square root K such that $K^2 = M$, and we denote $K = \sqrt{M}$.

A homomorphism can be defined between $\mathbb{C}^{d\times d}$ and $\mathbb{C}^{d^2}$ through vectorization. On the canonical basis $\{|j\rangle\}_{j=0}^{d-1}$, we define $\mathrm{vec}(|i\rangle\langle j|) := |j\rangle \otimes |i\rangle$. Vectorization for general matrices is defined by linearity. Furthermore, the matrix inner product can be equivalently written as the inner product on $\mathbb{C}^{d^2}$, $\langle A, B\rangle = \mathrm{vec}(A)^\dagger \mathrm{vec}(B)$.

(Linear) superoperators One can define linear operators over $\mathbb{C}^{d\times d}$, $\mathcal{N} : \mathbb{C}^{d\times d} \mapsto \mathbb{C}^{d\times d}$. Since each matrix itself can be viewed as an operator over $\mathbb{C}^d$, we refer to them as superoperators[3] to avoid confusion. For each superoperator $\mathcal{N}$, there exists a unique adjoint superoperator $\mathcal{N}^\dagger$ such that for all $X, Y \in \mathbb{C}^{d\times d}$, $\langle Y, \mathcal{N}(X)\rangle = \left\langle \mathcal{N}^\dagger(Y), X\right\rangle$. Similar to the trace of matrices, we define its trace as $\mathrm{Tr}[\mathcal{N}] = \sum_{i,j=1}^d \langle |i\rangle\langle j|, \mathcal{N}(|i\rangle\langle j|)\rangle$.

Schatten norms for linear (super)operators Let $\lambda_1, \ldots, \lambda_d \ge 0$ be the *singular values* of a linear operator A[4], then for $p \ge 1$, the *Schatten p-norm* is defined as $\|A\|_{S_p} := \left(\sum_{i=1}^d \lambda_i^p\right)^{1/p}$, which can be defined for both matrices and superoperators. Some important special cases are trace norm $\|A\|_1 := \|A\|_{S_1}$, Hilbert-Schmidt norm $\|A\|_{\mathrm{HS}} := \|A\|_{S_2}$, and operator norm $\|A\|_{\mathrm{op}} := \|A\|_{S_\infty} = \max_{i=1}^d \lambda_i$.. A standard fact is that $\|A\|_1 = \mathrm{Tr}[\sqrt{A^\dagger A}]$ and $\|A\|_{\mathrm{HS}} = \sqrt{\langle A, A\rangle}$.

3. Indeed an operator over $\mathbb{C}^{d\times d}$ i.e. superoperator need not be linear, but we only deal with linear superoperators in this work, so we drop the word "linear" for brevity.
4. For Hermitian matrices, the singular values are simply the absolute values of the eigenvalues.

3. Our techniques

Our main contribution is a unified lower bound framework for quantum state certification that works for both randomized and fixed non-adaptive schemes. Before we introduce the technical contributions, we provide a high-level explanation of why randomness is crucial in unentangled quantum testing and motivate with a simple example.

3.1. The disadvantage of fixed measurements through a simple example.

In randomized schemes, given the copies of the state, we then choose the measurements randomly. However, for fixed measurements, the measurement scheme is fixed and the state is then chosen. Thus, without randomness, nature would have the opportunity to *adversarially* design a quantum state that fools the pre-defined set of measurements. When randomness is available, we can avoid the bad effect of adversarial choice of quantum states. In principle, this qualitative gap is like the difference between randomized algorithms and deterministic algorithms.

We use a simple example to demonstrate this idea. Suppose we choose each measurement $\mathcal{M}_i$ simply to be the same canonical basis measurement, i.e. $\mathcal{M}_i = \{|x\rangle\langle x|\}_{x=0}^{d-1}$. Then nature can set ρ to be the "+" state where

$$\rho = |\phi\rangle\langle\phi|, \quad |\phi\rangle = \frac{1}{\sqrt{d}}\sum_{x=0}^{d-1}|x\rangle. \tag{3}$$

Note that the trace distance $\|\rho - \rho_{\mathrm{mm}}\|_1 = 2 - 2/d \simeq 2$. When the state is ρ, all measurement outcomes x_i would be independent samples from the uniform distribution over $\{0, \ldots, d-1\}$. However, if the state is the maximally mixed state ρ_{mm}, the distribution of each measurement outcome would also be the uniform distribution over $\{0, \ldots, d-1\}$. Thus, even though the trace distance between ρ and ρ_{mm} is large, the measurement scheme is completely fooled.

On the other hand, with randomness, one can (theoretically) sample a basis uniformly from the Haar measure as in Bubeck et al. (2020) to easily avoid this issue. No fixed ρ would be able to completely fool the randomized basis measurement sampled uniformly. In fact with high probability, the randomly sampled basis is good in the sense that the outcome distribution would be far enough when the two states ρ and ρ_{mm} are far (see (Chen et al., 2022b, Lemma 6.3)).

3.2. A novel lower bound construction.

The design of hard instances has to account for the difference illustrated above when proving lower bounds for randomized and fixed measurements. In particular, for randomized measurements, the lower bound construction is *measurement independent*. However, for fixed measurements, since the states can be chosen adversarially, the lower bound construction needs to be *measurement-dependent*.

Many prior works on testing and tomography O'Donnell and Wright (2015, 2016); Haah et al. (2017); Bubeck et al. (2020); Chen et al. (2021, 2022a) use *measurement-independent* distributions over states in $\mathbb{C}^{d\times d}$ to prove lower bounds. In particular, Bubeck et al. (2020); Chen et al. (2022a) show that testing within a specific class requires at least $n = \Omega(d^{3/2}/\varepsilon^2)$ when working with *randomized* and *adaptive* unentangled measurements respectively. Unfortunately, for these measurement-independent constructions, there must exist fixed measurement schemes whose copy complexity is $n = O(d^{3/2}/\varepsilon^2)$ due to standard derandomization arguments. To prove a stronger lower bound for fixed measurements, our task is different. We must show that for any fixed measurement scheme, we can design a hard instance of the testing problem that would require at least $n = \Omega(d^2/\varepsilon^2)$. We note that the lower bound construction in Yu (2023) is measurement-dependent, but specifically tailored to Pauli measurements and not general enough for our purpose.

Our generic *measurement-dependent* lower bound construction is a necessary and novel contribution that leads to tight lower bounds for fixed measurements.

Definition 2 *Let $\frac{d^2}{2} \le \ell \le d^2 - 1$ and $\mathcal{V} = (V_1, \ldots, V_{d^2})$ be an orthonormal basis of $\mathbb{H}_d$ with $V_{d^2} = \mathbb{I}_d/\sqrt{d}$. Define $\mathcal{D}_\ell(\mathcal{V})$ as follows. Let $z = (z_1, \ldots, z_\ell) \in \{-1, 1\}^\ell$ be uniformly drawn from the $\{-1, 1\}^\ell$ hypercube. Let c be a universal constant, then define*

$$\Delta_z = \frac{c\varepsilon}{\sqrt{d}} \cdot \frac{1}{\sqrt{\ell}} \left(\sum_{i=1}^{\ell} z_i V_i \right), \quad \bar{\Delta}_z = \Delta_z \min \left\{ 1, \frac{1}{d \|\Delta_z\|_{op}} \right\}. \tag{4}$$

$\bar{\Delta}_z$ normalizes Δ_z so that its maximum absolute eigenvalue is at most $1/d$.$\sigma_z = \rho_{mm} + \bar{\Delta}_z$. This defines a distribution over states (induced by the randomness in z) which we denote by $\mathcal{D}_\ell(\mathcal{V})$.

In essence, we perform independent binary perturbations along ℓ different trace-0 directions. We show with appropriate c, regardless of the basis $\mathcal{V}$, σ_z is ε-far in trace distance from ρ_{mm}. The proof is in Section B.

Proposition 3 *Let $d^2/2 \le \ell \le d^2 - 1$. Let z be drawn uniformly from $\{-1, 1\}^\ell$, and Δ_z, σ_z are as defined in Definition 2. Then, there exists a universal constant $c \le 10\sqrt{2}$, such that for $\varepsilon < \frac{1}{c^2}$, with probability at least $1 - 2\exp(-d)$, $\|\Delta_z\|_{op} \le 1/d$ and $\|\Delta_z\|_1 \ge \varepsilon$.*

The matrices V_i's can be chosen *dependent* on the fixed measurement scheme that we want to *fool*. In particular, we can pick directions $V_1, \ldots, V_{d^2/2}$ about which the fixed measurement schemes provide the least information. The matrices V_i's can also be fixed, in which case the construction is measurement-independent and our framework naturally leads to the lower bound for randomized non-adaptive measurements in Bubeck et al. (2020).

The binary perturbations in our construction are mathematically easier to handle. In prior works, Bubeck et al. (2020); Haah et al. (2017); Chen et al. (2021) designed the hard cases using random unitary transformations around the maximally-mixed state, which requires difficult calculations using Weingarten calculus (Weingarten, 1978; Collins, 2003). In contrast, our arguments avoid the difficult representation-theoretic tools. Chen et al. (2022a, 2023) used Gaussian orthogonal ensembles, which perturbs each matrix entry with independent Gaussian distributions. Binary perturbations share many statistical similarities with Gaussian since both are sub-gaussian distributions. However, the former is arguably simpler as the support is finite, and thus information-theoretic tools can be more easily applied.

We note that these constructions are all in spirit motivated by lower bounds in classical discrete distribution testing where the hard instances are constructed as perturbations around the uniform distribution (Paninski, 2008).

3.3. Lower bound framework using Lüders channel

For both fixed and randomized schemes, we find that perhaps very coincidentally or very naturally, the ability of a measurement scheme to distinguish between quantum states is characterized by the eigenvalues of the Lüders channel (DeBrota and Stacey, 2019)[5] which describes one possible state transformation after measurement. Let $\mathcal{M} = \{M_x\}_{x=1}^k$ be a rank-1 POVM where $M_x = |\psi_x\rangle\langle\psi_x|$. When $\mathcal{M}$ acts on ρ and we obtain an outcome x, then the *generalized Lüder's rule* (Lüders, 1950) gives one possible form of the post-measurement state[6],

$$\rho^x := \frac{K_x \rho K_x}{\mathrm{Tr}[M_x \rho]}, \quad K_x = \sqrt{M_x} = \frac{|\psi_x\rangle\langle\psi_x|}{\sqrt{\langle\psi_x|\psi_x\rangle}}.$$

If we lack the knowledge of measurement outcomes, we can view the underlying state as an expectation of all post-measurement states $\rho \mapsto \sum_x \Pr[\, X = x\,]\rho^x$. We can formulate the mapping as a quantum channel,

Definition 4 (Lüders channel) *Given a rank-1 POVM $\mathcal{M} = \{M_x = |\psi_x\rangle\langle\psi_x|\}_{x=1}^k$ where $\sum_x M_x = \mathbb{I}_d$ and $K_x = \sqrt{M_x}$, the Lüders channel $\mathcal{H}_{\mathcal{M}} : \mathbb{C}^{d\times d} \mapsto \mathbb{C}^{d\times d}$ is defined as*

$$\mathcal{H}_{\mathcal{M}}(X) := \sum_{x=1}^k K_x X K_x = \sum_{x=1}^k \frac{|\psi_x\rangle\langle\psi_x|X|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}. \tag{5}$$

Definition 4 gives its *Kraus representation*. It is helpful to consider the equivalent matrix form $\mathcal{C}_{\mathcal{M}} \in \mathbb{C}^{d^2\times d^2}$ which satisfies $\mathrm{vec}(\mathcal{H}_{\mathcal{M}}(X)) = \mathcal{C}_{\mathcal{M}}\mathrm{vec}(X)$ for all matrix $X \in \mathbb{C}^{d\times d}$. For rank-1 POVM, $\mathcal{C}_{\mathcal{M}}$ is the *Choi representation*, e.g. Khatri and Wilde (2021, Eq (4.3.8)) and defined as,

$$\mathcal{C}_{\mathcal{M}} := \sum_{x=1}^k \frac{|\bar{\psi}_x\rangle\langle\bar{\psi}_x| \otimes |\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}. \tag{6}$$

Indeed, if two quantum states ρ and σ yield the same post-measurement state, it is natural to believe that the two states cannot be distinguished by the measurement scheme. This is in spirit similar to the chi-squared contraction framework in Acharya et al. (2020b) which was used to derive a unified lower bound for information-constrained inference of classical distributions in the distributed setting. Depending on whether randomness is available, the copy complexity depends on different norms of the $\mathcal{H}$ channel, as shown in Table 2, where $\|\mathcal{H}\|_1$ and $\|\mathcal{H}\|_{\mathrm{HS}}$ are the trace and Hilbert-Schmidt/Frobenius norms of $\mathcal{H}$. The precise statement is stated in Theorem 11.

	Fixed	Randomized
Lower bound	$\frac{d^2}{\varepsilon^2}\cdot\frac{d}{\max_{\mathcal{H}}\|\mathcal{H}\|_1}$	$\frac{d^2}{\varepsilon^2 \max_{\mathcal{H}}\|\mathcal{H}\|_{\mathrm{HS}}}$

Table 2: Copy complexity lower bound of non-adaptive state certification in terms of the Lüders channel.

It is straightforward to prove that $\|\mathcal{H}\|_{\mathrm{HS}}^2 \le \|\mathcal{H}\|_1 \le d$, and thus we obtain tight copy complexity lower bounds for both fixed and randomized non-adaptive measurements.

Our lower bound results yield a very natural quantum interpretation. We believe that a similar characterization using Lüders channel could be applied to adaptive measurements and other problems such as quantum tomography, similar to Acharya et al. (2020b, 2022) which developed a unified framework for distributed learning and testing of discrete distributions.

5. This was referred to as the expected density operator in (Khatri and Wilde, 2021, Section 3.3)
6. The post-measurement states are undefined for general POVMs.

4. Lower bound framework for non-adaptive schemes

We first state the new lower bound for mixedness testing with fixed measurements in Theorem 5.

Theorem 5 *For $0 < \varepsilon < 1/200$ and $d \geq 16$, with fixed unentangled measurements, at least $n = \Omega(d^2/\varepsilon^2)$ copies are necessary to test whether $\rho = \rho_{mm}$ or $\|\rho - \rho_{mm}\|_1 > \varepsilon$ where $\rho_{mm} = \mathbb{I}_d/d$ is the maximally mixed state.*

Since mixedness testing is a special case of state certification, this theorem provides a worst-case lower bound for the problem, both when ρ_0 is known and unknown. Recall that $n = O(d^{3/2}/\varepsilon^2)$ copies are sufficient using randomized schemes. Thus there is a strict separation between algorithms with and without randomness, and the gap is a factor of $\Theta(\sqrt{d})$.

Theorem 5 is an immediate corollary of a unified theoretical framework that we establish for both randomized and fixed non-adaptive measurements which we will describe in this section.

4.1. Le Cam's method

The central tool to prove testing lower bounds is Le Cam's method LeCam (1973); Yu (1997). We first define *almost-ε perturbation*, which has constant probability mass over $\mathcal{P}_\varepsilon$.

Definition 6 *A distribution $\mathcal{D}$ is an* almost-ε perturbation *if* $\Pr_{\sigma\sim\mathcal{D}}[\,\sigma \in \mathcal{P}_\varepsilon\,] > \frac{4}{5}$. *Denote the set of* almost-ε perturbation *as* Γ_ε.

Recall in (2) that for a state ρ, the distribution of measurement outcomes $\mathbf{x} = (x_1, \ldots, x_n)$ when applying measurements $\mathcal{M}^n = (\mathcal{M}_1, \ldots, \mathcal{M}_n)$ is $\mathbf{P}_\rho$. For the mixedness testing problem, even if ρ is sampled from any distribution $\mathcal{D}$ over the $\mathcal{P}_\varepsilon := \{\rho : \|\rho - \rho_{\rm mm}\|_1 > \varepsilon\}$ (the set of states at least ε-far from $\rho_{\rm mm}$), a mixedness tester should still be able to distinguish it from the case when the state is $\rho_{\rm mm}$. When $\rho \sim \mathcal{D}$, the outcome distribution is $\mathbb{E}_{\rho\sim\mathcal{D}}[\,\mathbf{P}_\rho\,]$, and when $\rho = \rho_{\rm mm}$, then $\mathbf{x} \sim \mathbf{P}_{\rho_{\rm mm}}$. Thus, to guarantee a testing accuracy of at least $2/3$, we need $\frac{2}{3} < d_{\rm TV}(\mathbf{P}_{\rho_{\rm mm}}, \mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,])$. A similar argument also holds for $\mathcal{D}$ which is an *almost-ε perturbation*, which is stated in Lemma 7. The proof is in Section A.1.

Lemma 7 *Let $\mathcal{D}$ be an almost-ε perturbation. Suppose nature flips an unbiased coin $Y \in \{0, 1\}$. If $Y = 0$ then $\rho = \rho_{mm}$. Otherwise nature samples $\rho \sim \mathcal{D}$. Then, using a mixedness tester with success probability at least $2/3$ for n copies of ρ, we can obtain a guess $\hat{Y} \in \{0, 1\}$ with $\Pr[Y = \hat{Y}] \geq 3/5$. This implies*

$$\frac{1}{5} \leq d_{\rm TV}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,], \mathbf{P}_{\rho_{mm}}) \leq \sqrt{\frac{1}{2} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \mid\mid \mathbf{P}_{\rho_{mm}})}. \tag{7}$$

Thus to prove copy complexity lower bounds, we need to design an almost-ε perturbation and then upper bound (7) by some function of n, d, ε.

4.2. Min-max vs. max-min

Recall in Section 3.1 we discussed that the main difference between randomized and fixed measurements is whether nature can choose the hard state adversarially. In this section, we formalize the discussion under a rigorous game theory framework. The testing problem can be viewed as a two-party game played between nature and the algorithm designer, where the algorithm designer tries to design the best algorithms that can distinguish between two states, while nature tries to find hard states to fool the algorithm.

For a fixed measurement scheme $\mathcal{M}^n$, nature can choose a $\mathcal{D} \in \Gamma_\varepsilon$ that minimizes the chi-square divergence in (7). According to Lemma 7, if there exists a fixed $\mathcal{M}^n$ that achieves at least 2/3 probability in testing maximally mixed states, we must have

$$\frac{2}{25} \leq \max_{\mathcal{M}^n \text{ fixed}} \min_{\mathcal{D}\in\Gamma_\varepsilon} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{\mathrm{mm}}}). \tag{8}$$

Thus a max-min game is played between the two parties and nature has an advantage to decide its best action based on the choice of the algorithm designer.

With randomness, in principle, a max-min game is still played, but instead, the maximization is over all *distributions* of fixed (non-entangled) measurements. Using a similar argument as (Acharya et al., 2020b, Lemma IV.8), for the best distribution over all $\mathcal{M}^n$, the expected accuracy over $R \sim \mathcal{R}$ is at least 1/2 for all $\mathcal{D} \in \Gamma_\varepsilon$. Thus, for all $\mathcal{D}$, there must exist an instantiation $R(\mathcal{D})$ such that using the fixed measurement $\mathcal{M}^n(R(\mathcal{D}))$ the testing accuracy is at least 1/2. Therefore,

$$\frac{2}{25} \leq \min_{\mathcal{D}\in\Gamma_\varepsilon} \max_{\mathcal{M}^n \text{ fixed}} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{\mathrm{mm}}}), \tag{9}$$

which intuitively says that a min-max game is played and the algorithm designer has an advantage.

Therefore, to obtain a copy complexity lower bound for fixed measurements requires upper bounding (8), while for randomized schemes requires upper bounding (9). We can see that randomness is a "game changer" that changes a max-min game to a min-max game. Since min-max is no smaller than max-min, testing with randomness is easier than testing without it.

The min-max and max-min arguments in this section are similar to Acharya et al. (2020b) and we point to Acharya et al. (2020b, Lemma IV.8, IV.10) for additional reference.

4.3. The Lüders channel characterizes the hardness of testing

In the previous section, we give an abstract theoretical framework to prove tight lower bounds for fixed measurements. We now make it concrete and apply it to mixedness testing.

Our central contribution is to relate the hardness of testing (i.e., the min-max and max-min divergences) to the *average Lüders channel* defined by all the POVMs. Use the shorthand $\mathcal{H}_i := \mathcal{H}_{\mathcal{M}_i}$ where $\mathcal{H}_{\mathcal{M}_i}$ is from Definition 4, the average Lüders channel is defined as

$$\bar{\mathcal{H}} := \frac{1}{n}\sum_{i=1}^{n} \mathcal{H}_i \text{ (Kraus)}, \quad \bar{\mathcal{C}} := \frac{1}{n}\sum_{i=1}^{n} \mathcal{C}_i \text{ (Choi)}. \tag{10}$$

We again use the example in Section 3.1 to see why this superoperator is useful. Suppose ρ is the "+" state defined in (3). If $\mathcal{M}_i = \{|x\rangle\langle x|\}_{x=0}^{d-1}$, then $\bar{\mathcal{H}}(\cdot) = \sum_{x=0}^{d-1} |x\rangle\langle x|(\cdot)|x\rangle\langle x|$. It turns out that $\rho - \rho_{\mathrm{mm}}$ exactly falls into the 0-eigenspace of $\bar{\mathcal{H}}$,

$$\begin{aligned}\bar{\mathcal{H}}(\rho - \rho_{\mathrm{mm}}) &= \sum_{x=0}^{d-1} |x\rangle\langle x|(\rho - \rho_{\mathrm{mm}})|x\rangle\langle x| = \sum_{x=0}^{d-1} |x\rangle\langle x|\phi\rangle\langle\phi|x\rangle\langle x| - \sum_{x=0}^{d-1} |x\rangle\langle x|\frac{\mathbb{I}_d}{d}|x\rangle\langle x| \\ &= \sum_{x=0}^{d-1} |x\rangle\langle x|\frac{1}{d} - \frac{\mathbb{I}_d}{d} = 0.\end{aligned}$$

The third equality holds because $\langle x|\phi\rangle = \langle\phi|x\rangle = 1/\sqrt{d}$. This serves as an intuitive example that the eigenvalues of $\bar{\mathcal{H}}$ superoperator characterize the ability of the measurement scheme to distinguish between quantum states. If the difference $\rho - \rho_{\text{mm}}$ falls into the eigenspace of $\bar{\mathcal{H}}$ with small eigenvalues, then we can expect that the two states are hard to distinguish.

To formalize the intuition, we compute the chi-square divergence (7) between the outputs of the measurements in the cases when the input is the maximally mixed state, versus the case when it is chosen from an ε-perturbation. In our main technical result Lemma 8, we upper bound the divergence in terms of the average Lüders channel $\bar{\mathcal{H}}$. Thus, choosing Δ_σ from a subspace with small eigenvalues yields a small chi-square divergence and thus leads to tight copy complexity lower bounds.

Lemma 8 *Let σ, σ' be independently drawn from a distribution $\mathcal{D}$, and $\mathcal{M}_i$ be rank-1 POVM as in (1) for $i = 1, \ldots, n$. Define $\Delta_\sigma = \sigma - \rho_{mm}$. Then*

$$\mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{mm}}) \leq \mathbb{E}_{\sigma,\sigma'\sim\mathcal{D}}\left[\exp\left\{nd\left\langle\Delta_\sigma, \bar{\mathcal{H}}(\Delta_{\sigma'})\right\rangle\right\}\right] - 1 \tag{11}$$

where $\bar{\mathcal{H}}$ is the average Lüders channel defined in Eq. (10).

Proof The proof uses ideas from the decoupled chi-square fluctuations introduced in Acharya et al. (2020b). We can directly bound the chi-square distance using the following lemma which is from Pollard (2003).

Lemma 9 (Pollard (2003),(Acharya et al., 2020b, Lemma III.8)) *Let $\mathbf{P} = \mathbf{p}^{(1)} \otimes \cdots \otimes \mathbf{p}^{(n)}$ be a fixed product distribution and $\mathbf{Q}_\theta = \mathbf{q}_\theta^{(1)} \otimes \cdots \otimes \mathbf{q}_\theta^{(n)}$ be parameterized by a random variable θ. Then*

$$\mathrm{d}_{\chi^2}(\mathbb{E}_\theta[\,\mathbf{Q}_\theta\,] \,||\, \mathbf{P}) = \mathbb{E}_{\theta,\theta'}\left[\prod_{i=1}^n (1 + H_i(\theta, \theta'))\right] - 1,$$

where θ' is an independent copy of θ and

$$H_i(\theta, \theta') := \mathbb{E}_{x\sim\mathbf{p}^{(i)}}\left[\,\delta_\theta^{(i)}(x)\delta_{\theta'}^{(i)}(x)\,\right], \quad \delta_\theta^{(i)}(x) := \frac{\mathbf{q}_\theta^{(i)}(x) - \mathbf{p}^{(i)}(x)}{\mathbf{p}^{(i)}(x)}.$$

In our problem, $\mathbf{P}$ will be $\mathbf{P}_{\rho_{\text{mm}}}$, the distribution over the output of measurements across the n copies when the underlying state is maximally mixed, and $\mathbb{E}_\theta[\,\mathbf{Q}_\theta\,]$ will be $\mathbf{P}_\sigma$, the mixture distribution over the output of measurements when the underlying state is parameterized by a random density matrix σ induced by the perturbation. These are defined in (2).

We first compute the necessary quantities by appropriate substitution. Recall that $\mathbf{p}_\rho^i(\cdot)$ is the output distribution of the measurement on the ith copy.

$$\delta_\sigma^i(x) = \frac{\mathbf{p}_\sigma^i(x) - \mathbf{p}_{\rho_{\text{mm}}}^i(x)}{\mathbf{p}_{\rho_{\text{mm}}}^i(x)}, \; x \in [k].$$

We now evaluate $H_i(\sigma, \sigma')$ by expanding the probabilities using Born's rule.

$$H_i(\sigma, \sigma') = \mathbb{E}_{x\sim\mathbf{p}_{\rho_{\text{mm}}}^i}\left[\frac{(\mathbf{p}_\sigma^i(x) - \mathbf{p}_{\rho_{\text{mm}}}^i(x))(\mathbf{p}_{\sigma'}^i(x) - \mathbf{p}_{\rho_{\text{mm}}}^i(x))}{(\mathbf{p}_{\rho_{\text{mm}}}^i(x))^2}\right]$$

$$
\begin{aligned}
&= \sum_x \frac{(\mathbf{p}^i_\sigma(x) - \mathbf{p}^i_{\rho_{\text{mm}}}(x))(\mathbf{p}^i_{\sigma'}(x) - \mathbf{p}^i_{\rho_{\text{mm}}}(x))}{\mathbf{p}^i_{\rho_{\text{mm}}}(x)} \\
&= \sum_x \frac{\langle \psi^i_x | \Delta_\sigma | \psi^i_x \rangle \langle \psi^i_x | \Delta_{\sigma'} | \psi^i_x \rangle}{\langle \psi^i_x | \psi^i_x \rangle / d}.
\end{aligned}
$$

This expression can now be related to the Lüders channel. Adding trace to the numerator does not change the value, and from this we can apply cyclicity and linearity of trace,

$$
\begin{aligned}
H_i(\sigma, \sigma') &= d\left(\sum_x \frac{\operatorname{Tr}[\Delta_\sigma |\psi^i_x\rangle \langle \psi^i_x | \Delta_{\sigma'} | \psi^i_x \rangle \langle \psi^i_x |]}{\langle \psi^i_x | \psi^i_x \rangle} \right) \\
&= d \cdot \operatorname{Tr}\left[\sum_x \frac{\Delta_\sigma |\psi^i_x\rangle \langle \psi^i_x | \Delta_{\sigma'} | \psi^i_x \rangle \langle \psi^i_x |}{\langle \psi^i_x | \psi^i_x \rangle} \right] \\
&= d \cdot \operatorname{Tr}\left[\Delta_\sigma \sum_x \frac{|\psi^i_x\rangle \langle \psi^i_x | \Delta_{\sigma'} | \psi^i_x \rangle \langle \psi^i_x |}{\langle \psi^i_x | \psi^i_x \rangle} \right] \\
&= d \cdot \operatorname{Tr}\left[\Delta_\sigma \mathcal{H}_i(\Delta_{\sigma'}) \right] = d \left\langle \Delta_\sigma, \mathcal{H}_i(\Delta_{\sigma'}) \right\rangle \in \mathbb{R},
\end{aligned}
$$

where the last step uses the fact that Δ_σ is Hermitian.

Then, using Lemma 9, and the fact that $1 + x \leq \exp(x)$, we obtain

$$
\begin{aligned}
\mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\, \mathbf{P}_\sigma \,] \,||\, \mathbf{P}_{\rho_{\text{mm}}}) &= \mathbb{E}_{\sigma,\sigma'}\left[\prod_{i=1}^n (1 + H_i(\sigma, \sigma')) \right] - 1 \\
&\leq \mathbb{E}_{\sigma,\sigma'}\left[\exp\left\{ \sum_{i=1}^n H_i(\sigma, \sigma') \right\} \right] - 1 \\
&= \mathbb{E}_{\sigma,\sigma'}\left[\exp\left\{ d \sum_{i=1}^n \left\langle \Delta_\sigma, \mathcal{H}_i(\Delta_\sigma) \right\rangle \right\} \right] - 1.
\end{aligned}
$$

By linearity of the Hibert-Schmidt inner product and definition of $\bar{\mathcal{H}}$,

$$
\begin{aligned}
\mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\, \mathbf{P}_\sigma \,] \,||\, \mathbf{P}_{\rho_{\text{mm}}}) &\leq \mathbb{E}_{\sigma,\sigma'}\left[\exp\left\{ nd \left\langle \Delta_{\sigma'}, \frac{1}{n} \sum_{i=1}^n \mathcal{H}_i(\Delta_\sigma) \right\rangle \right\} \right] - 1 \\
&= \mathbb{E}_{\sigma,\sigma'}\left[\exp\{ nd \left\langle \Delta_\sigma, \bar{\mathcal{H}}(\Delta_{\sigma'}) \right\rangle \} \right] - 1.
\end{aligned}
$$

Using homomorphism $\operatorname{vec}(\mathcal{H}_{\mathcal{M}}(X)) = \mathcal{C}_{\mathcal{M}} \operatorname{vec}(X)$, we have $\left\langle \Delta_\sigma, \bar{\mathcal{H}}(\Delta_{\sigma'}) \right\rangle = \operatorname{vec}(\Delta_\sigma) \bar{\mathcal{C}} \operatorname{vec}(\Delta_{\sigma'})$, completing the proof. ■

Explaining the example in Section 3.1. We now use Lemma 8 to explain why choosing a fixed basis measurement $\{|x\rangle\langle x|\}_{x=0}^{d-1}$ for all copies as in Section 3.1 would fail. Since there are only d rank-1 projectors, the rank of $\bar{\mathcal{C}}$ is d, but $\bar{\mathcal{C}}$ has a dimension of $d^2 \times d^2$ and thus there are a total of $d^2 - d$ eigenvectors with 0 eigenvalues. From Proposition 3, we know that there must exist a trace-0 Δ in the 0-eigenspace such that $\sigma = \rho_{\text{mm}} + \Delta \in \mathcal{P}_\varepsilon$. For this particular σ the upper bound in (11) is 0, and thus it is impossible to distinguish ρ_{mm} and σ. This is consistent with the discussion in Section 3.1.

We can make a more general argument that to avoid the catastrophic failure similar to the dummy example in Section 3.1, $\bar{\mathcal{C}}$ has to be nearly full-rank: $\mathrm{rank}(\bar{\mathcal{C}}) \geq (1-o(1))d^2$. Thus $(1-o(1))d^2$ linearly independent rank-1 projectors are needed in all the POVMs. Indeed if otherwise, the dimension of the 0-eigenspace of $\bar{\mathcal{H}}$ is $\Omega(d^2)$, we can again invoke Proposition 3 (perhaps with some different constants) to find a *single fixed* σ that completely fools the measurement scheme.

Remark 10 *One can show that $\bar{\mathcal{H}}$ is the Lüders channel of a POVM $\mathcal{M} := \{\frac{1}{n}M_x^i\}_{x\in[k],i\in[n]}$ which is the ensemble of all measurements. One can define $\mathcal{M}^\dagger\mathcal{M}$ where we slightly abused the notation and treated $\mathcal{M} : \mathbb{H}_d \mapsto \mathbb{R}^k$ as a linear mapping from quantum states to probability vectors. $\bar{\mathcal{H}}$ and $\mathcal{M}^\dagger\mathcal{M}$ are similar but slightly different superoperators*[7]*. Guţă et al. (2020) used $\mathcal{M}^\dagger\mathcal{M}$ to derive upper bounds for quantum tomography for three specific types of measurements. Our result is orthogonal to their work in that we prove lower bounds for general rank-1 measurements.*

Applying Lemma 8 to our hard case construction in Definition 2, we can relate the eigenvalues of $\bar{\mathcal{H}}$ to the max-min and min-max distances in (8) (9) in Theorem 11. The proof is in Section C.

Theorem 11 *When $n = O(d^2/\varepsilon^2)$, the max-min chi-square divergence can be bounded as*

$$\max_{\mathcal{M}^n \text{ fixed}} \min_{\mathcal{D}\in\Gamma_\varepsilon} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{mm}}) = O\left(\frac{n^2\varepsilon^4}{d^4}\cdot\frac{\max_{\bar{\mathcal{H}}}\left\|\bar{\mathcal{H}}\right\|_1^2}{d^2}\right), \tag{12}$$

When $n = O(d^{3/2}/\varepsilon^2)$, the min-max chi-square divergence can be bounded as

$$\min_{\mathcal{D}\in\Gamma_\varepsilon} \max_{\mathcal{M}^n \text{ fixed}} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{mm}}) = O\left(\frac{n^2\varepsilon^4}{d^4}\max_{\bar{\mathcal{H}}}\left\|\bar{\mathcal{H}}\right\|_{HS}^2\right). \tag{13}$$

Different bounds are a result of how the basis $\mathcal{V}$ is chosen in Definition 2. To upper bound the min-max divergence, we choose $\mathcal{V}$ to be an arbitrary fixed basis. To upper bound the max-min divergence, we choose $\ell = d^2/2$ and $\mathcal{V}$ to be the eigenbasis of $\bar{\mathcal{H}}$, which has important properties stated in Lemma 12. These are standard results and we state their proofs in Section A.2.1.

Lemma 12 *$\bar{\mathcal{H}}$ has an orthonormal eigenbasis $\mathcal{V}_{\bar{\mathcal{H}}} = (V_1, \ldots, V_{d^2})$ with eigenvalues $0 \leq \lambda_1 \leq \ldots \leq \lambda_{d^2} = 1$ where $V_i \in \mathbb{C}^{d\times d}$ is trace-0 Hermitian for $i \leq d^2-1$ and $V_{d^2} = \mathbb{I}_d/\sqrt{d}$. Furthermore, $\mathrm{Tr}[\bar{\mathcal{H}}] = \sum_{i=1}^{d^2}\lambda_i = d$.*

Using Lemma 12, the copy complexity lower bounds for both fixed and randomized schemes are immediate corollaries of Theorem 11. From Lemma 12, we have $\left\|\bar{\mathcal{H}}\right\|_1 = d$. Moreover, $\left\|\bar{\mathcal{H}}\right\|_{\mathrm{HS}}^2 = \sum_{i=1}^{d^2}\lambda_i^2 \leq (\max_i \lambda_i)\sum_{i=1}^{d^2}\lambda_i \leq d$ (since $\lambda_i \leq 1$). Thus,

$$\frac{\max_{\bar{\mathcal{H}}}\left\|\bar{\mathcal{H}}\right\|_1^2}{d^2} = 1, \quad \max_{\bar{\mathcal{H}}}\left\|\bar{\mathcal{H}}\right\|_{\mathrm{HS}}^2 \leq d.$$

Combining (8) and (12), we conclude that for fixed measurements $n = \Omega(d^2/\varepsilon^2)$ and prove Theorem 5. Combining (9) and (13), we recover the $n = \Omega(d^{3/2}/\varepsilon^2)$ lower bound for randomized non-adaptive schemes, which was shown in Bubeck et al. (2020).

7. The differ by a scalar factor if $|\psi_x^i\rangle$ have equal norms, but can be very different otherwise.

Acknowledgments

This work was supported by NSF-CCF-1846300 (CAREER) grant, NSF-CCF-1815893, and by a Google Research Scholar Program. The authors thank Mark Wilde for helpful discussions.

References

Scott Aaronson. Shadow tomography of quantum states. *SIAM J. Comput.*, 49(5), 2020. doi: 10.1137/18M120275X. URL https://doi.org/10.1137/18M120275X.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints II: Communication constraints and shared randomness. *IEEE Trans. Inform. Theory*, 66(12):7856–7877, 2020a. ISSN 0018-9448. doi: 10.1109/TIT.2020.3028439. URL https://doi.org/10.1109/TIT.2020.3028439. Available at abs/1905.08302.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints I: lower bounds from chi-square contraction. *IEEE Trans. Inf. Theory*, 66(12):7835–7855, 2020b. doi: 10.1109/TIT.2020.3028440. URL https://doi.org/10.1109/TIT.2020.3028440.

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi. Inference under information constraints iii: Local privacy constraints. *IEEE Journal on Selected Areas in Information Theory*, 2(1):253–267, 2021. doi: 10.1109/JSAIT.2021.3053569.

Jayadev Acharya, Clément L. Canonne, Yuhan Liu, Ziteng Sun, and Himanshu Tyagi. Interactive inference under information constraints. *IEEE Transactions on Information Theory*, 68(1):502–516, 2022. doi: 10.1109/TIT.2021.3123905.

Costin Badescu and Ryan O'Donnell. Improved quantum data analysis. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1398–1411. ACM, 2021. doi: 10.1145/3406325.3451109. URL https://doi.org/10.1145/3406325.3451109.

Costin Badescu, Ryan O'Donnell, and John Wright. Quantum state certification. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 503–514. ACM, 2019. doi: 10.1145/3313276.3316344. URL https://doi.org/10.1145/3313276.3316344.

Leighton Pate Barnes, Yanjun Han, and Ayfer Ozgur. Lower bounds for learning distributions under communication constraints via fisher information. *Journal of Machine Learning Research*, 21(236):1–30, 2020. URL http://jmlr.org/papers/v21/19-737.html.

Tugkan Batu, Lance Fortnow, Eldar Fischer, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 442–451. IEEE Computer Society, 2001. doi: 10.1109/SFCS.2001.959920. URL https://doi.org/10.1109/SFCS.2001.959920.

Fernando G. S. L. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu. Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, volume 132 of *LIPIcs*, pages 27:1–27:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. doi: 10.4230/LIPICS.ICALP.2019.27. URL https://doi.org/10.4230/LIPIcs.ICALP.2019.27.

Fernando G. S. L. Brandão, Aram W. Harrow, James R. Lee, and Yuval Peres. Adversarial hypothesis testing and a quantum stein's lemma for restricted measurements. *IEEE Transactions on Information Theory*, 66(8):5037–5054, 2020. doi: 10.1109/TIT.2020.2979704.

Sébastien Bubeck, Sitan Chen, and Jerry Li. Entanglement is necessary for optimal quantum property testing. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 692–703. IEEE, 2020. doi: 10.1109/FOCS46700.2020.00070. URL https://doi.org/10.1109/FOCS46700.2020.00070.

Tony Cai, Donggyu Kim, Yazhen Wang, Ming Yuan, and Harrison H. Zhou. Optimal large-scale quantum state tomography with Pauli measurements. *The Annals of Statistics*, 44(2):682 – 712, 2016. doi: 10.1214/15-AOS1382. URL https://doi.org/10.1214/15-AOS1382.

Clément L. Canonne. Topics and techniques in distribution testing: A biased but representative sample. *Foundations and Trends® in Communications and Information Theory*, 19(6):1032–1198, 2022. ISSN 1567-2190. doi: 10.1561/0100000114. URL http://dx.doi.org/10.1561/0100000114.

Siu On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1193–1203. SIAM, 2014. doi: 10.1137/1.9781611973402.88. URL https://doi.org/10.1137/1.9781611973402.88.

Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 574–585. IEEE, 2021. doi: 10.1109/FOCS52979.2021.00063. URL https://doi.org/10.1109/FOCS52979.2021.00063.

Sitan Chen, Jerry Li, Brice Huang, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 1205–1213. IEEE, 2022a. doi: 10.1109/FOCS54457.2022.00118. URL https://doi.org/10.1109/FOCS54457.2022.00118.

Sitan Chen, Jerry Li, and Ryan O'Donnell. Toward instance-optimal state certification with incoherent measurements. In Po-Ling Loh and Maxim Raginsky, editors, *Conference on Learning Theory, 2-5 July 2022, London, UK*, volume 178 of *Proceedings of Machine Learning Re-*

search, pages 2541–2596. PMLR, 2022b. URL https://proceedings.mlr.press/v178/chen22b.html.

Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning? In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 391–404. IEEE, 2023. doi: 10.1109/FOCS57990.2023.00029. URL https://doi.org/10.1109/FOCS57990.2023.00029.

Benoît Collins. Moments and cumulants of polynomial random variables on unitarygroups, the itzykson-zuber integral, and free probability. *International Mathematics Research Notices*, 2003 (17):953–982, 2003.

Benoît Collins and Piotr Śniady. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

John B DeBrota. *Informationally Complete Measurements and Optimal Representations of Quantum Theory*. PhD thesis, University of Massachusetts Boston, 2020.

John B DeBrota and Blake C Stacey. Lüders channels and the existence of symmetric-informationally-complete measurements. *Physical Review A*, 100(6):062327, 2019.

Ilias Diakonikolas and Daniel M. Kane. A new approach for testing properties of discrete distributions. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 685–694. IEEE Computer Society, 2016. doi: 10.1109/FOCS.2016.78. URL https://doi.org/10.1109/FOCS.2016.78.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 429–438. IEEE Computer Society, 2013. doi: 10.1109/FOCS.2013.53. URL https://doi.org/10.1109/FOCS.2013.53.

Thomas Durt, Berthold-Georg Englert, Ingemar Bengtsson, and Karol Życzkowski. On mutually unbiased bases. *International journal of quantum information*, 8(04):535–640, 2010.

Omar Fawzi, Nicolas Flammarion, Aurélien Garivier, and Aadil Oufkir. On Adaptivity in Quantum Testing. *Transactions on Machine Learning Research Journal*, pages 1–33, September 2023. URL https://hal.science/hal-04107265.

Steven T. Flammia and Yi-Kai Liu. Direct fidelity estimation from few pauli measurements. *Phys. Rev. Lett.*, 106:230501, Jun 2011. doi: 10.1103/PhysRevLett.106.230501. URL https://link.aps.org/doi/10.1103/PhysRevLett.106.230501.

Steven T. Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *CoRR*, abs/2305.18519, 2023. doi: 10.48550/ARXIV.2305.18519. URL https://doi.org/10.48550/arXiv.2305.18519.

Madalin Guţă, Jonas Kahn, Richard Kueng, and Joel A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, 2020.

Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Theory*, 63(9):5628–5641, 2017. doi: 10.1109/TIT.2017.2719044. URL https://doi.org/10.1109/TIT.2017.2719044.

Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Minimax estimation of discrete distributions under l1 loss. *IEEE Transactions on Information Theory*, 61(11):6343–6354, 2015.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

Sumeet Khatri and Mark M Wilde. Principles of quantum communication theory: A modern approach. 2021. URL https://markwilde.com/PQCT-khatri-wilde.pdf.

Andreas Klappenecker and Martin Rotteler. Mutually unbiased bases are complex projective 2-designs. In *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, pages 1740–1744. IEEE, 2005.

Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017.

L. LeCam. Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, 1 (1):38–53, 1973. ISSN 00905364. URL http://www.jstor.org/stable/2958155.

Yi-Kai Liu. Universal low-rank matrix recovery from pauli measurements. In John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, pages 1638–1646, 2011. URL https://proceedings.neurips.cc/paper/2011/hash/e820a45f1dfc7b95282d10b6087e11c0-Abstract.html.

Gerhart Lüders. Über die zustandsänderung durch den meßprozeß. *Annalen der Physik*, 443(5-8): 322–328, 1950.

Ryan O'Donnell and John Wright. Quantum spectrum testing. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 529–538. ACM, 2015. doi: 10.1145/2746539.2746582. URL https://doi.org/10.1145/2746539.2746582.

Ryan O'Donnell and John Wright. Efficient quantum tomography. In Daniel Wichs and Yishay Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 899–912. ACM, 2016. doi: 10.1145/2897518.2897544. URL https://doi.org/10.1145/2897518.2897544.

Ryan O'Donnell and John Wright. Efficient quantum tomography II. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 962–974. ACM, 2017. doi: 10.1145/3055399.3055454. URL https://doi.org/10.1145/3055399.3055454.

T. Ogawa and H. Nagaoka. Strong converse and stein's lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory*, 46(7):2428–2433, 2000. doi: 10.1109/18.887855.

Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Trans. Inf. Theory*, 54(10):4750–4755, 2008. doi: 10.1109/TIT.2008.928987. URL https://doi.org/10.1109/TIT.2008.928987.

David Pollard. Asymptopia, 2003. URL http://www.stat.yale.edu/~pollard/Books/Asymptopia/. Manuscript.

Bartosz Regula, Ludovico Lami, and Mark M. Wilde. Postselected quantum hypothesis testing. *IEEE Transactions on Information Theory*, pages 1–1, 2023. doi: 10.1109/TIT.2023.3299870.

Joseph M Renes, Robin Blume-Kohout, Andrew J Scott, and Carlton M Caves. Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics*, 45(6):2171–2180, 2004.

Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Society, 2023.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Don Weingarten. Asymptotic behavior of group integrals in the limit of infinite rank. *Journal of Mathematical Physics*, 19(5):999–1001, 1978.

John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016.

Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, pages 423–435. Springer, 1997.

Nengkun Yu. Sample efficient identity testing and independence testing of quantum states. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference*, volume 185 of *LIPIcs*, pages 11:1–11:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPICS.ITCS.2021.11. URL https://doi.org/10.4230/LIPIcs.ITCS.2021.11.

Nengkun Yu. Almost tight sample complexity analysis of quantum identity testing by pauli measurements. *IEEE Transactions on Information Theory*, 69(8):5060–5068, 2023. doi: 10.1109/TIT.2023.3271206.

Gerhard Zauner. Grundzüge einer nichtkommutativen designtheorie. *Ph. D. dissertation, PhD thesis*, 1999.

Appendix A. Proofs of technical lemmas in Section 4

A.1. Proof of Lemma 7

Proof Recall that $Y = 0$ and $\rho = \rho_{\text{mm}}$ with probability 1/2 and $Y = 1$ and $\rho \sim \mathcal{D}$ with probability $1/2$. In the former case when the state is ρ_{mm} and $Y = 0$, then the tester outputs the correct answer with probability at least $2/3$,

$$\Pr[\hat{Y} = 0 | Y = 0] \geq 2/3.$$

When $\rho \sim \mathcal{D}$, note that by the definition of almost-ε perturbations, the probability that $\|\sigma_z - \rho_{\text{mm}}\|_1 > \varepsilon$ is at least $4/5$. Denote this event as E, then $\Pr[E|Y = 1] \geq 4/5$. We can lower bound the success probability as

$$\Pr[\hat{Y} = 1 | Y = 1] \geq \Pr[Y = 1 | E, Y = 1)] \Pr[E | Y = 1] \geq \frac{2}{3} \cdot \frac{4}{5} = \frac{8}{15}.$$

Combining the two parts,

$$\Pr[Y = \hat{Y}] = \frac{1}{2} \Pr[\hat{Y} = 0 | Y = 0] + \frac{1}{2} \Pr[\hat{Y} = 1 | Y = 1] \geq \frac{1}{2}\left(\frac{2}{3} + \frac{8}{15}\right) = \frac{3}{5}.$$

By standard argument on the distinguishability of two distributions (Yu, 1997, Lemma 1),

$$1 - \frac{3}{5} \geq \frac{1}{2}(1 - d_{\text{TV}}(\mathbb{E}_z[\mathbf{P}_{\sigma_z}], \mathbf{P}_{\rho_{\text{mm}}})) \implies d_{\text{TV}}(\mathbb{E}_z[\mathbf{P}_{\sigma_z}], \mathbf{P}_{\rho_{\text{mm}}}) \geq \frac{1}{5}.$$

Finally, the inequality follows by Pinsker's inequality and the relation between KL and chi-square divergences.

$$d_{\text{TV}}(\mathbb{E}_\sigma[\mathbf{P}_\sigma], \mathbf{P}_{\rho_{\text{mm}}}) \leq \sqrt{\frac{1}{2}\text{KL}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\mathbf{P}_\sigma] \,||\, \mathbf{P}_{\rho_{\text{mm}}})} \leq \sqrt{\frac{1}{2}\text{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\mathbf{P}_\sigma] \,||\, \mathbf{P}_{\rho_{\text{mm}}})}.$$

■

A.2. Proof of Lemma 12

Let us recall the lemma.

Lemma 12 *$\bar{\mathcal{H}}$ has an orthonormal eigenbasis $\mathcal{V}_{\bar{\mathcal{H}}} = (V_1, \ldots, V_{d^2})$ with eigenvalues $0 \leq \lambda_1 \leq \ldots \leq \lambda_{d^2} = 1$ where $V_i \in \mathbb{C}^{d\times d}$ is trace-0 Hermitian for $i \leq d^2 - 1$ and $V_{d^2} = \mathbb{I}_d/\sqrt{d}$. Furthermore, $\text{Tr}[\bar{\mathcal{H}}] = \sum_{i=1}^{d^2} \lambda_i = d$.*

The proof is broken into two parts. In A.2.1 we state some properties of superoperators, and in A.2.2 we provide a proof of the lemma.

A.2.1. IMPORTANT PROPERTIES OF $\mathcal{H}_{\mathcal{M}}$

We start with some useful definitions.

Definition 13 *Let $\mathcal{N} : \mathbb{C}^{d\times d} \mapsto \mathbb{C}^{d\times d}$ be a superoperator.*

1. *$\mathcal{N}$ is called Hermitian if $\mathcal{N} = \mathcal{N}^\dagger$.*
2. *$\mathcal{N}$ is Hermiticity preserving if for all Hermitian $X \in \mathbb{H}_d$, $\mathcal{N}(X)$ is also Hermitian.*
3. *$\mathcal{N}$ is trace-preserving if for all $X \in \mathbb{C}^{d\times d}$, $\mathrm{Tr}[X] = \mathrm{Tr}[\mathcal{N}(X)]$.*
4. *$\mathcal{N}$ is unital if $\mathcal{N}(\mathbb{I}_d) = \mathbb{I}_d$.*

We have the following fact about the Lüders channel.

Fact 14 *$\mathcal{H}_{\mathcal{M}}$ is a superoperator over $\mathbb{C}^{d\times d}$ that satisfies all properties in Definition 13.*

Proof The proof follows from Definition 13, Definition 4, and the definition of POVMs. Nevertheless, we provide the proof for completeness.

1. Hermitian:
$$\begin{aligned}\langle Y, \mathcal{H}_{\mathcal{M}}(X)\rangle &= \mathrm{Tr}\left[Y^\dagger \sum_x \frac{|\psi_x\rangle\langle\psi_x|X|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}\right] = \sum_x \mathrm{Tr}\left[\frac{Y^\dagger|\psi_x\rangle\langle\psi_x|X|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}\right] \\ &= \sum_x \mathrm{Tr}\left[\frac{|\psi_x\rangle\langle\psi_x|Y^\dagger|\psi_x\rangle\langle\psi_x|X}{\langle\psi_x|\psi_x\rangle}\right] = \mathrm{Tr}\left[\sum_x \frac{|\psi_x\rangle\langle\psi_x|Y^\dagger|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle} X\right] \\ &= \langle \mathcal{H}_{\mathcal{M}}(Y), X\rangle\end{aligned}$$
2. Hermiticity preserving: let X be Hermitian, then
$$\mathcal{H}_{\mathcal{M}}(X)^\dagger = \sum_x \frac{|\psi_x\rangle\langle\psi_x|X^\dagger|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle} = \sum_x \frac{|\psi_x\rangle\langle\psi_x|X|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle} = \mathcal{H}_{\mathcal{M}}(X)$$
3. Trace preserving:
$$\begin{aligned}\mathrm{Tr}\left[\mathcal{H}_{\mathcal{M}}(X)\right] &= \mathrm{Tr}\left[\sum_x \frac{|\psi_x\rangle\langle\psi_x|X|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}\right] \\ &= \sum_x \mathrm{Tr}\left[\frac{|\psi_x\rangle\langle\psi_x|X|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}\right] \\ &= \sum_x \langle\psi_x|X|\psi_x\rangle = \sum_x \mathrm{Tr}[|\psi_x\rangle\langle\psi_x|[X]] \\ &= \mathrm{Tr}\left[\sum_x |\psi_x\rangle\langle\psi_x|X\right] = \mathrm{Tr}[X].\end{aligned}$$
4. Unital:
$$\mathcal{H}_{\mathcal{M}}(\mathbb{I}_d) = \sum_x \frac{|\psi_x\rangle\langle\psi_x|\mathbb{I}_d|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle} = \sum_x |\psi_x\rangle\langle\psi_x| = \mathbb{I}_d.$$

∎

A.2.2. PROOF OF THE LEMMA

Hermitian eigenvectors. By linearity, $\bar{\mathcal{H}}$ satisfies all properties in Fact 14. Since $\bar{\mathcal{H}}$ is Hermiticity preserving, $\bar{\mathcal{H}}$ is also a linear superoperator over the subspace of all Hermitian matrices $\mathbb{H}_d$.

Since $\bar{\mathcal{H}}$ is a Hermitian operator on $\mathbb{H}_d$, the eigenvectors of $\bar{\mathcal{H}}$ form an orthonormal basis $\{V_i\}_{i=1}^{d^2}$ of $\mathbb{H}_d$. Note that $\mathbb{I}_d$ is an eigenvector of $\bar{\mathcal{H}}$ with eigenvalue 1 since

$$\bar{\mathcal{H}}\left(\mathbb{I}_d\right) = \frac{1}{n}\left(\sum_{i=1}^{n} \mathcal{H}_i(\mathbb{I}_d)\right) = \frac{1}{n}\left(\sum_{i=1}^{n} \mathbb{I}_d = \mathbb{I}_d\right).$$

We then set $V_{d^2} = \mathbb{I}_d/\sqrt{d}$. Thus, all other eigenvectors $V_1, \dots, V_{d^2-1}$ must lie in the space orthogonal to $\mathrm{span}\{\mathbb{I}_d\}$, which is exactly the space of trace-0 Hermitian matrices since

$$\langle A, \mathbb{I}_d\rangle = 0 \iff \mathrm{Tr}[A^\dagger \mathbb{I}_d] = \mathrm{Tr}[A^\dagger] = 0 = \mathrm{Tr}[A].$$

Non-negative eigenvalues. To show that all eigenvalues are non-negative, we just need to show that $\bar{\mathcal{H}}$ is positive semi-definite, i.e. for all matrix $X \in C^{d\times d}$,

$$\left\langle X, \bar{\mathcal{H}}(X)\right\rangle \geq 0.$$

Due to linearity, we just need to prove that each $\mathcal{H}_i$ as defined in 5 is p.s.d.,

$$\begin{aligned} \langle X, \mathcal{H}_i(X)\rangle &= \mathrm{Tr}\left[X^\dagger \sum_{x=1}^{k} \frac{|\psi_x^i\rangle\langle\psi_x^i|X|\psi_x^i\rangle\langle\psi_x^i|}{\langle\psi_x^i|\psi_x^i\rangle}\right] \\ &= \sum_{x=1}^{k} \frac{\mathrm{Tr}[X^\dagger|\psi_x^i\rangle\langle\psi_x^i|X|\psi_x^i\rangle\langle\psi_x^i|]}{\langle\psi_x^i|\psi_x^i\rangle} \\ &= \sum_{x=1}^{k} \frac{\langle\psi_x^i|X^\dagger|\psi_x^i\rangle\langle\psi_x^i|X|\psi_x^i\rangle}{\langle\psi_x^i|\psi_x^i\rangle} \\ &= \sum_{x=1}^{k} \frac{|\langle\psi_x^i|X|\psi_x^i\rangle|^2}{\langle\psi_x^i|\psi_x^i\rangle} \geq 0. \end{aligned} \tag{14}$$

The last line is due to

$$\overline{\langle\psi_x^i|X|\psi_x^i\rangle} = \langle\psi_x^i|X|\psi_x^i\rangle^\dagger = \langle\psi_x^i|X^\dagger|\psi_x^i\rangle.$$

Upper bound on eigenvalues. Finally, we show that all eigenvalues are at most 1. This is equivalent to $\left\|\bar{\mathcal{H}}\right\|_{\mathrm{op}} \leq 1$. By the convexity of norms, it suffices to prove that $\|\mathcal{H}_i\|_{\mathrm{op}} \leq 1$. Starting from (14),

$$\begin{aligned} \langle X, \mathcal{H}_i(X)\rangle &= \sum_{x=1}^{k} \frac{\mathrm{Tr}[X^\dagger|\psi_x^i\rangle\langle\psi_x^i|X|\psi_x^i\rangle\langle\psi_x^i|]}{\langle\psi_x^i|\psi_x^i\rangle} \\ &\leq \sum_{x=1}^{k} \frac{\sqrt{\mathrm{Tr}[X^\dagger|\psi_x^i\rangle\langle\psi_x^i|\psi_x^i\rangle\langle\psi_x^i|X]\,\mathrm{Tr}[|\psi_x^i\rangle\langle\psi_x^i|X^\dagger X|\psi_x^i\rangle\langle\psi_x^i|]}}{\langle\psi_x^i|\psi_x^i\rangle} && \text{Cauchy-Schwarz} \\ &= \sum_{x=1}^{k} \sqrt{\mathrm{Tr}[X^\dagger X|\psi_i^x\rangle\langle\psi_i^x|]\,\mathrm{Tr}[XX^\dagger|\psi_i^x\rangle\langle\psi_i^x|]} && \text{Cyclicity of trace} \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{x=1}^{k} \mathrm{Tr}\left[\frac{X^\dagger X + XX^\dagger}{2} |\psi_i^x\rangle\langle\psi_i^x|\right] && \text{AM-GM} \\
&= \mathrm{Tr}\left[\frac{X^\dagger X + XX^\dagger}{2} \sum_{x=1}^{k} |\psi_i^x\rangle\langle\psi_i^x|\right] \\
&= \mathrm{Tr}\left[\frac{X^\dagger X + XX^\dagger}{2}\right] && \text{POVM} \\
&= \mathrm{Tr}[X^\dagger X] \\
&= \langle X, X\rangle \,.
\end{aligned}
$$

Trace. Again due to linearity, we only need to prove that $\mathrm{Tr}[\mathcal{H}_l] = d$ for each $l = 1, \ldots, n$.

$$
\begin{aligned}
\mathrm{Tr}[\mathcal{H}_l] &= \sum_{i,j=1}^{d} \mathrm{Tr}\left[\sum_{x=1}^{k} \frac{|j\rangle\langle i|\psi_x^l\rangle\langle\psi_x^l|i\rangle\langle j|\psi_x^l\rangle\langle\psi_x^l|}{\langle\psi_x^l|\psi_x^l\rangle}\right] \\
&= \sum_{i,j=1}^{d}\sum_{x=1}^{k} \frac{\langle\psi_x^l|j\rangle\langle i|\psi_x^l\rangle\langle\psi_x^l|i\rangle\langle j|\psi_x^l\rangle}{\langle\psi_x^l|\psi_x^l\rangle} \\
&= \sum_{x=1}^{k} \frac{1}{\langle\psi_x^l|\psi_x^l\rangle} \sum_{i=1}^{d} |\langle i|\psi_x^l\rangle|^2 \sum_{j=1}^{d} |\langle j|\psi_x^l\rangle|^2 \\
&= \sum_{x=1}^{k} \frac{\langle\psi_x^l|\psi_x^l\rangle^2}{\langle\psi_x^l|\psi_x^l\rangle} \\
&= \sum_{x=1}^{k} \langle\psi_x^l|\psi_x^l\rangle = d.
\end{aligned}
$$

The final equality is due to $\sum_x |\psi_x^l\rangle\langle\psi_x^l| = \mathbb{I}_d$ and thus $\mathrm{Tr}[\sum_x |\psi_x^l\rangle\langle\psi_x^l|] = \sum_{x=1}^{k}\langle\psi_x^l|\psi_x^l\rangle = \mathrm{Tr}[\mathbb{I}_d] = d$

Appendix B. Proof of Proposition 3

The central claim is Theorem 15 which states that the operator norm of a random matrix with independently perturbed orthogonal components is $O(\sqrt{d})$ with high probability. The proof is in Section B.1.

Theorem 15 *Let $V_1, \ldots, V_{d^2} \in \mathbb{C}^{d\times d}$ be an orthonormal basis of $\mathbb{C}^{d\times d}$ and $z_1, \ldots, z_{d^2} \in \{-1, 1\}$ be independent symmetric Bernoulli random variables. Let $W = \sum_{i=1}^{\ell} z_i V_i$ where $\ell \leq d^2$. For all $\alpha > 0$, there exists κ_α, which is increasing in α such that*

$$
\Pr\Big[\, \|W\|_{op} > \kappa_\alpha\sqrt{d}\,\Big] \leq 2\exp\{-\alpha d\}.
$$

Remark 16 *Standard random matrix theory (e.g. Tao (2023)[Corollary 2.3.5]) states that if each entry of W is independent and uniform from $\{-1, 1\}$, i.e. $W = \sum_{i,j} z_{ij} E_{ij}$ where E_{ij} is a matrix with 1 at position (i, j) and 0 everywhere else, then $\|W\|_{op} = O(\sqrt{d})$ with high probability. Theorem 15 generalizes this argument to arbitrary basis $\{V_i\}_{i=1}^{d^2}$. This could be of independent interest.*

Proposition 3 is an immediate corollary of Theorem 15.

Proposition 3 *Let $d^2/2 \le \ell \le d^2-1$. Let z be drawn uniformly from $\{-1,1\}^\ell$, and Δ_z, σ_z are as defined in Definition 2. Then, there exists a universal constant $c \le 10\sqrt{2}$, such that for $\varepsilon < \frac{1}{c^2}$, with probability at least $1-2\exp(-d)$, $\|\Delta_z\|_{op} \le 1/d$ and $\|\Delta_z\|_1 \ge \varepsilon$.*

Proof By Hölder's inequality, we have that for all matrices A,

$$\|A\|_{\text{op}}\|A\|_1 \ge \|A\|_{\text{HS}}^2.$$

Note that $\Delta_z = \frac{c\varepsilon}{\sqrt{d\ell}}W$ and $\|\Delta_z\|_{\text{HS}} = \frac{c\varepsilon}{\sqrt{d}}$. Thus setting $\alpha = 1$ and $\kappa = \kappa_1$ in Theorem 15, with probability at least $1-2\exp(-d)$,

$$\|\Delta_z\|_{\text{op}} \le \frac{c\varepsilon}{\sqrt{d\ell}} \cdot \kappa\sqrt{d} = \frac{c\kappa\varepsilon}{\sqrt{\ell}}.$$

This implies that

$$\|\Delta_z\|_1 \ge \|\Delta_z\|_{\text{HS}}^2/\|\Delta_z\|_{\text{op}} \ge \frac{c\varepsilon}{\kappa} \cdot \frac{\sqrt{\ell}}{d}.$$

In the proof of Theorem 15 in Section B.1, we can show that $\kappa = \kappa_1 \le 10$. Thus choosing $c = \sqrt{2}\kappa \le 10\sqrt{2}$, we guarantee that $\|\Delta_z\|_1 > \varepsilon$ due to $\ell \ge d^2/2$. As long as $\varepsilon \le \frac{1}{200}$, we have $\|\Delta_z\|_{\text{op}} \le 1/d$ and thus $\sigma_z = \rho_{\text{mm}} + \Delta_z$ is a valid density matrix. This completes the proof of Proposition 3. ∎

Different bounds for min-max and max-min divergences in Theorem 11 are due to whether or not nature can choose $\mathcal{V}$ dependent on $\bar{\mathcal{H}}$, which in turn depends on the measurements $\mathcal{M}^n$. For randomized schemes, we need to upper bound the min-max divergence, and we can simply choose a fixed $\mathcal{V}$ that is uniformly bad for all $\mathcal{M}^n$. For fixed measurements however, under the max-min framework, nature could choose the hard distribution depending on $\mathcal{M}^n$. Specifically, with $\mathcal{V} = \mathcal{V}_{\bar{\mathcal{H}}}$ and ℓ small, $\sigma_z - \rho_{\text{mm}}$ completely lies in an eigenspace of $\bar{\mathcal{H}}$ with the ℓ smallest eigenvalues, thus generalizing the intuition from the toy example in Section 3.1.

B.1. Proof of Theorem 15

Proof We first prove that for any fixed unit vector $x \in \mathbb{C}^d$, the norm of Wx is at most $O(\sqrt{d})$ with high probability. Then we use an ϵ-net argument to show that the probability is also high for *all* unit vectors. We start with the following lemma.

Lemma 17 *Let $\{z_i\}_{i=1}^{d^2}, \{V_i\}_{i=1}^{d^2}$ and W be defined in Theorem 15. Then there exists a universal constant c' for any fixed unit vector x and all $s > 0$,*

$$\Pr\left[\|Wx\|_2 \ge (1+s)\sqrt{d}\right] \le 2\exp\{-c's^2 d\}.$$

Proof Let $z = (z_1, \ldots, z_{d^2}) \in \mathbb{R}^{d^2}$, and $\Pi_\ell \in \mathbb{R}^{d^2 \times d^2}$ be a diagonal matrix with 1 in the first ℓ diagonal entries and 0 everywhere else. Then

$$Wx = \sum_{i=1}^{\ell} z_i V_i x = V_x \Pi_\ell z,$$

where

$$V_x := [V_1 x, \ldots, V_{d^2} x] \in \mathbb{C}^{d \times d^2}$$

which is an isometry, i.e. $V_x V_x^\dagger = \mathbb{I}_d$, as stated in Claim 22 which will be proved at the end of this section. Therefore,

$$\|V_x\|_{\text{op}} = 1, \quad \|V_x\|_{\text{HS}}^2 = \text{Tr}[V_x V_x^\dagger] = d.$$

From this, we can apply concentration for linear transforms of independent sub-Gaussian random variables.

Theorem 18 (Vershynin (2018, Theorem 6.3.2)) *Let $B \in \mathbb{C}^{m\times n}$ be a fixed $m \times n$ matrix and let $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ be a random vector with independent, mean zero, unit variance, and sub-Gaussian coordinates with Orlicz-2 norm $\|X_i\|_{\psi_2} \le K$. Then there exists a universal constant $C = \frac{3}{8}$ such that for all $t > 0$,*

$$\Pr\left[\, |\|BX\|_2 - \|B\|_{HS}| > t \,\right] \le 2\exp\left\{ -\frac{Ct^2}{K^4 \|B\|_{op}^2} \right\}.$$

Remark 19 *The original (Vershynin, 2018, Theorem 6.3.2) was stated for real matrix B. However, it is straightforward to extend the argument to complex B by considering $\tilde{B} = \begin{bmatrix} Re(B) \\ Im(B) \end{bmatrix}$. Then $\left\|\tilde{B}\right\|_{op} = \|B\|_{op}$, $\left\|\tilde{B}\right\|_{HS} = \|B\|_{HS}$, and $\|BX\|_2 = \|\tilde{B}X\|_2$.*

Setting $B = V_x \Pi_\ell$, we observe that

$$\|B\|_{\text{op}} \le \|V_x\|_{\text{op}} \|\Pi_\ell\|_{\text{op}} = 1, \quad \|B\|_{\text{HS}} \le \|V_x\|_{\text{HS}} = \sqrt{d}.$$

Thus, plugging $t = s\sqrt{d}$, and noting that $\|z_i\|_{\psi_2} = 1/\sqrt{\ln 2} = K$, we have

$$\Pr\left[\, \|Wx\|_2 > (1+s)\sqrt{d} \,\right] \le \Pr\left[\, \|Bz\|_2 > s\sqrt{d} + \|B\|_{\text{HS}} \,\right] \le 2\exp\left\{-Cd(\ln 2)^2 s^2\right\}.$$

Setting $c' = C(\ln 2)^2 = \frac{3(\ln 2)^2}{8}$ completes the proof. ■

We can then proceed to use the ϵ-net argument, which follows closely to (Tao, 2023, Section 2.3).

Lemma 20 ((Tao, 2023, Lemma 2.3.2)) *Let Σ be a maximal $1/2$-net of the unitary sphere, i.e., a maximal set of points that are separated from each other by at least $1/2$. Then for any matrix $M \in \mathbb{C}^{d\times d}$ and $\lambda > 0$,*

$$\Pr\left[\, \|M\|_{op} > \lambda \,\right] \le \sum_{y \in \Sigma} \Pr[\, \|My\|_2 > \lambda/2 \,].$$

By standard volume packing argument, the size of Σ is at most $\exp(O(d))$,

Lemma 21 ((Tao, 2023, Lemma 2.3.4)) *Let $\epsilon \in (0, 1)$ and let Σ be an ϵ-net of the unit sphere. Then $|\Sigma| \le (C'/\epsilon)^d$ where $C' = 3$.*

Thus with c' defined in Lemma 17 and C' defined in Lemma 21 we conclude that

$$\Pr\Big[\,\|W\|_{\mathrm{op}} > 2(1+s)\sqrt{d}\,\Big] \le 2(2C')^d \exp\{-c's^2 d\} = 2\exp\left\{-(c's^2 - \ln(2C'))d\right\}.$$

Thus choosing s sufficiently large, we can guarantee that the tail probability decays exponentially in d. Specifically, let $\alpha > 0$ and $s^2 = \frac{\alpha+\ln(2C')}{c'}$, then we have

$$\Pr\Big[\,\|W\|_{\mathrm{op}} > 2(1+s)\sqrt{d}\,\Big] \le 2e^{-\alpha d}.$$

Setting $\kappa_\alpha = 2(1+s) = 2\left(1+\sqrt{\frac{\alpha+\ln(2C')}{c'}}\right)$ proves the theorem. In particular, $\kappa_1 \le 10$ when substituting the values of c' and C'. ■

We end this section with the proof of the isometry claim.

Claim 22 *Let $V_1, \dots, V_{d^2}$ be an orthonormal basis of $\mathbb{C}^{d\times d}$ and $x \in \mathbb{C}^d$ be a unit vector. Then $V_x := [V_1 x, \dots, V_{d^2} x] \in \mathbb{C}^{d\times d^2}$ is an isometry: $V_x V_x^\dagger = \mathbb{I}_d$.*

Proof Let $V_x^{(k)}$ be the kth row of V_x written as row vector. It suffices to prove that

$$V_x^{(k)}(V_x^{(l)})^\dagger = \delta_{kl}$$

Let $V_i^{(k)}$ be the kth row of V_i, written as a row vector. Then the kth element of $V_i x$ is

$$v_i^{(k)} := V_i^{(k)} x.$$

Since $V_1, \dots, V_{d^2}$ are orthonormal, we know that

$$V := [\mathrm{vec}(V_1), \dots, \mathrm{vec}(V_{d^2})]$$

is a unitary matrix in $\mathbb{C}^{d^2\times d^2}$. Let V^j be the jth row of V, then because V is unitary, the vector dot product $\langle V^j, V^i\rangle = \delta_{ij}$. Let

$$V^{(k)} = [(V^k)^\dagger, (V^{k+d})^\dagger, \dots (V^{k+d(j-1)})^\dagger, \dots, (V^{k+d(d-1)})^\dagger]^\dagger$$

which picks out the kth row of all $V_1, \dots, V_{d^2}$. Then, we have

$$V^{(k)} = [(V_1^{(k)})^\top, \dots, (V_{d^2}^{(k)})^\top].$$

Thus,

$$\sum_{i=1}^{d^2} (V_i^{(k)})^\dagger V_i^{(k)} = \overline{V^{(k)}(V^{(k)})^\dagger} = \mathbb{I}_d,$$

and for $k \ne l$,

$$\sum_{i=1}^{d^2} (V_i^{(k)})^\dagger V_i^{(l)} = \overline{V^{(k)}(V^{(l)})^\dagger} = 0.$$

Therefore,

$$V_x^{(k)}(V_x^{(l)})^\dagger = \sum_{i=1}^{d^2} v_i^{(k)}(v_i^{(l)})^\dagger = \sum_{i=1}^{d^2} x^\dagger (V_i^{(l)})^\dagger V_i^{(k)} x = x^\dagger \delta_{kl}\mathbb{I}_d x = \delta_{kl},$$

exactly as desired, completing the proof. ■

Appendix C. Proof of Theorem 11

Let $\mathcal{V} = (V_1, \ldots, V_{d^2} = \mathbb{I}_d/\sqrt{d})$ be an orthonormal basis of $\mathbb{H}_d$. We now upper bound the expression (11) in Lemma 8 when $\mathcal{D} = \mathcal{D}_\ell(\mathcal{V})$, defined in Definition 2. The result is in Theorem 23. The central claim is that the chi-squared divergence is related to the Hilbert-Schmidt norm of the projection of $\bar{\mathcal{H}}$ onto the subspace defined by $V_1, \ldots, V_\ell$.

Theorem 23 *Let $\frac{d^2}{2} \le \ell \le d^2-1$, $\mathcal{V} = (V_1, \ldots, V_{d^2} = \mathbb{I}_d/\sqrt{d})$ be an orthonormal basis of $\mathbb{H}_d$, $V := [vec(V_1), \ldots, vec(V_\ell)]$ and $\sigma_z, \sigma_{z'} \sim \mathcal{D}_\ell(\mathcal{V})$ defined in Definition 2. Then for $n < \frac{d^2}{6c^2\varepsilon^2}$,*

$$\mathbb{E}_{\sigma_z,\sigma_{z'}}\left[\exp\left\{nd\left\langle \bar{\Delta}_{z'}, \bar{\mathcal{H}}(\bar{\Delta}_z)\right\rangle\right\}\right] - 1 \le \exp\left\{\frac{c^2n^2\varepsilon^4}{\ell^2}\left\|V^\dagger \bar{\mathcal{C}} V\right\|_{HS}^2\right\} - 1 + \frac{4}{e^d}. \tag{15}$$

We now bound $\left\|V^\dagger\bar{\mathcal{C}}V\right\|_{\mathrm{HS}}^2$, which depends on how the basis $\mathcal{V}$ is chosen.

Observation 24 *For all orthonormal basis $\mathcal{V}$, we have $\left\|V^\dagger\bar{\mathcal{C}}V\right\|_{HS} \le \left\|\bar{\mathcal{H}}\right\|_{HS}$. However when $\mathcal{V} = \mathcal{V}_{\bar{\mathcal{H}}}$ in Lemma 12, for all $\frac{d^2}{2} \le \ell \le d^2-1$, $\left\|V^\dagger\bar{\mathcal{C}}V\right\|_{HS} = \left\|\bar{\mathcal{H}}_\ell\right\|_{HS} := \sqrt{\sum_{i=1}^{\ell}\lambda_i^2}$ and $0 \le \lambda_1 \le \ldots \le \lambda_{d^2} = 1$ are the eigenvalues of $\bar{\mathcal{H}}$.*

The proof of Theorem 23 is in Section C.1 and the proof of Observation 24 is in Section C.2. We can now prove Theorem 11. It is more straightforward to prove the min-max upper bound (9) by setting $\mathcal{V}$ as an arbitrary fixed basis that satisfies Definition 2. For example, one can choose the generalized Gell-Mann basis,

$$\begin{aligned}
\sigma_{0,0} &:= \frac{\mathbb{I}_d}{\sqrt{d}}, \\
\sigma_{k,l}^{(+)} &:= \frac{1}{\sqrt{2}}(|k\rangle\langle l| + |l\rangle\langle k|), && 0 \le k < l \le d-1, \\
\sigma_{k,l}^{(i)} &:= \frac{1}{\sqrt{2}}(-i|k\rangle\langle l| + i|l\rangle\langle k|), && 0 \le k < l \le d-1, \\
\sigma_{k,k} &:= \frac{k}{k+1}\left(-k|k\rangle\langle k| + \sum_{j=0}^{k-1}|j\rangle\langle j|\right), && 1 \le k \le d-1.
\end{aligned}$$

We can relabel them as $V_1, \ldots, V_{d^2}$ where $V_{d^2} = \sigma_{0,0}$. This is a natural extension of Pauli matrices for $d=2$. It can be easily verified that these d^2 matrices indeed form an orthonormal basis over $\mathbb{H}_d$. Using Lemma 8 and Theorem 23, setting $\ell = d^2-1$,

$$\begin{aligned}
\min_{\mathcal{D}\in\Gamma_\varepsilon}\max_{\mathcal{M}^n \text{ fixed}} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \;||\; \mathbf{P}_{\rho_{\mathrm{mm}}}) &\le \max_{\mathcal{M}^n \text{ fixed}} \mathrm{d}_{\chi^2}\left(\mathbb{E}_{\sigma\sim\mathcal{D}_\ell(\mathcal{V})}[\,\mathbf{P}_\sigma\,] \;||\; \mathbf{P}_{\rho_{\mathrm{mm}}}\right) \\
&\le O\left(\frac{n^2\varepsilon^4}{d^4}\max_{\bar{\mathcal{H}}}\left\|\bar{\mathcal{H}}\right\|_{\mathrm{HS}}^2\right).
\end{aligned}$$

When upper bounding the max-min divergence (12), we would have the freedom to choose a basis $\mathcal{V}$ that depends on $\bar{\mathcal{H}}$, which is determined by the measurement $\mathcal{M}^n$. More precisely, we can set $\mathcal{V} = \mathcal{V}_{\bar{\mathcal{H}}}$ and $\ell = d^2/2$, and the perturbations $\mathcal{D}_\ell(\mathcal{V}_{\bar{\mathcal{H}}})$ would be along directions that are least sensitive for the measurement scheme, which leads to the extra d factor in the chi-square divergence upper bound,

$$\max_{\mathcal{M}^n \text{ fixed}} \min_{\mathcal{D}\in\Gamma_\varepsilon} \mathrm{d}_{\chi^2}(\mathbb{E}_{\sigma\sim\mathcal{D}}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{\text{mm}}}) \le \max_{\mathcal{M}^n \text{ fixed}} \mathrm{d}_{\chi^2}\big(\mathbb{E}_{\sigma\sim\mathcal{D}_\ell(\mathcal{V}_{\bar{\mathcal{H}}})}[\,\mathbf{P}_\sigma\,] \,||\, \mathbf{P}_{\rho_{\text{mm}}}\big)$$
$$\le O\left(\frac{n^2\varepsilon^4}{d^4}\max_{\bar{\mathcal{H}}}\big\|\bar{\mathcal{H}}_\ell\big\|_{\text{HS}}^2\right).$$

The square-sum of the smallest eigenvalues can be bounded in terms of $\mathrm{Tr}[\bar{\mathcal{H}}]$,

$$\big\|\bar{\mathcal{H}}_\ell\big\|_{\text{HS}}^2 = \sum_{i=1}^{\ell}\lambda_i^2 \le \ell\lambda_\ell^2 \le \ell\left(\frac{\mathrm{Tr}[\bar{\mathcal{H}}]}{d^2-\ell}\right)^2 = \frac{2\big\|\bar{\mathcal{H}}\big\|_1^2}{d^2} = 2.$$

The second inequality is because all eigenvalues are sorted in increasing order, and thus λ_ℓ is no greater than the average of $\lambda_{\ell+1},\dots,\lambda_{d^2}$, which is at most $\mathrm{Tr}[\bar{\mathcal{H}}]/(d^2-\ell)$. The proof is complete.

C.1. Proof of Theorem 23

We first recall the theorem.

Theorem 23 *Let $\frac{d^2}{2} \le \ell \le d^2-1$, $\mathcal{V} = (V_1,\dots,V_{d^2} = \mathbb{I}_d/\sqrt{d})$ be an orthonormal basis of $\mathbb{H}_d$, $V := [vec(V_1),\dots,vec(V_\ell)]$ and $\sigma_z,\sigma_{z'} \sim \mathcal{D}_\ell(\mathcal{V})$ defined in Definition 2. Then for $n < \frac{d^2}{6c^2\varepsilon^2}$,*

$$\mathbb{E}_{\sigma_z,\sigma_{z'}}\left[\,\exp\left\{nd\left\langle\bar{\Delta}_{z'},\bar{\mathcal{H}}(\bar{\Delta}_z)\right\rangle\right\}\,\right]-1 \le \exp\left\{\frac{c^2n^2\varepsilon^4}{\ell^2}\left\|V^\dagger\bar{\mathcal{C}}V\right\|_{HS}^2\right\}-1+\frac{4}{e^d}. \tag{15}$$

Proof First, we claim that due to the exponentially small probability of the bad event $\Delta_z+\rho_{\text{mm}}\notin\mathcal{P}_\varepsilon$ as stated in Proposition 3, we can consider Δ_z instead of the normalized perturbation $\bar{\Delta}_z$. The claim is proved at the end of this section.

Claim 25 *Let $\bar{\Delta}_z$ and Δ_z be defined in Definition 2, then*

$$\mathbb{E}_{z,z'}\left[\,\exp\left\{nd\left\langle\bar{\Delta}_{z'},\bar{\mathcal{H}}(\bar{\Delta}_z)\right\rangle\right\}\,\right] \le \mathbb{E}_{z,z'}\left[\,\exp\left\{nd\left\langle\Delta_{z'},\bar{\mathcal{H}}(\Delta_z)\right\rangle\right\}\,\right]+\frac{4}{e^d}.$$

We then apply a standard result on the moment generating function of Radamacher chaos.

Lemma 26 (Acharya et al. (2020b, Claim IV.17)) *Let θ,θ' be two independent random vectors distributed uniformly over $\{-1,1\}^\ell$. Then for any positive semi-definite real matrix H,*

$$\log\mathbb{E}_{\theta,\theta'}\Big[\exp\{\lambda\theta^\top H\theta'\}\Big] \le \frac{\lambda^2}{2}\frac{\|H\|_{HS}^2}{1-4\lambda^2\|H\|_{op}^2}, \quad \textit{for } 0\le\lambda<\frac{1}{2\|H\|_{op}}.$$

We now evaluate the inner product. Recall the Choi representation of $\bar{\mathcal{H}}$ is $\bar{\mathcal{C}} = \frac{1}{n}\sum_{i=1}^n\mathcal{C}_i$. Note that $\mathcal{C}_i$ and $\bar{\mathcal{C}}$ are p.s.d. Hermitian matrices, and the eigenvalues exactly match those of $\mathcal{H}_i$ and $\bar{\mathcal{H}}$ due to the homomorphism between $\mathbb{C}^{d^2}$ and $\mathbb{C}^{d\times d}$.

Setting $V = [\mathrm{vec}(V_1),\dots,\mathrm{vec}(V_\ell)] \in \mathbb{C}^{d^2\times\ell}$, we have $\mathrm{vec}(\Delta_z) = \frac{c\varepsilon}{\sqrt{d\ell}}Vz$. Thus,

$$\left\langle\Delta_z,\bar{\mathcal{H}}(\Delta_{z'})\right\rangle = \mathrm{vec}(\Delta_{z'})^\dagger\bar{\mathcal{C}}\mathrm{vec}(\Delta_{z'})$$

$$= \frac{c^2\varepsilon^2}{d\ell} z^\dagger V^\dagger \bar{\mathcal{C}} V z'.$$

We now show that $H := V^\dagger \bar{\mathcal{C}} V$ is a real matrix when each V_i is a Hermitian matrix. First note that $\bar{\mathcal{C}} V = [\mathrm{vec}(\bar{\mathcal{H}}(V_1)), \ldots, \mathrm{vec}(\bar{\mathcal{H}}(V_\ell))]$. Therefore the i, j the element in H is

$$H_{ij} = \mathrm{vec}(V_i)^\dagger \mathrm{vec}(\bar{\mathcal{H}}(V_j)) = \left\langle V_i, \bar{\mathcal{H}}(V_j) \right\rangle \in \mathbb{R}. \tag{16}$$

We use the fact that $\bar{\mathcal{H}}$ is Hermiticity preserving and thus $\bar{\mathcal{H}}(V_j)$ is Hermitian. Since $\mathbb{H}_d$ is a real Hilbert space, the inner product is a real number.

We then set $\lambda = \frac{c^2 n \varepsilon^2}{\ell}$ and $H = V^\dagger \bar{\mathcal{C}} V$ in Lemma 26. Then $\|H\|_{\mathrm{op}} \leq \left\|\bar{\mathcal{C}}\right\|_{\mathrm{op}} = \left\|\bar{\mathcal{H}}\right\|_{\mathrm{op}} \leq 1$ due to Lemma 12. Thus for $n < \frac{\ell}{3c^2\varepsilon^2}$, we have

$$\lambda \|H\|_{\mathrm{op}} \leq \lambda < \frac{1}{3} \implies \frac{\lambda^2}{2(1 - 4\lambda^2 \|H\|_{\mathrm{op}}^2)} \leq \frac{9\lambda^2}{10} < \lambda^2.$$

Hence, applying Lemma 26

$$\mathbb{E}_{z,z'}\left[\exp\left\{\frac{c^2 n \varepsilon^2}{\ell} z^\top H z'\right\}\right] \leq \exp\{\lambda^2 \|H\|_{\mathrm{HS}}^2\} = \exp\left\{\frac{c^4 n^2 \varepsilon^4}{\ell^2} \|H\|_{\mathrm{HS}}^2\right\}.$$

Combining with Claim 25 proves Theorem 23. ■

Proof [Claim 25] Note that $\bar{\Delta}_z = a_z \Delta_z$, where

$$a_z := \min\left\{1, \frac{1}{d\|\Delta_z\|_{\mathrm{op}}}\right\} \in [0, 1].$$

Therefore,

$$\left\langle \bar{\Delta}_{z'}, \bar{\mathcal{H}}(\bar{\Delta}_z) \right\rangle = a_z a_{z'} \left\langle \Delta_{z'}, \bar{\mathcal{H}}(\Delta_z) \right\rangle.$$

As a short hand let $f(z, z') = nd \left\langle \Delta_{z'}, \bar{\mathcal{H}}(\Delta_z) \right\rangle$. Denote event E as $f(z, z') < 0$ and $a_z a_{z'} < 1$. When this event occors, $\exp\{a_z a_{z'} f(z, z')\} \leq 1$. Using Proposition 3, let $\delta = 2\exp(-d)$,

$$\Pr[\, a_z < 1 \,] \leq \delta.$$

Thus, by the union bound,

$$\Pr[\, E \,] = \Pr[\, a_z a_{z'} < 1 \,] = \Pr[\, a_z < 1 \text{ or } a_{z'} < 1 \,] \leq 2\delta.$$

Note that E^c denotes the event that $f(z, z') \geq 0$ or $a_z a_z' = 1$. When this occurs, $a_z a_{z'} f(z, z') \leq f(z, z')$. Thus,

$$\begin{aligned}
&\mathbb{E}_{z,z'}\left[\, \exp\left\{a_z a_{z'} f(z, z')\right\} \,\right] \\
&= \mathbb{E}_{z,z'}\left[\, \exp\left\{a_z a_{z'} f(z, z')\right\} \,\middle|\, E^c \,\right] \Pr[\, E^c \,] + \mathbb{E}_{z,z'}\left[\, \exp\left\{a_z a_{z'} f(z, z')\right\} \,\middle|\, E \,\right] \Pr[\, E \,] \\
&\leq \mathbb{E}_{z,z'}\left[\, \exp\left\{f(z, z')\right\} \,\middle|\, E^c \,\right] \Pr[\, E^c \,] + 2\delta \\
&\leq \mathbb{E}_{z,z'}\left[\, \exp\left\{f(z, z')\right\} \,\right] + 2\delta,
\end{aligned}$$

as desired. The second-to-last inequality uses $a_z f_z' f(z, z') \leq 0$ when event E happens, and the final inequality uses $\exp\{f(z, z')\} > 0$ and therefore

$$\begin{aligned}
\mathbb{E}_{z,z'}\left[\, \exp\left\{f(z, z')\right\} \,\right] &= \mathbb{E}_{z,z'}\left[\, \exp\left\{f(z, z')\right\} \,\middle|\, E^c \,\right] \Pr[\, E^c \,] + \mathbb{E}_{z,z'}\left[\, \exp\left\{f(z, z')\right\} \,\middle|\, E \,\right] \Pr[\, E \,] \\
&\geq \mathbb{E}_{z,z'}\left[\, \exp\left\{f(z, z')\right\} \,\middle|\, E^c \,\right] \Pr[\, E^c \,].
\end{aligned}$$

Plugging in the definition of a_z and $f(z, z')$ completes the proof. ■

C.2. Proof of Observation 24

Recall the statement of the observation.

Observation 24 *For all orthonormal basis $\mathcal{V}$, we have $\left\|V^\dagger \bar{\mathcal{C}} V\right\|_{HS} \leq \left\|\bar{\mathcal{H}}\right\|_{HS}$. However when $\mathcal{V} = \mathcal{V}_{\bar{\mathcal{H}}}$ in Lemma 12, for all $\frac{d^2}{2} \leq \ell \leq d^2 - 1$, $\left\|V^\dagger \bar{\mathcal{C}} V\right\|_{HS} = \left\|\bar{\mathcal{H}}_\ell\right\|_{HS} := \sqrt{\sum_{i=1}^{\ell} \lambda_i^2}$ and $0 \leq \lambda_1 \leq \ldots \leq \lambda_{d^2} = 1$ are the eigenvalues of $\bar{\mathcal{H}}$.*

Proof Since $V_1, \ldots, V_{d^2}$ is an orthonormal basis, we have $V^\dagger V = \mathbb{I}_\ell$, and thus V is an isometry and $\|V\|_{\text{op}} = \left\|V^\dagger\right\|_{\text{op}} = 1$. Using $\|AB\|_{\text{HS}} \leq \|A\|_{\text{op}} \|B\|_{\text{HS}}$, we obtain

$$\left\|V^\dagger \bar{\mathcal{C}} V\right\|_{\text{HS}} \leq \|\bar{\mathcal{C}}\|_{\text{HS}} = \|\bar{\mathcal{H}}\|_{\text{HS}}.$$

When $\mathcal{V} = \mathcal{V}_{\bar{\mathcal{H}}}$, we note that $V^\dagger \bar{\mathcal{C}} V = D_\ell := \text{diag}\{\lambda_1, \ldots, \lambda_\ell\}$, and $\|D_\ell\|_{\text{HS}}^2 = \sum_{i=1}^{\ell} \lambda_i^2$. Indeed, as derived in (16),

$$H_{ij} = \left\langle V_i, \bar{\mathcal{H}}(V_j) \right\rangle = \lambda_j \left\langle V_i, V_j \right\rangle = \lambda_j \delta_{ij}.$$

Therefore,

$$\|H\|_{\text{HS}}^2 \leq \begin{cases} \|D_\ell\|_{\text{HS}}^2 = \sum_{i=1}^{\ell} \lambda_i^2, & \mathcal{V} = \mathcal{V}_{\bar{\mathcal{H}}}, \\ \left\|\bar{\mathcal{H}}\right\|_{\text{HS}}^2, & \text{otherwise.} \end{cases}$$

■

Appendix D. Upper bound for fixed measurements

The algorithm we present is similar to an algorithm proposed in (Yu, 2021, Algorithm 4)[8]. They specifically work with maximal mutually unbiased bases Klappenecker and Rotteler (2005), and we work with quantum 2-designs, which are generalizations of the former. For completeness, we present the algorithm and its copy complexity guarantee.

The algorithm is based on quantum 2-designs, a finite set of vectors that preserves the second moment of the Haar measure and yields a rank-1 POVM with appropriate scaling. The same measurement is applied to *all* copies. Since it preserves the statistics of the Haar measure, one can show that when ρ and ρ_0 are far, then the outcome distribution on each copy is also far in ℓ_2 distance. From this, we apply classical closeness testing to the outcomes. As long as the 2-design has size at most $O(d^2)$, then we can achieve the desired $O(d^2/\varepsilon^2)$ copy complexity. For d that are prime powers, such 2-design exists due to maximal mutually unbiased bases Klappenecker and Rotteler (2005). This is already general enough since the system dimension d is 2^N for quantum computers implemented in qubits. Moreover, the algorithm can be easily generalized to the problem of *closedness* testing, where the goal is to test whether two *unknown* states ρ and σ are close in trace distance given n copies from each.

8. We came across the result after writing a draft of the paper. However, given the similarity of the algorithms, it should be attributed to Yu (2021).

D.1. Preliminaries

Quantum t-designs. At a high level, for an integer $t > 0$, t-design is a finite set of unit vectors such that the average of any polynomial f of degree at most t is the same as the expectation of f over the Haar measure.

Definition 27 (Quantum t-design) *Let t be a positive integer, we say that a finite set of normalized vectors $\{|\psi_x\rangle\}_{x=1}^k$ in $\mathbb{C}^d$ and a discrete distribution $q = (q_1, \ldots, q_k)$ over $[k]$ a quantum t-design if*

$$\sum_{x=1}^{k} q_x |\psi_x\rangle\langle\psi_x|^{\otimes t} = \int |\psi\rangle\langle\psi|^{\otimes t} d\mu(\psi),$$

where μ is the Haar measure on the unit sphere in $\mathbb{C}^d$. If $q_x = 1/k$, then the t-design is proper and we may omit the distribution q when describing proper t-designs.

By taking the partial trace on both sides, we can easily see that a t-design is naturally a t'-design for all $t' \leq t$. Moreover, when $t = 1$, the right-hand side is $\mathbb{I}_d/d$ and thus $\{dq_x|\psi_x\rangle\langle\psi_x|\}$ is a POVM. An important example of spherical 2-design is based on mutually unbiased bases (MUB) (see Durt et al. (2010) for a survey).

Theorem 28 (Klappenecker and Rotteler (2005)) *Let d be a prime power, then there exists a maximal MUB, i.e. $d + 1$ orthonormal bases $\{|\psi_x^l\rangle\}_{x=1}^d, l = 1, \ldots, d + 1$ such that the collection of all vectors $\{|\psi_x^l\rangle\}_{x,l}$ is a proper 2-design.*

Classical distribution testing We will use the classical closeness testing algorithm for discrete distributions as a sub-routine. Given two distributions $\mathbf{p}$ and $\mathbf{q}$ and n samples from each, the goal is to test whether $\mathbf{p} = \mathbf{q}$ or $\|\mathbf{p} - \mathbf{q}\|_2 \geq \varepsilon$. The sample complexity guarantee is given by the following theorem.

Theorem 29 ((Diakonikolas and Kane, 2016, Lemma 2.3),(Chan et al., 2014, Proposition 3.1)) *Let $\mathbf{p}, \mathbf{q}$ be unknown distributions over k such that $\min\{\|p\|_2, \|q\|_2\} \leq b$. There exists an algorithm TestClosenessL2($\mathbf{x}, \mathbf{x}', \varepsilon$) that distinguishes whether $\mathbf{p} = \mathbf{q}$ or $\|\mathbf{p} - \mathbf{q}\|_2 > \varepsilon$, where $\mathbf{x}$ and $\mathbf{x}'$ are $O(b/\varepsilon^2)$ samples from $\mathbf{p}$ and $\mathbf{q}$ respectively.*

D.2. Algorithm

The algorithm applies a proper 2-design for all copies, with suitable coefficients to make the projection matrices a POVM. 2-designs preserve the statistics of the Haar measure up to order 2, and therefore should be a good choice for fixed measurements.

Theorem 30 *Let k be the size of the proper 2-design used in Algorithm 1. With $n = O\left(d\sqrt{k}/\varepsilon^2\right)$ copies from each unknown state, Algorithm 1 can test whether $\rho = \rho_0$ or $\|\rho - \rho_0\|_1 > \varepsilon$ with probability at least $2/3$.*

Proof Let $\Delta = \rho - \rho_0$ and $\mathbf{p}_\rho$ be the distribution of a single measurement outcome for $\mathcal{M}$. When $\|\rho - \rho_0\|_1 = \|\Delta\|_1 \geq \varepsilon$, we have $\|\Delta\|_{\text{HS}} = \sqrt{\text{Tr}[\Delta^2]} \geq \varepsilon/\sqrt{d}$.

Algorithm 1: State certification/closedness testing without shared randomness

Input: n copies of unknown state ρ. If ρ_0 is unknown, n copies of ρ_0 as well.
Output YES if $\rho = \rho_0$, NO if $\|\rho - \rho_0\|_1 > \varepsilon$.
Let $\{|\psi_x\rangle\}_{x=1}^k$ be a proper 2-design and $\mathcal{M} = \{\frac{d}{k}|\psi_x\rangle\langle\psi_x|\}_{x=1}^k$
Apply the measurement $\mathcal{M}$ for all copies of ρ and obtain outcomes $\mathbf{x} = (x_1, \ldots, x_n)$.
Obtain n samples $\mathbf{x}' = (x'_1, \ldots, x'_n)$ from $\mathbf{p}_{\rho_0}$. If ρ_0 is known, then $\mathbf{x}'$ is obtained by measuring each copy with $\mathcal{M}$. Else, $\mathbf{x}'$ is sampled using classical randomness.
return TestClosenessL2$(\mathbf{x}, \mathbf{x}', \varepsilon/\sqrt{k(d+1)})$.

We can compute the $\|\mathbf{p}_\rho\|_2$ and $\|\mathbf{p}_\rho - \mathbf{p}_{\rho_0}\|_2$ in terms of Δ.

$$\|\mathbf{p}_\rho\|_2^2 = \frac{d^2}{k^2}\sum_{x=1}^k \langle\psi_x|\rho|\psi_x\rangle^2, \quad \|\mathbf{p}_\rho - \mathbf{p}_{\rho_{\text{mm}}}\|_2^2 = \frac{d^2}{k^2}\sum_{x=1}^k \langle\psi_x|\Delta|\psi_x\rangle^2.$$

Note that $\{|\psi_x\rangle\}_{x=1}^k$ is a proper 2-design, and thus by definition for all Hermitian matrices M,

$$\begin{aligned}
\frac{1}{k}\sum_{x=1}^k \langle\psi_x|M|\psi_x\rangle^2 &= \frac{1}{k}\sum_x \mathrm{Tr}[|\psi_x\rangle\langle\psi_x|M]^2 = \mathrm{Tr}\left[\frac{1}{k}\sum_x |\psi_x\rangle\langle\psi_x|^{\otimes 2} M^{\otimes 2}\right] \\
&= \mathrm{Tr}\left[\mathbb{E}_{\psi\sim\mu}\left[\,|\psi\rangle\langle\psi|^{\otimes 2}\right] M^{\otimes 2}\,\right] = \mathbb{E}_{\psi\sim\mu}\left[\,\mathrm{Tr}[|\psi\rangle\langle\psi|^{\otimes 2} M^{\otimes 2}]\,\right] \\
&= \mathbb{E}_{\psi\sim\mu}\left[\,\mathrm{Tr}[|\psi\rangle\langle\psi|M]^2\,\right] \\
&= \mathbb{E}_{\psi\sim\mu}\left[\,\langle\psi|M|\psi\rangle^2\,\right],
\end{aligned}$$

where μ is the Haar measure. The expectation can be computed using Weingarten calculus Collins (2003); Collins and Śniady (2006).

Lemma 31 *For any Hermitian $M \in \mathbb{C}^{d\times d}$ and $|\psi\rangle \sim \mu$ the Haar measure, we have,*

$$\mathbb{E}_{\psi\sim\mu}\left[\,\langle\psi|M|\psi\rangle^2\,\right] = \frac{1}{d(d+1)}(\mathrm{Tr}[M]^2 + \mathrm{Tr}[M^2]).$$

The proof can be found in (Chen et al., 2022b, Lemma 6.4). Since $\mathrm{Tr}[\rho^2] \le \mathrm{Tr}[\rho] = 1$ and $\mathrm{Tr}[\Delta] = 0$, from this lemma we conclude that

$$\|\mathbf{p}_\rho\|_2^2 \le \frac{2d}{k(d+1)}, \quad \|\mathbf{p}_\rho - \mathbf{p}_{\rho_{\text{mm}}}\|_2^2 = \frac{d^2\,\mathrm{Tr}[\Delta^2]}{kd(d+1)} \ge \frac{\varepsilon^2}{k(d+1)}.$$

Therefore, we can apply Theorem 29 with domain size k, $b \leftarrow \sqrt{\frac{2d}{k(d+1)}}$, and $\varepsilon \leftarrow \frac{\varepsilon}{\sqrt{k(d+1)}}$. The number of samples n required is

$$n = O\left(\sqrt{\frac{2d}{k(d+1)}}\cdot\frac{k(d+1)}{\varepsilon^2}\right) = O\left(\frac{\sqrt{kd(d+1)}}{\varepsilon}\right).$$

■

The upper bound part of Theorem 1 is an immediate corollary of the above theorem.

Corollary 32 *If the size of the proper 2-design in Algorithm 1 is $k = O(d^2)$, then $n = O(d^2/\varepsilon^2)$ copies are sufficient for Algorithm 1. Specifically, when d is a prime power, such 2-design exists due to maximal MUB which satisfies $k = d(d+1)$.*

This result suggests that the optimal copy complexity of $O(d^2/\varepsilon^2)$ can be generalized to dimensions d other than prime powers. For example, SIC-POVM Zauner (1999); Renes et al. (2004) is a minimal 2-design with $k = d^2$ and is known to exist for $d = 2$ to 28 and as high as $d = 1299$ DeBrota (2020). It has been conjectured in Zauner (1999) that SIC-POVMs exist for all d. If the conjecture is proved, then Algorithm 1 naturally generalizes to all d.